



# BRAIN STATES AND NEURAL MECHANISMS OF CONSCIOUSNESS

EDITED BY: Alain Destexhe, Kathinka Evers, Tommaso Fellin, Olivia Gosseries,  
Steven Laureys, Marcello Massimini, Maurizio Mattia,  
Cyriel Pennartz, Maria V. Sanchez-Vives, Johan F. Storm,  
Igor Timofeev and Vladyslav Vyazovski

PUBLISHED IN: Frontiers in Systems Neuroscience



# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-83250-123-8

DOI 10.3389/978-2-83250-123-8

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [frontiersin.org/about/contact](https://frontiersin.org/about/contact)



# BRAIN STATES AND NEURAL MECHANISMS OF CONSCIOUSNESS

Topic Editors:

**Alain Destexhe**, FRE3693 Unité de Neurosciences, Information et Complexité (UNIC), France

**Kathinka Evers**, Uppsala University, Sweden

**Tommaso Fellin**, Italian Institute of Technology (IIT), Italy

**Olivia Gosseries**, University of Liège, Belgium

**Steven Laureys**, University of Liège, Belgium

**Marcello Massimini**, University of Milan, Italy

**Maurizio Mattia**, National Institute of Health (ISS), Italy

**Cyriel Pennartz**, University of Amsterdam, Netherlands

**Maria V. Sanchez-Vives**, Institut de Recerca Biomèdica August Pi i Sunyer (IDIBAPS), Spain

**Johan F. Storm**, University of Oslo, Norway

**Igor Timofeev**, Laval University, Canada

**Vladislav Vyazovskiy**, University of Oxford, United Kingdom

**Citation:** Destexhe, A., Evers, K., Fellin, T., Gosseries, O., Laureys, S., Massimini, M., Mattia, M., Pennartz, C., Sanchez-Vives, M. V., Storm, J. F., Timofeev, I., Vyazovskiy, V., eds. (2022). Brain States and Neural Mechanisms of Consciousness. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-123-8

# Table of Contents

- 06    *Reappearance of Command-Following Is Associated With the Recovery of Language and Internal-Awareness Networks: A Longitudinal Multiple-Case Report***  
Charlène Aubinet, Rajanikant Panda, Stephen Karl Larroque, Helena Cassol, Mohamed Ali Bahri, Manon Carrière, Sarah Wannez, Steve Majerus, Steven Laureys and Aurore Thibaut
- 12    *Attention to Monocular Images Bias Binocular Rivalry***  
Manuel Moreno-Sánchez, J. Antonio Aznar-Casanova and Fernando Valle-Inclán
- 20    *Outer Brain Oscillations in Down Syndrome***  
Marcel Ruiz-Mejias
- 28    *Coherence of Visual-Evoked Gamma Oscillations Is Disrupted by Propofol but Preserved Under Equipotent Doses of Isoflurane***  
Adeeti Aggarwal, Connor Brennan, Brenna Shortal, Diego Contreras, Max B. Kelz and Alex Proekt
- 41    *Evaluating Complexity of Fetal MEG Signals: A Comparison of Different Metrics and Their Applicability***  
Julia Moser, Siouar Bensaid, Eleni Kroupi, Franziska Schleger, Fabrice Wendling, Giulio Ruffini and Hubert Preißl
- 56    *Physically Sufficient Neural Mechanisms of Consciousness***  
Matthew Owen and Mihretu P. Guta
- 70    *Characterizing the Dynamical Complexity Underlying Meditation***  
Anira Escrichs, Ana Sanjuán, Selen Atasoy, Ane López-González, César Garrido, Estela Càmarà and Gustavo Deco
- 76    *Indicators and Criteria of Consciousness in Animals and Intelligent Machines: An Inside-Out Approach***  
Cyriel M. A. Pennartz, Michele Farisco and Kathinka Evers
- 99    *Scaling of a Large-Scale Simulation of Synchronous Slow-Wave and Asynchronous Awake-Like Activity of a Cortical Model With Long-Range Interconnections***  
Elena Pastorelli, Cristiano Capone, Francesco Simula, Maria V. Sanchez-Vives, Paolo Del Giudice, Maurizio Mattia and Pier Stanislao Paolucci
- 115    *The Role of Top-Down Modulation in Shaping Sensory Processing Across Brain States: Implications for Consciousness***  
Tom Sikkens, Conrado A. Bosman and Umberto Olcese
- 130    *General Anesthesia: A Probe to Explore Consciousness***  
Vincent Bonhomme, Cécile Staquet, Javier Montupil, Aline Defresne, Murielle Kirsch, Charlotte Martial, Audrey Vanhaudenhuyse, Camille Chatelle, Stephen Karl Larroque, Federico Raimondo, Athena Demertzi, Olivier Bodart, Steven Laureys and Olivia Gosseries

- 145 ***Exploring the Phase-Locking Mechanisms Yielding Delayed and Anticipated Synchronization in Neuronal Circuits***  
Leonardo Dalla Porta, Fernanda S. Matias, Alfredo J. dos Santos, Ana Alonso, Pedro V. Carelli, Mauro Copelli and Claudio R. Mirasso
- 154 ***Coupling the State and Contents of Consciousness***  
Jaan Aru, Mototaka Suzuki, Renate Rutiku, Matthew E. Larkum and Talis Bachmann
- 163 ***Sleep–Wake Cycle in Young and Older Mice***  
Sara Soltani, Sylvain Chauvette, Olga Bukhtiyarova, Jean-Marc Lina, Jonathan Dubé, Josée Seigneur, Julie Carrier and Igor Timofeev
- 177 ***A Naturalistic Approach to the Hard Problem of Consciousness***  
Wolf Singer
- 186 ***COALIA: A Computational Model of Human EEG for Consciousness Research***  
Siouar Bensaid, Julien Modolo, Isabelle Merlet, Fabrice Wendling and Pascal Benquet
- 204 ***Analysis Pipeline for Extracting Features of Cortical Slow Oscillations***  
Giulia De Bonis, Miguel Dasilva, Antonio Paziienti, Maria V. Sanchez-Vives, Maurizio Mattia and Pier Stanislao Paolucci
- 219 ***Bridging Single Neuron Dynamics to Global Brain States***  
Jennifer S. Goldman, Núria Tort-Colet, Matteo di Volo, Eduarda Susin, Jules Bouté, Melissa Dali, Mallory Carlu, Trang-Anh Nghiem, Tomasz Górski and Alain Destexhe
- 228 ***Principles of Neurorehabilitation After Stroke Based on Motor Learning and Brain Plasticity Mechanisms***  
Martina Maier, Belén Rubio Ballester and Paul F. M. J. Verschure
- 246 ***The Molecular Organization of Self-awareness: Paralimbic Dopamine-GABA Interaction***  
Hans C. Lou, Kristine Rømer Thomsen and Jean-Pierre Changeux
- 251 ***Multiple Network Disconnection in Anosognosia for Hemiplegia***  
Elena Monai, Francesca Bernocchi, Marta Bisio, Antonio Luigi Bisogno, Alessandro Salvalaggio and Maurizio Corbetta
- 263 ***Perspective on the Multiple Pathways to Changing Brain States***  
Malinda L. S. Tantirigama, Timothy Zolnik, Benjamin Judkewitz, Matthew E. Larkum and Robert N. S. Sachdev
- 273 ***Up and Down States and Memory Consolidation Across Somatosensory, Entorhinal, and Hippocampal Cortices***  
John J. Tukker, Prateep Beed, Dietmar Schmitz, Matthew E. Larkum and Robert N. S. Sachdev
- 290 ***The (Un)Conscious Mouse as a Model for Human Brain Functions: Key Principles of Anesthesia and Their Impact on Translational Neuroimaging***  
Henning M. Reimann and Thoralf Niendorf
- 332 ***Late Positivity Does Not Meet the Criteria to be Considered a Proper Neural Correlate of Perceptual Awareness***  
Chiara Mazzi, Gaetano Mazzeo and Silvia Savazzi
- 346 ***Effects of Cortical Cooling on Activity Across Layers of the Rat Barrel Cortex***  
Gulshat Burkhanova, Kseniya Chernova, Roustem Khazipov and Maxim Sheroziya

- 363** *Accounting for Changing Structure in Functional Network Analysis of TBI Patients*  
John Dell'Italia, Micah A. Johnson, Paul M. Vespa and Martin M. Monti
- 381** *Decreased Evoked Slow-Activity After tDCS in Disorders of Consciousness*  
Armand Mensen, Olivier Bodart, Aurore Thibaut, Sarah Wannez, Jitka Annen, Steven Laureys and Olivia Gosseries
- 391** *Segregated Co-activation Patterns in the Emergence of Decision Confidence During Visual Perception*  
Cilia Jaeger, Sarah Glim, Cristiana Dimulescu, Anja Ries, Christian Sorg and Afra Wohlschläger
- 403** *Detecting (Un)seen Change: The Neural Underpinnings of (Un)conscious Prediction Errors*  
Elise G. Rowe, Naotsugu Tsuchiya and Marta I. Garrido
- 418** *Up and Down States During Slow Oscillations in Slow-Wave Sleep and Different Levels of Anesthesia*  
Melody Torao-Angosto, Arnau Manasanch, Maurizio Mattia and Maria V. Sanchez-Vives
- 427** *Analyzing the Loss and the Recovery of Consciousness: Functional Connectivity Patterns and Changes in Heart Rate Variability During Propofol-Induced Anesthesia*  
Davide Sattin, Dunja Duran, Sergio Visintini, Elena Schiaffi, Ferruccio Panzica, Carla Carozzi, Davide Rossi Sebastiano, Elisa Visani, Eleonora Tobaldini, Angelica Carandina, Valeria Citterio, Francesca Giulia Magnani, Martina Cacciatore, Eleonora Orena, Nicola Montano, Dario Caldiroli, Silvana Franceschetti, Mario Picozzi and Leonardi Matilde
- 443** *Dynamic Patterns of Global Brain Communication Differentiate Conscious From Unconscious Patients After Severe Brain Injury*  
Daniel Golkowski, Rebecca Willnecker, Jennifer Rösler, Andreas Ranft, Gerhard Schneider, Denis Jordan and Rüdiger Ilg
- 452** *Criticality Creates a Functional Platform for Network Transitions Between Internal and External Processing Modes in the Human Brain*  
Minkyung Kim, Hyoungkyu Kim, Zirui Huang, George A. Mashour, Denis Jordan, Rüdiger Ilg and UnCheol Lee
- 471** *EEG-Microstates Reflect Auditory Distraction After Attentive Audiovisual Perception Recruitment of Cognitive Control Networks*  
Ute Korn, Marina Krylova, Kilian L. Heck, Florian B. Hübner, Robert S. Stark, Sarah Alizadeh, Hamidreza Jamalabadi, Martin Walter, Ralf A. W. Galuske and Matthias H. J. Munk
- 480** *Holographic Duality and the Physics of Consciousness*  
Uziel Awret
- 496** *Brain Activity Characteristics of Patients With Disorders of Consciousness in the EEG Resting State Paradigm: A Review*  
Anna Duszyk-Bogorodzka, Magdalena Zieleniewska and Kamila Jankowiak-Siuda



# Reappearance of Command-Following Is Associated With the Recovery of Language and Internal-Awareness Networks: A Longitudinal Multiple-Case Report

Charlène Aubinet<sup>1\*</sup>, Rajanikant Panda<sup>1</sup>, Stephen Karl Larroque<sup>1</sup>, Helena Cassol<sup>1</sup>, Mohamed Ali Bahri<sup>2</sup>, Manon Carrière<sup>1</sup>, Sarah Wannez<sup>1</sup>, Steve Majerus<sup>3</sup>, Steven Laureys<sup>3†</sup> and Aurore Thibaut<sup>1\*†</sup>

<sup>1</sup> Coma Science Group, Department of GIGA Consciousness and Neurology, University and University Hospital of Liège, Liège, Belgium, <sup>2</sup> GIGA-Cyclotron Research Centre In Vivo Imaging, University of Liège, Liège, Belgium, <sup>3</sup> Psychology and Neuroscience of Cognition Research Unit, University of Liège, Liège, Belgium

## OPEN ACCESS

### Edited by:

Mikhail Lebedev,  
Duke University, United States

### Reviewed by:

Marco Cambiaghi,  
Columbia University, United States  
Marek Binder,  
Jagiellonian University, Poland

### \*Correspondence:

Charlène Aubinet  
caubinet@uliege.be  
orcid.org/0000-0001-5095-5583  
Aurore Thibaut  
athibaut@uliege.be

<sup>†</sup> Co-last authors

**Received:** 17 September 2018

**Accepted:** 31 January 2019

**Published:** 26 February 2019

### Citation:

Aubinet C, Panda R, Larroque SK, Cassol H, Bahri MA, Carrière M, Wannez S, Majerus S, Laureys S and Thibaut A (2019) Reappearance of Command-Following Is Associated With the Recovery of Language and Internal-Awareness Networks: A Longitudinal Multiple-Case Report. *Front. Syst. Neurosci.* 13:8. doi: 10.3389/fnsys.2019.00008

The recovery of patients with disorders of consciousness is a real challenge, especially at the chronic stage. After a severe brain injury, patients can regain some slight signs of consciousness, while not being able to functionally communicate. This entity is called the minimally conscious state (MCS), which has been divided into MCS– and MCS+, respectively based on the absence or presence of language-related signs of consciousness. In this series of cases we aimed to describe retrospectively the longitudinal recovery of specific language-related behaviors using neuroimaging measurement in severely brain-injured patients. Among 209 chronic MCS patients admitted to our center from 2008 to 2018, 19 were assessed at two time points by means of behavioral and neuroimaging assessments. Three of them met our inclusion criteria and were diagnosed as MCS– during their first stay and had recovered command-following when they were reassessed (i.e., MCS+). As compared to their first assessments, when the three patients were in a MCS+, they showed less hypometabolism and/or higher gray matter volume in brain regions such as the precuneus and thalamus, as well as the left caudate and temporal/angular cortices known to be involved in various aspects of semantics. According to these preliminary results, the reappearance of language-related behaviors was concomitant with the recovery of metabolism and gray matter in neural regions that have been associated with self-consciousness and language processing. Prospective studies should be conducted to deepen our understanding of the neural correlates of the recovery of language-related behaviors in chronic MCS.

**Keywords:** minimally conscious state, language, consciousness, positron emission tomography, structural magnetic resonance imaging

## INTRODUCTION

The minimally conscious state (MCS) was defined in 2002 to distinguish unconscious patients from those presenting reproducible behavioral signs of consciousness (Giacino et al., 2002), and was later subcategorized on the basis of the absence or presence of language-related behaviors (Bruno et al., 2011). Patients in MCS– show non-reflexive behaviors, most frequently visual pursuit and fixation, oriented movements or localization to pain (Wannez et al., 2017a). Patients in MCS+ additionally demonstrate command following, intelligible verbalization or intentional communication (Bruno et al., 2011). These criteria require a certain residual language. For example, patients need to be able to understand what is asked in order to respond to verbal commands, and residual semantic processing is necessary.

A left-lateralized network was identified in a review investigating the neural correlates of semantics using functional neuroimaging studies. It is comprised of seven regions: posterior inferior parietal lobule (including angular gyrus), middle temporal gyrus, fusiform and parahippocampal gyri, dorsomedial prefrontal cortex, inferior frontal gyrus, ventromedial prefrontal cortex, and posterior cingulate gyrus (Binder et al., 2009). As regards to severely brain-injured patients, two neuroimaging studies, using either positron emission tomography (PET), or functional magnetic resonance imaging (MRI), demonstrated that MCS+ patients, as compared to MCS– patients, present a higher brain metabolism and functional connectivity in language-related areas, such as Broca's and Wernicke's regions or the left fusiform gyrus (Bruno et al., 2012; Aubinet et al., 2018). Nevertheless, it is still undetermined if behavioral improvement from MCS– to MCS+ is solely the consequence of the recovery of language-related functions, and therefore, due to an increase in brain functions in related areas, or if it is the combination of language and consciousness recuperation.

In the present study, we examine whether the MCS sub-categorization also characterizes the trajectory of recovery of individual patients when followed longitudinally. We demonstrate here in three MCS patients the transition from the MCS– to MCS+ based on repeated neuroimaging and behavioral assessments.

## MATERIALS AND METHODS

Between 2008 and 2018, 209 brain-damaged patients subsequently diagnosed as MCS were admitted into the University Hospital of Liège, including 19 patients who were assessed at two time points using neuroimaging techniques and repeated Coma Recovery Scale-Revised (CRS-R) (Giacino et al., 2004; Wannez et al., 2017b). Inclusion criteria were: >28 days post-injury when first assessed; diagnosis made based on at least 5 CRS-R, diagnosed in MCS– during the first week of assessment and later diagnosed in MCS+ during the second week of assessments. Three patients met our inclusion criteria (age: 23–37 years-old, two TBI, time since onset: 10 months to 5 years). These three patients in MCS– during their first week of

assessment, later recovered command following when reassessed during a second week of evaluations (26–31 months later).

The study was approved by the Ethics Committee of the Faculty of Medicine of the University of Liège and written informed consents for study participation and data publication have been obtained from the patients' legal representatives as well as from the healthy control subjects.

Fluorodeoxyglucose-PET data were acquired on a Gemini TF CT scanner (Philips Medical Systems) and preprocessed as described elsewhere (Thibaut et al., 2012). The data were smoothed with an isotropic 14 mm full-width at half-maximum (FWHM) Gaussian kernel and analyzed using statistical parametric mapping 12 (SPM12). A patient specific template has been used as proposed in a previous study (Phillips et al., 2011). The design matrices included the two repeated scans of each patient and the scans of healthy controls ( $n = 34$ , age range 19–70 years old, 15 women). Sample *t*-tests were used to assess the fluorodeoxyglucose metabolism differences between patients and controls. Results were considered significant at  $p < 0.05$  false discovery rate (FDR)-corrected.

Structural MRI data were obtained with T1-weighted 3D gradient echo sequence (120 slices, repetition time 2.3 s, echo time 2.47 ms, voxel size  $1 \text{ mm}^3 \times 1 \text{ mm}^3 \times 1.2 \text{ mm}^3$ , flip angle  $9^\circ$ , field of view  $256 \text{ mm} \times 256 \text{ mm}^2$ ). A T1 voxel-based morphometry (VBM) analysis (Ashburner and Friston, 2000) was carried with VBM8<sup>a</sup> toolbox, with non-linear modulation and a study-specific DARTEL (Ashburner, 2007) template generated on 50 patients of same population as our patients, other parameters set to defaults. Data were smoothed with a Gaussian kernel of 12 mm FWHM and then modeled using a factorial design with the total intracranial volume as covariate. An exclusive mask of the cerebrospinal fluid mask was used. A design matrix was constructed for each patient, including the two patient's scans (pre and post) and the scans of MRI-specific control subjects ( $n = 36$ , age range 20–75 years old, 13 women). Sample *t*-tests were conducted to assess the gray matter differences between patient and controls. Results were considered significant at FDR-corrected  $p < 0.05$  voxel-wise over the whole brain.

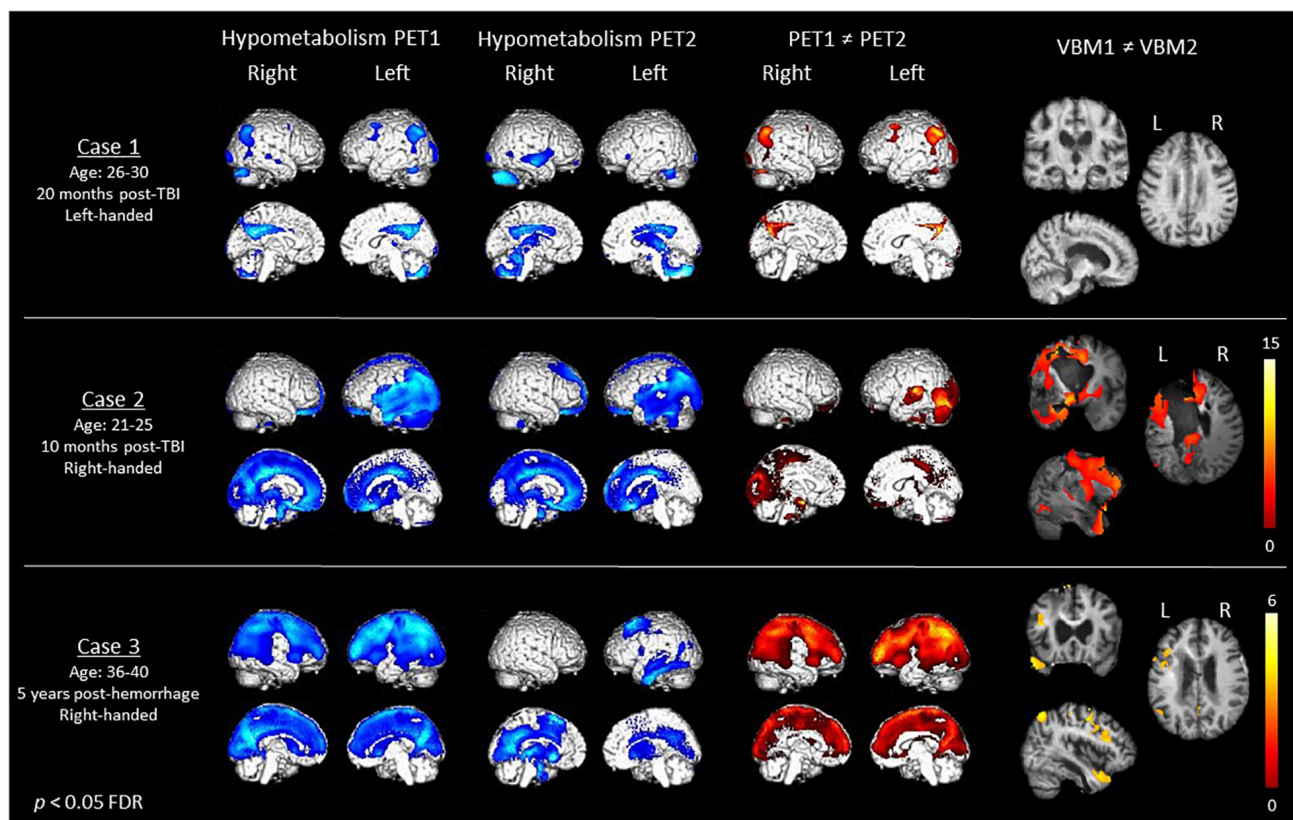
Statistical analyses identified: (a) brain areas showing hypometabolism (as compared with healthy control subjects) when the patient was in MCS–; (b) brain areas showing hypometabolism when the patient was in MCS+; (c) brain areas showing less hypometabolism or gray matter impairment when the patient was in MCS+ as compared to the scan when the patient was in MCS–.

## RESULTS

Demographical and neuroimaging results are presented in **Figure 1** and **Table 1**. Behavioral (i.e., CRS-R sub-scores) and neuroimaging details are reported in **Supplementary Materials I, II**.

At Time 1, the left-handed case 1 presented hypometabolism in the bilateral angular, fusiform and middle frontal gyri, left calcarine gyrus and right thalamus, cerebellum, inferior occipital, and middle temporal gyri. When





**FIGURE 1 |** Demographical data (first column), cerebral hypometabolism at Time 1 and Time 2 (second and third column in blue), recovery of metabolism (fourth column in red) as assessed with fluorodeoxyglucose-PET, as well as recovery of gray matter volume (fifth column) as assessed with MRI voxel-based morphometry in the three MCS patients. Color bars represent the T-values.

this patient was able to respond to simple commands 27 months later, we identified a significant recovery of metabolism in the bilateral angular and middle frontal gyri, left precuneus, middle occipital and calcarine gyri, right inferior occipital gyrus and cerebellum. Gray matter volume did not significantly increase or decrease between both examinations.

At Time 1, Case 2 presented hypometabolism mostly in the left superior medial gyrus and temporo-parieto-occipital cortex. Thirty-one months later, we observed command following, along with a significant increase of metabolism in the left temporal lobule, cerebellum and superior medial gyrus, as well as the thalamus. Gray matter volume was shown to be significantly increased in MCS+ than in MCS− in the bilateral caudate and the left fusiform, angular and middle/inferior temporal gyri.

Finally, at Time 1, case 3 presented hypometabolism in the left middle frontal and temporal gyri, left inferior parietal lobule, left rolandic operculum, and right thalamus. When he had recovered command following 26 months later, we identified a recovery of metabolism in the bilateral caudate and the left middle frontal gyrus. Between the two scans, gray matter volume was found to be more preserved in bilateral parahippocampal gyri, in the left inferior and superior temporal

cortex, caudate, inferior parietal lobule, precuneus and frontal cortex, as well as in the right thalamus, middle frontal, and temporal cortex.

## DISCUSSION

The present findings offer a new perspective to understand the neural correlates of the recovery of language-related behaviors, and more specifically command following (behavior recovered in all three patients). We here report that the recovery of this ability in three chronic MCS patients is paralleled with an increase in brain metabolism and gray matter in regions previously related to language and/or consciousness.

First, all patients presented an increase of brain metabolism in regions that have been associated with language processing. For example, Binder et al. (2009) considered the left angular gyrus as particularly involved in semantics (i.e., sentence comprehension, complex information integration and knowledge retrieval), and this region was less hypometabolic when case 1 was MCS+ as compared to MCS−. Moreover, a recovery of its contralateral homolog was reported. According to the same authors, the left temporal areas highlighted in case 2 were also shown to be activated in language tasks and to

**TABLE 1** | Coordinates of peak voxels of areas less hypometabolic/atrophic at PET2 as compared to PET1.

	Brain region	X (mm)	Y (mm)	Z (mm)	p value*	Z value
<b>Case 1</b>						
PET1 < PET2	Left precuneus	−4	−56	24	0.001	4.447
	Right angular	44	−66	48	0.006	3.889
	Left angular	−40	−72	44	0.006	3.872
	Left middle occipital gyrus	−34	−92	18	0.009	3.747
	Right inferior occipital gyrus	34	−94	−2	0.020	3.329
	Right cerebellum	38	−72	−26	0.023	3.260
	Left middle frontal gyrus	−32	10	54	0.023	3.249
	Left calcarine gyrus	−6	−102	−2	0.028	3.138
	Right middle frontal gyrus	48	8	52	0.044	2.867
<b>Case 2</b>						
PET1 < PET2	Left temporal pole	−32	2	−18	0.000	7.222
	Left cerebellum	−46	−42	−36	0.000	4.847
	Left superior medial frontal gyrus	0	52	20	0.003	3.089
	Right thalamus	16	−14	22	0.006	2.900
VBM1 < VBM2	Left caudate	−8	10	−12	0.000	5.733
	Right caudate	23	9	18	0.001	4.669
	Left temporal fusiform gyrus	−41	−65	−12	0.001	4.115
	Left middle temporal gyrus	−63	−53	−12	0.004	3.566
	Left inferior temporal gyrus	−68	−27	−20	0.006	3.430
	Left angular gyrus	−50	−65	39	0.029	2.672
<b>Case 3</b>						
PET1 < PET2	Left middle frontal gyrus	−40	4	56	0.000	6.147
	Right caudate	18	−8	24	0.001	3.505
	Left caudate	−14	8	14	0.002	3.346
VBM1 < VBM2	Right thalamus	17	−7	14	0.019	3.961
	Left inferior temporal gyrus	−66	−56	−9	0.019	3.928
	Left precuneus	2	−42	59	0.019	3.901
	Left superior frontal gyrus	−11	17	63	0.019	3.792
	Right middle temporal gyrus	66	−51	−2	0.020	3.759
	Right parahippocampal gyrus	21	−4	−18	0.020	3.745
	Left superior temporal gyrus	−51	18	−30	0.020	3.718
	Left parahippocampal gyrus	−23	−24	−35	0.020	3.638
	Left precentral gyrus	−38	−4	62	0.021	3.509
	Left inferior parietal lobule	−44	−68	51	0.021	3.495
	Left parahippocampal gyrus	−24	−24	−12	0.023	3.335
	Left caudate	−17	9	15	0.024	3.240
	Right middle frontal gyrus	32	48	3	0.027	3.123

\*FDR corrected. <sup>a</sup><http://www.neuro.uni-jena.de/vbm/>.

support concept retrieval. Finally, case 3 presented a better metabolism in the left caudate, which was associated with language control skills (Crinion et al., 2006; Friederici, 2006). Using structural MRI, patients 2 and 3 also showed a larger gray matter volume in language comprehension areas (i.e., left fusiform, angular, and temporal cortex) (Binder et al., 2009) when having recovered command following. Thus our findings are congruent with previous studies in MCS patients, showing a relationship between the ability to respond to commands and left hemisphere functions (Bruno et al., 2012; Aubinet et al., 2018). Nevertheless, we found that the structural and functional changes between MCS− and MCS+ might be bilateral. The non-dominant hemisphere could therefore contribute to the recovery

of command following. This is also consistent with previous studies showing compensatory mechanisms in the contralateral right hemisphere in conscious aphasic patients (Teki et al., 2013; Artzi et al., 2016).

Furthermore, the thalamus and/or precuneus showed an improvement of metabolism in the two first patients and an increase of gray matter structure in the third case. These regions are critical for consciousness processes, and especially self-consciousness as they are part of the internal consciousness network (Vanhaudenhuyse et al., 2011; Di Perri et al., 2016). Consequently, our results tend to confirm that, besides language processing differences, MCS− and MCS+ categories also reflect differentiated levels of consciousness.

It is however important to note that other higher cognitive functions might be involved in the sub-categorization of the MCS. Indeed, for all three cases an improvement was shown in various frontal areas related to executive functions (Miyake et al., 2000), or in regions such as the caudate, which is implicated in learning or cognitive control (Chiu et al., 2017). These skills could also be involved in the recovery of command following as assessed with the CRS-R.

Our results are not representative of the population and prospective studies including a large sample of patients are needed in order to overcome statistical limitations. Future longitudinal studies should also investigate the clinical implications on long-term outcomes of the transition from MCS– to MCS+, as it could be hypothesized that patients in MCS+ could have a better outcome than patients in MCS–, making the distinction between these two entities crucial. In addition, single subject VBM analysis is prone to false positive/type 1 error and group studies may limit this risk of false positive. Nevertheless, our findings point-out the possibility of command following recovery even in chronic MCS– patients, which seems in line with a recovery of brain functions in the ipsi- or contra-lateral language networks, as well as in the consciousness network. This is particularly relevant given the importance of this behavior for a potential communication (Giacino et al., 2002, 2004).

## CONCLUSION

The reappearance of command following in the three patients was concomitant with a recovery of metabolism and gray matter structure in language-related areas, mainly in the left temporal lobule, angular gyrus, fusiform gyrus, and caudate. It was also concomitant with functional and structural recovery of structures such as the thalamus and the precuneus involved in self-consciousness. The present results indicate that the transition

from MCS– to MCS+ involves recovery in networks particularly associated with both language and consciousness.

## AUTHOR CONTRIBUTIONS

CA, SW, and AT conceived and planned the presented research. CA, HC, and MC contributed to data acquirement. CA, RP, SL, MB, SM, SL, and AT worked on data analyses and interpretation. CA drafted the manuscript under AT's supervision. All authors provided critical feedback and helped to shape the manuscript.

## FUNDING

The study was supported by the University and University Hospital of Liège, the French Speaking Community Concerted Research Action (ARC 12-17/01), the Belgian National Funds for Scientific Research (FRS-FNRS), Human Brain Project (EU-H2020-fetflagship-hbp-sga1-ga720270), Luminous Project (EU-H2020-fetopen-ga686764), the James McDonnell Foundation, Mind Science Foundation, IAP research network P7/06 of the Belgian Government (Belgian Science Policy), the European Commission, the Public Utility Foundation “Université Européenne du Travail,” “Fondazione Europea di Ricerca Biomedica,” the Bial Foundation, Belgian National Plan Cancer (139).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnsys.2019.00008/full#supplementary-material>

## REFERENCES

- Artzi, M., Shiran, S. I., Weinstein, M., Myers, V., and Tarrasch, R. (2016). Cortical reorganization following injury early in life list of abbreviation. *Neural Plast.* 2016:8615872. doi: 10.1155/2016/8615872
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage* 38, 95–113. doi: 10.1016/j.neuroimage.2007.07.007
- Ashburner, J., and Friston, K. J. (2000). Voxel-based morphometry-the methods. *NeuroImage* 11, 805–821. doi: 10.1006/nimg.2000.0582
- Aubinet, C., Larroque, S. K., Heine, L., Martial, C., Majerus, S., Laureys, S., et al. (2018). Clinical subcategorization of minimally conscious state according to resting functional connectivity. *Hum. Brain Mapp.* 39, 4519–4532. doi: 10.1002/hbm.24303
- Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055
- Bruno, M. A., Majerus, S., Boly, M., Vanhaudenhuyse, A., Schnakers, C., Gosseries, O., et al. (2012). Functional neuroanatomy underlying the clinical subcategorization of minimally conscious state patients. *J. Neurol.* 259, 1087–1098. doi: 10.1007/s00415-011-6303-7
- Bruno, M. A., Vanhaudenhuyse, A., Thibaut, A., Moonen, G., and Laureys, S. (2011). From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: recent advances in our understanding of disorders of consciousness. *J. Neurol.* 258, 1373–1384. doi: 10.1007/s00415-011-6114-x
- Chiu, Y.-C., Jiang, J., and Egner, T. (2017). The caudate nucleus mediates learning of stimulus-control state associations. *J. Neurosci.* 37, 1028–1038. doi: 10.1523/JNEUROSCI.0778-16.2016
- Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., et al. (2006). Language control in the bilingual brain. *Science* 312, 1537–1540. doi: 10.1126/science.1127761
- Di Perri, C., Bahri, M. A., Amico, E., Thibaut, A., Heine, L., Antonopoulos, G., et al. (2016). Neural correlates of consciousness in patients who have emerged from a minimally conscious state: a cross-sectional multimodal imaging study. *Lancet Neurol.* 15, 830–842. doi: 10.1016/S1474-4422(16)00111-3
- Friederici, A. D. (2006). What's in control of language? *Nat. Neurosci.* 9, 991–992. doi: 10.1038/nn0806-991
- Giacino, J. T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., and Katz, D. I. (2002). The minimally conscious state. *Neurology* 58, 349–353. doi: 10.1212/WNL.58.3.349
- Giacino, J. T., Kalmar, K., and Whyte, J. (2004). The JFK coma recovery scale-revised: measurement characteristics and diagnostic utility. *Arch. Phys. Med. Rehabil.* 85, 2020–2029. doi: 10.1016/j.apmr.2004.02.033

- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734
- Phillips, C. L., Bruno, M.-A., Maquet, P., Boly, M., Noirhomme, Q., Schnakers, C., et al. (2011). “Relevance vector machine” consciousness classifier applied to cerebral metabolism of vegetative and locked-in patients. *Neuroimage* 56, 797–808. doi: 10.1016/j.NEUROIMAGE.2010.05.083
- Teki, S., Barnes, G. R., Penny, W. D., Iverson, P., Woodhead, Z. V. J., Griffiths, T. D., et al. (2013). The right hemisphere supports but does not replace left hemisphere auditory function in patients with persisting aphasia. *Brain* 136, 1901–1912. doi: 10.1093/brain/awt087
- Thibaut, A., Bruno, M. A., Chatelle, C., Gosseries, O., Vanhaudenhuyse, A., Demertzi, A., et al. (2012). Metabolic activity in external and internal awareness networks in severely brain-damaged patients. *J. Rehabil. Med.* 44, 487–494. doi: 10.2340/16501977-0940
- Vanhaudenhuyse, A., Demertzi, A., Schabus, M., Noirhomme, Q., Bredart, S., Boly, M., et al. (2011). Two distinct neuronal networks mediate the awareness of environment and of self. *J. Cogn. Neurosci.* 23, 570–578. doi: 10.1162/jocn.2010.21488
- Wannez, S., Gosseries, O., Azzolini, D., Martial, C., Cassol, H., Aubinet, C., et al. (2017a). Prevalence of coma-recovery scale-revised signs of consciousness in patients in minimally conscious state. *Neuropsychol. Rehabil.* 28, 1350–1359. doi: 10.1080/09602011.2017.1310656
- Wannez, S., Heine, L., Thonnard, M., Gosseries, O., and Laureys, S. (2017b). The repetition of behavioral assessments in diagnosis of disorders of consciousness. *Ann. Neurol.* 81, 883–889. doi: 10.1002/ana.24962

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Aubinet, Panda, Larroque, Cassol, Bahri, Carrière, Wannez, Majerus, Laureys and Thibaut. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Attention to Monocular Images Bias Binocular Rivalry

Manuel Moreno-Sánchez<sup>1</sup>, J. Antonio Aznar-Casanova<sup>1\*</sup> and Fernando Valle-Inclán<sup>2</sup>

<sup>1</sup>Department of Cognitive Processes, Universitat de Barcelona, Barcelona, Spain, <sup>2</sup>Department of Psychology, Universidad de La Coruña, La Coruña, Spain

When monocular images cannot be fused, perception alternates between the two (or more) possible images. This phenomenon, binocular rivalry (BR), is driven by the physical properties of the stimuli (size, contrast, spatial frequency, etc.) but it can also be modulated by attention to features of one of the rival stimuli (Chong et al., 2005; Dieter et al., 2016) and by attentional demands independent of the BR assessment (Paffen et al., 2008). Instead of the perceptually demanding tasks previously used to bias BR, we designed a simple counting task. We monocularly presented a number of trials (around 10 min) with a set of symbols and asked participants to count them. We found that after this task, dominance durations decreased for the unattended channel, and did not change for the attended channel. The results parallel those of Paffen et al. (2008) and square nicely with Levelt's second proposition, suggesting that the counting task effectively increased the sensibility of one channel which led to increased strength of the images presented to that channel. Alternatively, the results could be explained assuming that the non-attended channel was inhibited during the counting task, and the inhibition was carried over to the BR task.

## OPEN ACCESS

### Edited by:

Maria V. Sanchez-Vives,  
August Pi i Sunyer Biomedical  
Research Institute (IDIBAPS), Spain

### Reviewed by:

Jitendra Sharma,  
Massachusetts Institute of  
Technology, United States  
Ruben Moreno-Bote,  
Universidad Pompeu Fabra, Spain

### \*Correspondence:

J. Antonio Aznar-Casanova  
jaznar2@ub.edu

**Received:** 28 September 2018

**Accepted:** 27 February 2019

**Published:** 18 March 2019

### Citation:

Moreno-Sánchez M,  
Aznar-Casanova JA and  
Valle-Inclán F (2019) Attention to  
Monocular Images Bias  
Binocular Rivalry.  
*Front. Syst. Neurosci.* 13:12.  
doi: 10.3389/fnsys.2019.00012

**Keywords:** visual awareness, visual attention, binocular rivalry, endogenous attention, attention

## INTRODUCTION

When unfusable images are dichoptically presented, perception alternates between the possible percepts, which are usually two. This phenomenon, binocular rivalry (BR), is essentially stochastic and arises from competition at multiple levels in the visual system (Blake and Logothetis, 2002). The perceptual alternations in BR depend on the physical characteristics of stimuli (size, contrast, motion, etc.; Levelt, 1965), but are also under the influence of attentional processes. Helmholtz (1925) was the first to note that attending to one of the rival stimuli (i.e., counting the lines in one of the images) prolonged dominance durations for that stimulus. More recent research has shown that images immediately presented before the rivalry display tend to be the initial dominant image (Meng and Tong, 2004; Mitchell et al., 2004; Chong and Blake, 2006; Hancock and Andrews, 2007). Also, when observers have to track subtle changes in one of the rival stimuli, dominance durations increase for that image/channel (Chong et al., 2005; Hancock and Andrews, 2007; Dieter et al., 2015, 2016). Finally, when subjects attend to relevant stimuli mixed with irrelevant stimuli in a non-rivalry task, and then the relevant irrelevant stimuli are dichoptically presented, the dominance durations of the previously unattended stimulus decrease but the dominance of the attended stimulus does not change (Paffen et al., 2008).

In the experiments above, the biases in dominance duration were induced by tracking of changes in one of the rivalrous stimuli (Chong et al., 2005; Hancock and Andrews, 2007; Dieter et al., 2016), or after training on a binocular visual discrimination task (Paffen et al., 2008). It is not totally clear if less perceptually demanding tasks, performed outside the BR assessment



task, would produce similar effects. We tested this idea using a simple monocular counting task and assessing BR before and after. The use of simple, easily discriminable and static to-be-attended stimuli provides a stricter test of the effects of endogenous attention on BR.

## EXPERIMENT 1

### Participants

All subjects gave written informed consent in accordance with the Declaration of Helsinki. All the experiments were approved by the local ethics committee (Bio-ethics committee of the University of Barcelona) and conducted in accordance with the Declaration of Helsinki of 1975 (as revised in Fortaleza, Brazil, October 2013). In all the experiments, participants had normal (or corrected to normal) visual acuity (20/20), and stereoacuity (at least 60 s arc, according to TNO test).

Fifty-two students (40 women) between 19 and 26 years (mean = 22.45; SD = 2.63), volunteered for the experiment. Two subjects did not show perceptual changes in some of the BR tasks, and were excluded. Participants were randomly assigned to left-attended ( $N = 16$ ), right-attended ( $N = 19$ ), and control groups ( $N = 15$ ).

### Materials and Methods

#### Stimuli and Apparatus

The stimuli were presented on a 19-inch TFT screen (1,280 × 768 pixels). The rivalrous stimuli were anaglyphs with red and cyan square-wave gratings orthogonally oriented ( $\pm 45^\circ$ ), with a spatial frequency of 0.82 cycle  $\text{deg}^{-1}$  and 70% contrast. They subtended a visual angle of 2.81 deg.

The stimuli to be attended were sets of “Os” and “Xs.” The stimulus presentation and response collection were controlled with a C++ (Open-GL API) program running on a desktop with Windows 7.

### Procedure and Data Analysis

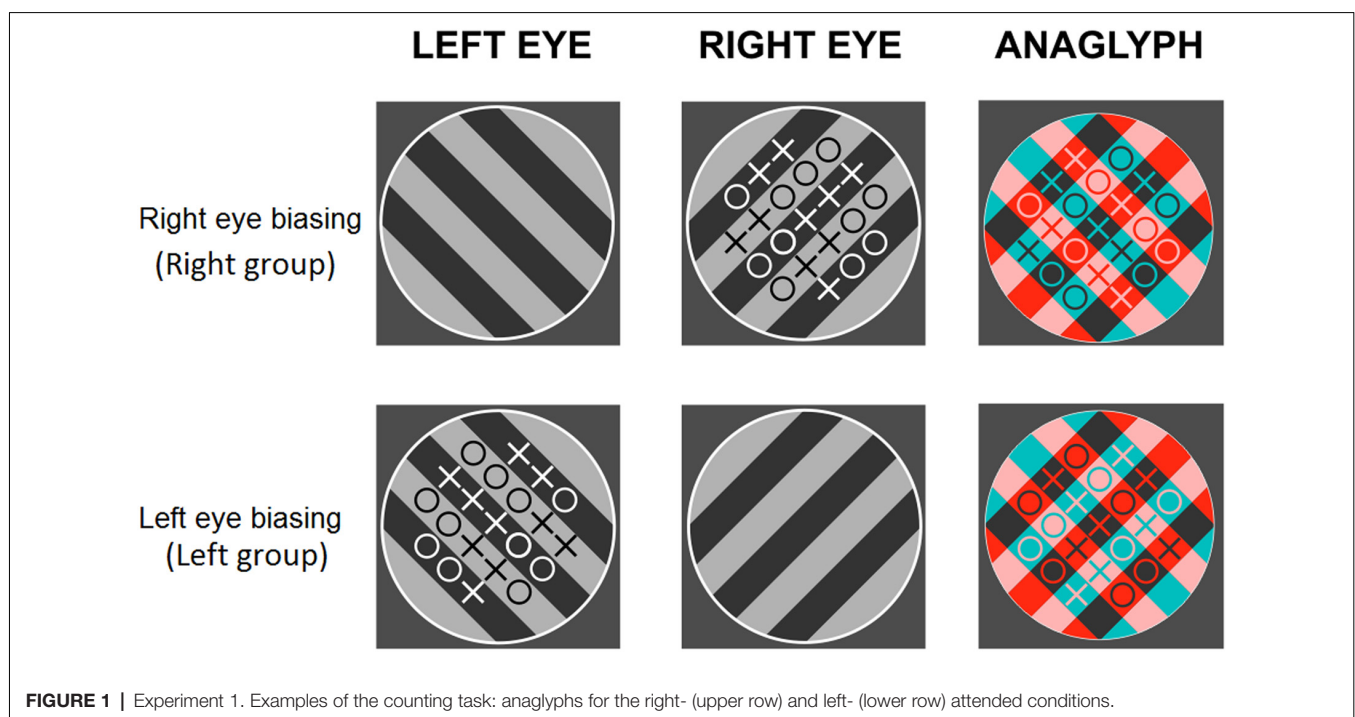
The experimental procedure was the same in all experiments, except the attentional task. There were two BR sessions (each of them comprising four 1-min periods) in which the subjects pressed keys to indicate their perception. The rival display was a red and cyan anaglyph seen with the red filter on the left eye and the cyan filter on the right eye.

In between the two BR sessions, subjects saw the same rival targets with a set of O's and X's (21 in total) overlaid on one of them (see **Figure 1**). The difference between the two items varied randomly from trial to trial. Subjects were instructed to count those elements and indicate which one was more numerous by pressing a key. Following their response, a blank screen was presented for 1 s and the next display appeared. Subjects performed this task for around 10 min. In different groups, the symbols were presented to the left eye, right eye, or alternate between the eyes in different trials (control group). **Figure 2** summarizes the procedure.

### Results










In the counting task, there were 101.37 trials per subject (SD = 6.59) with a duration of 8.19 s/trial, (SD = 1.24). Mean accuracy was 0.96.

The interocular ratio (right eye/left eye) of dominance durations were calculated for the pre and the post BR tasks. **Figure 3** shows that after performing the monocular counting task, the interocular ratio decreased for the left-attended group

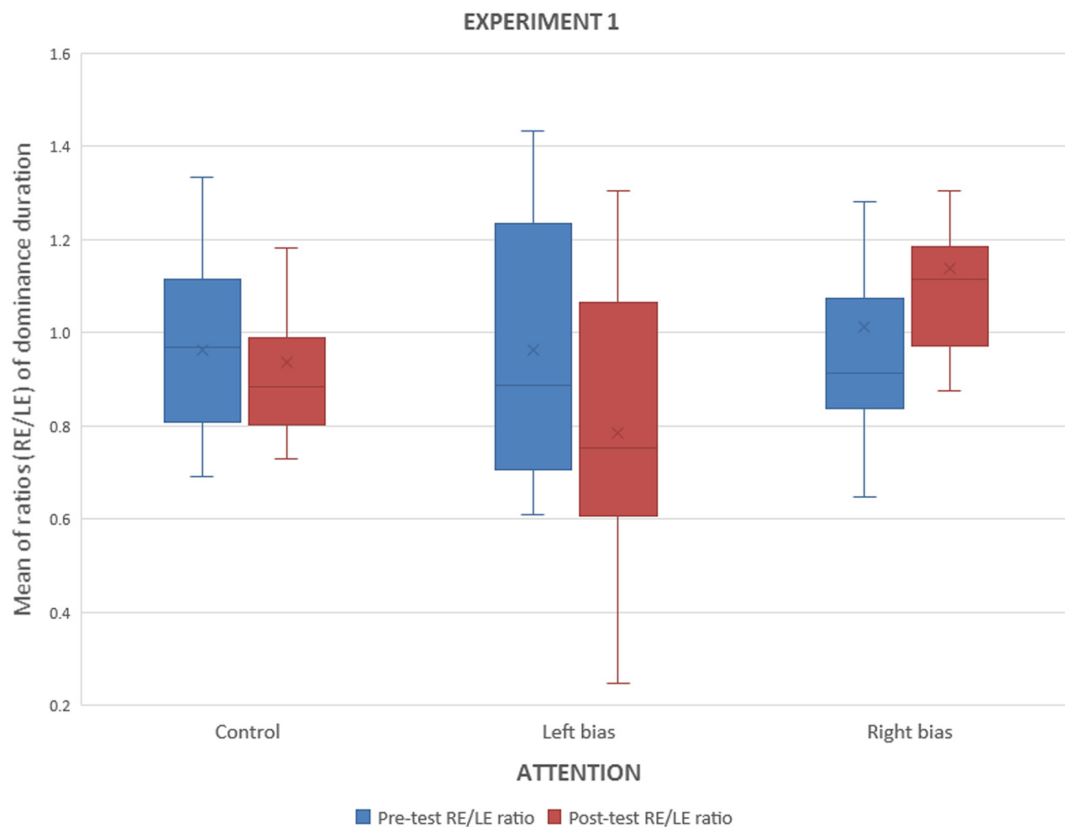


**FIGURE 1 |** Experiment 1. Examples of the counting task: anaglyphs for the right- (upper row) and left- (lower row) attended conditions.



	Binocular Rivalry task (pre) 4x1min	Counting task 15 min	Binocular rivalry task (post) 4x1min
GROUP 1 right eye biased			
GROUP 2 left eye biased			
GROUP 3 control both eyes biased			

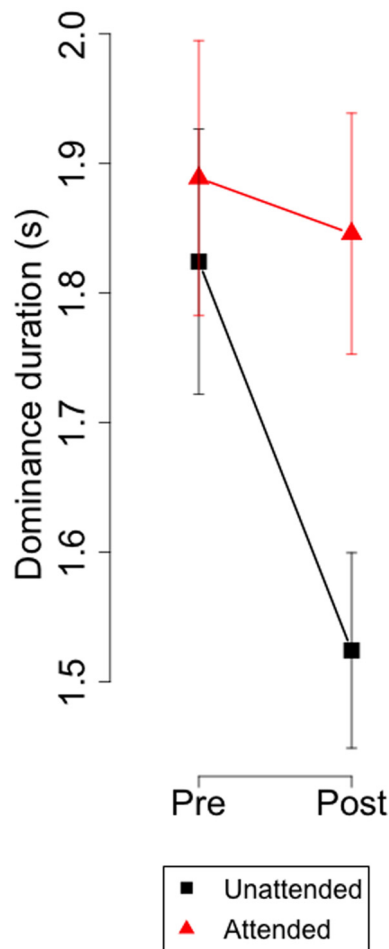
**FIGURE 2** | Experiment 1. Sequence of tasks for each group (row).



**FIGURE 3** | Experiment 1. Interocular ratios (right/left) of median dominance durations before and after the monocular counting task.

and increased for the right-attended. An ANOVA with factors attention (control, left attended, right attended) and time (before, after the BR task) showed an interaction of attention and time for dominance durations ( $F_{(2,49)} = 5.90$ ,  $p < 0.005$ ) and for alternations ( $F_{(2,49)} = 4.32$ ,  $p < 0.02$ ).

The biases shown in **Figure 3** could result from increases in attended eye, from decrements in the unattended eye, or a combination of both. **Figure 4** depicts the dominance durations for the attended and unattended channels during the monocular attentional task. It shows that, after paying attention

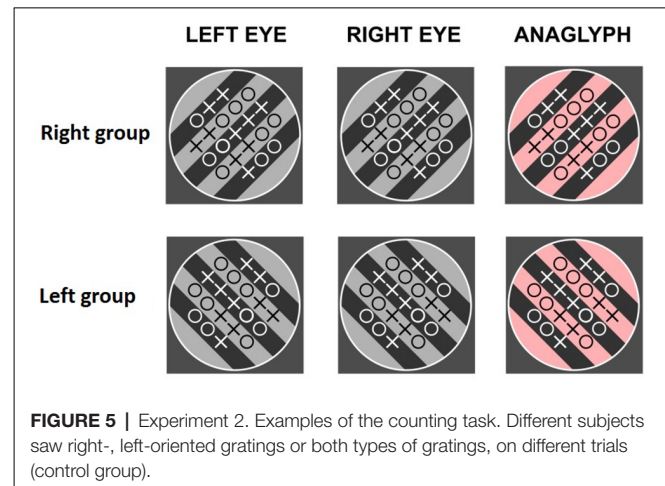


**FIGURE 4 |** Experiment 1. Group means and 95% confidence intervals of dominance durations for the attended and unattended channel before and after the monocular attentional task.

to a monocular set of symbols, dominance duration on the unattended eye decreased but did not change on the attended eye. These observations were confirmed with an ANOVA with factors: attention (attended, not attended) and time (before, after the attentional task) which revealed a significant interaction between attention and time ( $F_{(1,35)} = 9.53, p < 0.004$ ).

## Discussion

The experiment showed that minutes later after paying attention to monocular sets of O's and X's overlaid on a rivalrous display, the dominance durations decreased on the unattended channel and did not change on the attended channel. This result differs from Chong et al. (2005), Hancock and Andrews (2007), and Dieter et al. (2016) who found that the dominance duration of the attended image increased while it did not change or decreased for the unattended stimulus. The main difference with our procedure is that these experiments demanded attention to some feature of the rivalrous display that was under continuous and subtle change. It seems plausible that perceptual load was much higher in those experiments than in ours, and



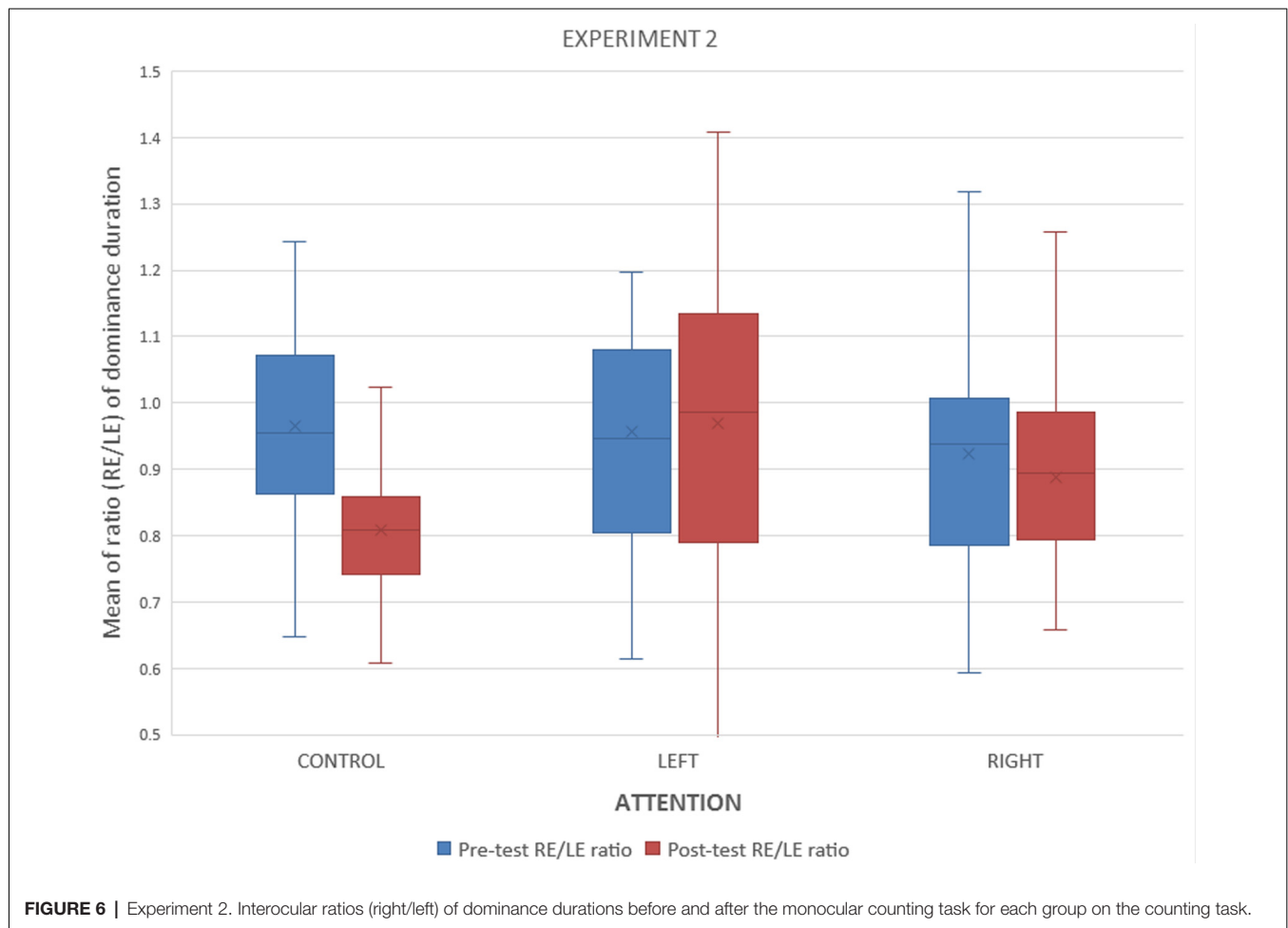
**FIGURE 5 |** Experiment 2. Examples of the counting task. Different subjects saw right-, left-oriented gratings or both types of gratings, on different trials (control group).

if this were the critical difference, the results suggest that mild perceptual loads inhibit the unattended channel, while higher loads enhance neural responses in the attended channel (Zhang et al., 2012).

Our findings dovetail with those of Paffen et al. (2008) who trained subjects in a binocular motion discrimination task with relevant and irrelevant directions. The relevant and irrelevant stimuli (random dots) were used on a BR task before and after the perpetual discrimination task. They found the same pattern described here: previously not attended stimuli became less dominant, while the attended stimuli were unaffected. A straight forward interpretation of these findings would be that the irrelevant stimuli were suppressed in the presence of relevant stimuli (Paffen et al., 2008). On our experiment, the same logic would state that, in order to perform the counting task, the unattended channel and/or the background image, were inhibited.

Alternatively, it is tentative to suggest that monocular attention increased the effective contrast for the images presented to that channel as suggested by Chong and Blake (2006). This interpretation squares nicely with Levelt's proposition 2 (Levelt, 1965) which predicts that increasing the contrast of a stimulus will decrease the dominance of the other stimulus and will not affect the dominance of the stimulus with increased contrast. The generality of Levelt's propositions has been limited by recent research (see Brascamp et al., 2015, for a review) and, in particular proposition two does not seem hold when contrast of one stimulus is fixed at a low level and the other stimulus is varied over higher contrast levels (Brascamp et al., 2006; Moreno-Bote et al., 2010). These findings led to reformulate Levelt's proposition as follows: changes in contrast of one eye affect the mean dominance duration of the highest contrast eye. According to this modified second proposition, we should expect longer dominances on the attended than on the unattended. However, the violations of Levelt's second proposition have been generally found with smaller stimuli than the confirmations (Kang, 2009), and the size of our stimuli ( $2.8^\circ$ ) were in the range where confirmations of Levelt's 2nd propositions have been reported.

There is no way we can decide whether monocular attention enhanced neural responses and boosted the contrast in the



attended channel, or simply inhibited the unattended channel. However, there is a previous question we can explore. Were the effects due to functional changes in the monocular channels, independently of the images presented, or were they due to stimulus adaptation. To test this idea, the counting task in Experiment 2 presented the same grating and sets of O's and X's to both eyes, and Experiment 3 used different orientations in the attentional and in the BR tasks.

## EXPERIMENT 2

The experiment was designed to test whether indirect attention to an image, in a task not involving rivalry nor monocular selection, could replicate Experiment 1. It has been found that the initial dominance phase is affected by images shortly presented before the rivalry display (Mitchell et al., 2004; Chong and Blake, 2006; Hancock and Andrews, 2007), but we are not aware of studies showing similar effects on dominance durations.

## Method

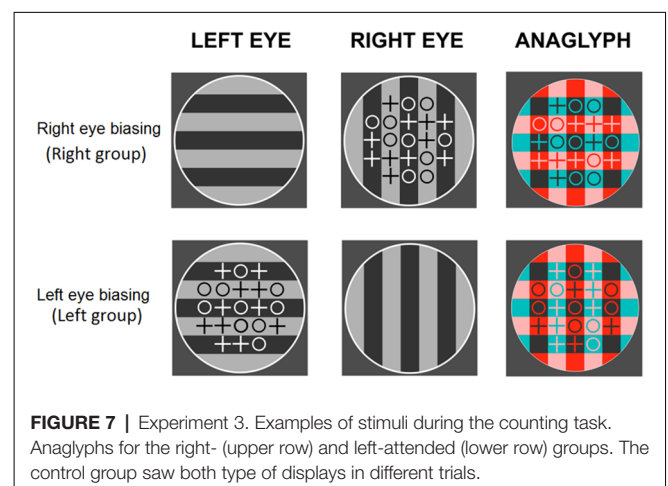
### Participants

Forty-five volunteers, students of Psychology (38 women) between 19 and 26 years (mean = 22.36 years; SD = 2.51 years),

were randomly assigned to one of these groups: left-attended ( $N = 16$ ), right-attended ( $N = 16$ ), and control ( $N = 13$ ).

### Stimuli, Procedure, and Data Analysis

The experiment was identical to Experiment 1, except that the counting task was binocular and the gratings had the



same orientation on both eyes. Different groups of subjects saw left-oriented gratings, right-oriented gratings, or both on different trials (control group). **Figure 5** illustrates these conditions. The data treatment and statistical analyzes were identical to those in Experiment 1.

## Results and Discussion

In the counting task, there were 129.56 trials per subject ( $SD = 6.35$ ) with a duration of 6.24 s per trial, ( $SD = 1.16$ ). Mean accuracy was 0.94. **Figure 6** summarizes the results obtained after calculating the interocular ratios (see Experiment 1) and indicates that the attentional task had no influence. The ANOVA confirmed this observation for the dominance durations ( $F_{(2,42)} = 2.351$ ;  $p < 0.1$ ).

Although it is difficult to argue based on null effects, these results suggest that indirect attention to an image minutes before the BR task, does not affect dominance durations neither alternations. It seems reasonable to assume that the presence of BR and the attention to a monocular image, or channel, are requisites to induce the biases in found in Experiment 1.

## EXPERIMENT 3

Knowing that mere exposure to a stimulus does not affect BR, we tried to dissociate the effects of the background image (the grating) from those of the attention within a monocular

channel. Remember that in Experiment 1 the gratings were the same during the counting and BR tasks. Changing the rivalrous displays during the counting and the BR tasks should not affect the findings of Experiment 1 if the critical factor is the monocular presentation of the to-be-attended stimuli. In addition, we tried to control the possible cross-talks between the color filters using two groups of subjects, each with a different location of the cyan and red filters.

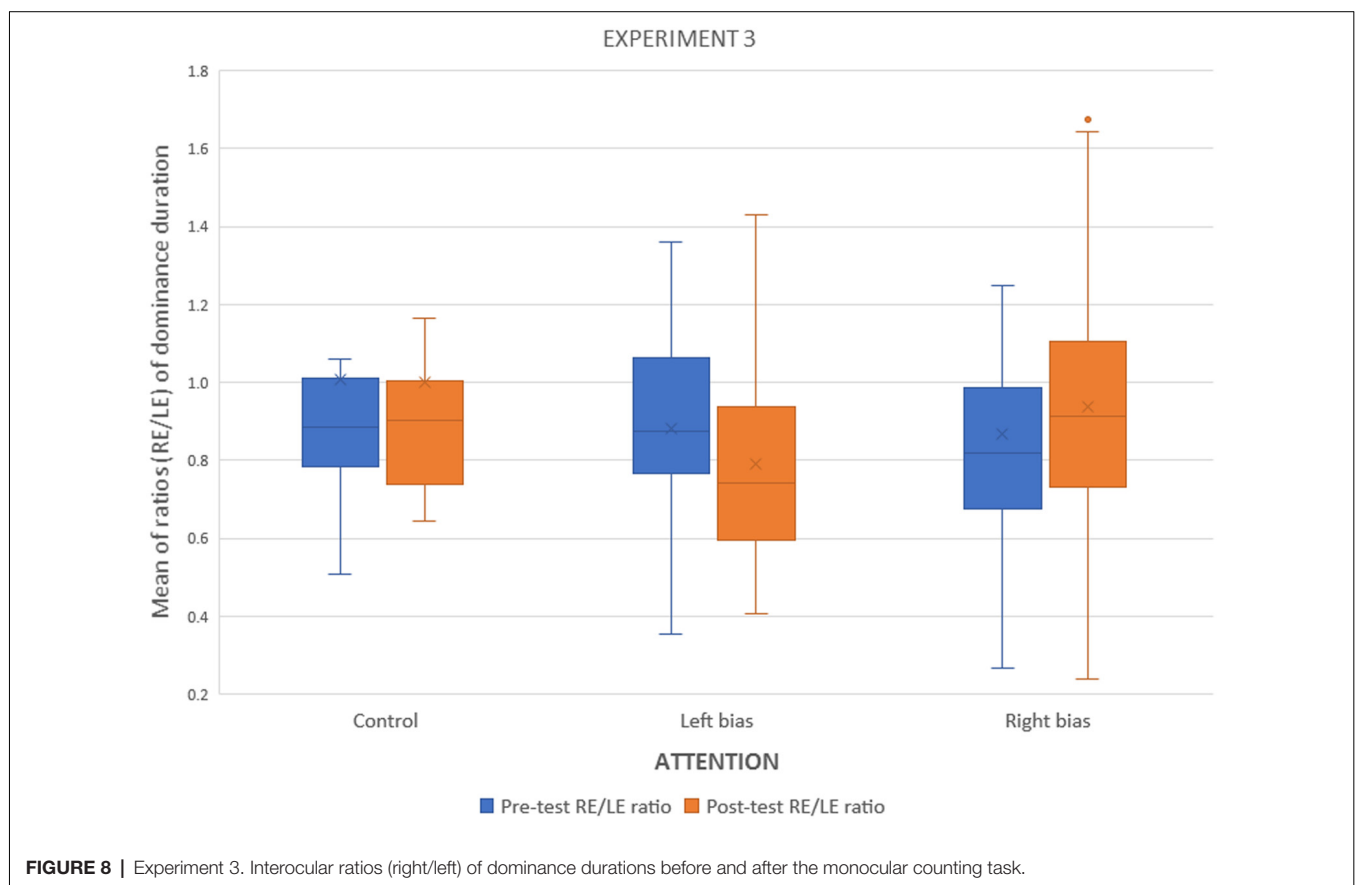
## Method

### Participants

Eighty-four volunteers (70 women), between 19 and 26 years (mean = 22.36 years;  $SD = 2.51$  years) participated in the experiment. Forty-four of them performed the experiment with the red and cyan filters over the left and right eyes, respectively. For the other 40 subjects, the location of the filtered was reversed. Participants were randomly assigned to left attend, right attend or attend both, on different trials, conditions during the counting task.

### Stimuli, Procedure, and Data Analysis

The only differences with Experiment 1 were the orientation of gratings (0 and 90°) during the counting task (see **Figure 7**) and the addition of a second group with the locations of the color filters reversed. Since preliminary analyzes showed no differences related to the location of the filters, the two groups were collapsed.



## Results and Discussion

In the counting task, there were 105.57 trials per subject ( $SD = 7.08$ ) with a duration of 7.79 s/trial, ( $SD = 1.10$ ). Mean accuracy was 0.94.

The interocular ratios (right eye/left eye) for dominance and alternations were analyzed as in Experiment 1. The only significant effect for the alternation rate was an increase in the number of alternations in the second BR tasks ( $F_{(1,83)} = 43.59$ ,  $p < 0.0001$ ). **Figure 8** depicts dominance results, which show the pattern seen in Experiment 1, confirmed by an interaction of attention condition  $\times$  time  $\times$  eye ( $F_{(2,83)} = 5.27$ ,  $p < 0.007$ ).

Again, to elucidate whether these biases were caused by increments in the attended channel, decrement in the unattended, or a combination of both, a second analysis including only the left and right attended groups, resulted in a strong interaction between attention condition and time ( $F_{(1,57)} = 10.83$ ,  $p < 0.002$ ; **Figure 9**).

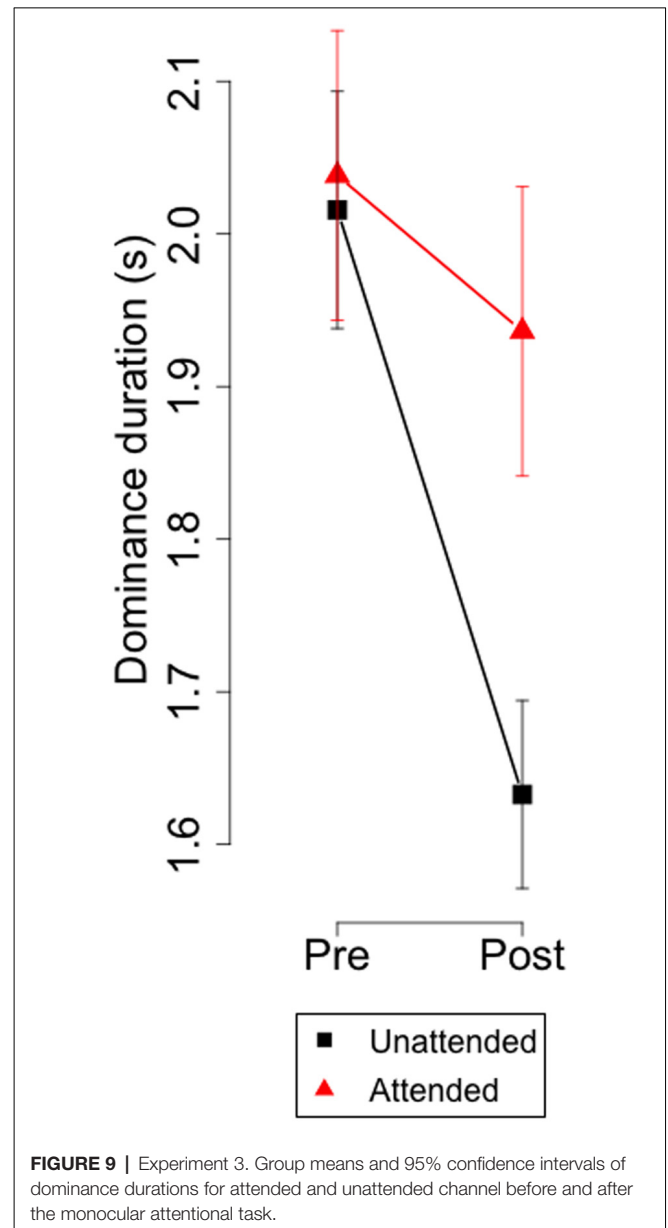
The results for dominance duration replicate the main finding of Experiment 1, namely, the decreased dominance durations for the unattended monocular channel. Since the gratings were different in the attentional and the BR tasks, the conclusion is that attention to a monocular channel is the critical variable. However, the size of the effect in Experiment 3 was slightly smaller than in Experiment 1, making it reasonable to think that adaptation to the grating played some role in Experiment 1. We analyzed both experiments together looking for some effect related to the experiment, and there were none.

The main difference with Experiment 1 was the absence of effects related to attention on the alternation rate, which lead us to consider the previously reported finding as unreliable, since the sample size was double in this experiment.

## GENERAL DISCUSSION

After some minutes seeing short-term (8 s, as an average) rivalry displays with monocular elements to be attended, dominance was reduced for the non-attended monocular channel, without changes in the attended channel. Although the task required attention, it is possible that the results just reflect the differences between the two images. It is known that stimuli with higher density contour tend to dominate (Levelt, 1965) and subjects would have learned to use one of the monocular channels when confronted with rival displays. Within this interpretation, the experiments are closer to a low-level perceptual learning than to an endogenous attentional effect. However, were this the case, the attended channel should show longer dominances after the counting task, in overt contradiction with the results.

Discarding the low-level perceptual learning leads to interpretations related to endogenous attention. A plausible interpretation would be that the non-attended channel was inhibited during the counting task, and this effect was carried over to the BR task (as Paffen et al., 2008; Vergeer et al., 2016, have proposed in different experimental setups). Alternatively, the counting task could have enhanced neural responses in one channel and as a consequence, the stimuli presented to that channel would be have higher contrast than their



contralateral partners. Contrary to common sense, increasing the contrast of one stimulus does not affect dominance for that stimulus, instead, it reduces dominance of the opposite stimulus. This is the second proposition of Levelt (1965) which is valid for relatively large stimuli (Kang, 2009) and for smaller stimuli within a range of contrasts (Moreno-Bote et al., 2010; Brascamp et al., 2015).

In summary, we found that monocular attention prior to a BR task, reduces the dominance in the non-attended channel, without changes in the attended channel. It should be noted that the attentional effects here reported last for minutes (as in Paffen et al., 2008). These results are consistent with either or both interpretations. First, the neural response in the attended channel was enhanced, which is equivalent to increase the contrast of the stimuli presented to that channel. Second,

monocular attention suppressed the neural responses on the unattended channel.

## AUTHOR CONTRIBUTIONS

MM-S, JA-C and FV-I designed the experiment, conducted the research, analyzed the data and wrote the manuscript. All authors contributed to the manuscript.

## REFERENCES

- Blake, R., and Logothetis, N. K. (2002). Visual competition. *Nat. Rev. Neurosci.* 3, 13–23. doi: 10.1038/nrn701
- Brascamp, J. W., Klink, P. C., and Levelt, W. J. M. (2015). The 'laws' of binocular rivalry: 50 years of Levelt's propositions. *Vision Res.* 109, 20–37. doi: 10.1016/j.visres.2015.02.019
- Brascamp, J. W., van Ee, R., Noest, A. J., Jacobs, R., and van den Berg, A. V. (2006). The time course of binocular rivalry reveals a fundamental role of noise. *J. Vis.* 6, 1244–1256. doi: 10.1167/6.11.8
- Chong, S. C., and Blake, R. (2006). Exogenous attention and endogenous attention influence initial dominance in binocular rivalry. *Vision Res.* 46, 1794–1803. doi: 10.1016/j.visres.2005.10.031
- Chong, S. C., Tadin, D., and Blake, R. (2005). Endogenous attention prolongs dominance durations in binocular rivalry. *J. Vis.* 5, 1004–1012. doi: 10.1167/5.11.6
- Dieter, K. C., Melnick, M. D., and Tadin, D. (2015). When can attention influence binocular rivalry? *Atten. Percept. Psychophys.* 77, 1908–1918. doi: 10.3758/s13414-015-0905-6
- Dieter, K. C., Melnick, M. D., and Tadin, D. (2016). Perceptual training profoundly alters binocular rivalry through both sensory and attentional enhancements. *Proc. Natl. Acad. Sci. U S A* 113, 12874–12879. doi: 10.1073/pnas.1602722113
- Hancock, S., and Andrews, T. J. (2007). The role of voluntary and involuntary attention in selecting perceptual dominance during binocular rivalry. *Perception* 36, 288–298. doi: 10.1068/p5494
- Kang, M. S. (2009). Size matters: a study of binocular rivalry dynamics. *J. Vis.* 9:17. doi: 10.1167/9.1.17
- Levelt, W. J. M. (1965). *On Binocular Rivalry*. Soesterberg: Institute for Perception RVO-TNO.
- Meng, M., and Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *J. Vis.* 4, 539–551. doi: 10.1167/4.7.2
- Mitchell, J. F., Stoner, G. R., and Reynolds, J. H. (2004). Object-based attention determines dominance in binocular rivalry. *Nature* 429, 410–413. doi: 10.1038/nature02584
- Moreno-Bote, R., Shpiro, A., Rinzel, J., and Rubin, N. (2010). Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *J. Vis.* 10:1. doi: 10.1167/10.11.1
- Paffen, C. L. E., Verstraten, F. A. J., and Vidnyánszky, Z. (2008). Attention-based perceptual learning increases binocular rivalry suppression of irrelevant visual features. *J. Vis.* 8:25. doi: 10.1167/8.4.25
- Vergeer, M., Wagemans, J., and van Ee, R. (2016). Training of binocular rivalry suppression suggests stimulus-specific plasticity in monocular and binocular visual areas. *Sci. Rep.* 6:25753. doi: 10.1038/srep25753
- Helmholtz, H. (1925). *Helmholtz's Treatise on Physiological Optics*, Vol. 3. Trans. J. P. C. Southall. (New York, NY: Optical Society of America).
- Zhang, P., Jiang, Y., and He, S. (2012). Voluntary attention modulates processing of eye-specific visual information. *Psychol. Sci.* 23, 254–260. doi: 10.1177/0956797611424289

## FUNDING

This work was funded by grants 2017SGR0048 from the Catalan government (Generalitat de Catalunya) and PSI2017-83493-R from the Ministry of Economy, Industry and Competitiveness, Government of Spain (Ministerio de Economía, Industria y Competitividad, Gobierno de España; grant no. AEI/FEDER, UE).

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Moreno-Sánchez, Aznar-Casanova and Valle-Inclán. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Outer Brain Oscillations in Down Syndrome

**Marcel Ruiz-Mejias\***

*Systems Neuroscience, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain*

The present article reviews the relationship between sleep and oscillatory activity in Down Syndrome (DS), as well as the featuring emergent rhythmic activity across different brain states. A comprehensive discussion of the data from electroencephalographic studies in DS humans and transgenic/trisomic mouse models is provided, as well as data from signals collected from local field potentials (LFP) and intracellular recordings in DS mouse models. The first sections focus specially on the alpha phenotype consistently observed in DS subjects, as well as its description in DS childhood and aging. Subsequently, a review of the data reported in DS mouse models is presented with the aim to deepen on the mechanisms underlying altered rhythmic patterns. Further sections situate the state-of-the-art of the field, with a discussion on the possible circuit alterations that may underlie impaired alpha and gamma oscillatory activity. A further aim is to highlight the importance of studying network oscillatory activity in mouse models to infer alterations in the underlying circuits related to cognition, such as in intellectual disability. In this direction, a view of alpha and gamma rhythms generated by the cerebral cortex as a tool for evaluating an unbalance between excitation and inhibition in DS is claimed, which points out toward an over-inhibited network. A final aim is to situate oscillatory activity as a key phenomenon that may be used as a biomarker for monitoring as well the effect of novel therapeutic strategies.

**Keywords:** down syndrome, oscillation, sleep, alpha, gamma, inhibition, mouse model, EEG

## OPEN ACCESS

### Edited by:

Igor Timofeev,  
Laval University, Canada

### Reviewed by:

Fabio Placidi,  
Policlinico Tor Vergata, Italy  
Marco Cambiaghi,  
Columbia University, United States

### \*Correspondence:

Marcel Ruiz-Mejias  
marcel.ruiz.mejias@gmail.com

**Received:** 07 January 2019

**Accepted:** 12 April 2019

**Published:** 07 May 2019

### Citation:

Ruiz-Mejias M (2019) Outer Brain Oscillations in Down Syndrome. *Front. Syst. Neurosci.* 13:17. doi: 10.3389/fnsys.2019.00017

## INTRODUCTION

Unraveling the alterations of the underlying circuits in neurobiological disorders is with no doubt one of the most interesting applications of studying brain oscillatory activity. Sleep and brain oscillatory activity in Down Syndrome (DS) is a particularly explored field of study, with early findings in sleep alterations in DS patients. Nowadays it is well accepted the relationship of sleep with cognitive function, due to its implication in processes such as memory consolidation. As well, rhythmic activity of neural networks such as alpha or gamma waves has been related widely to cognition in the last decades. Both sleep and oscillatory activity are tightly related, as the major oscillatory feature of sleep -the slow oscillation ( $\sim 1$  Hz)- shows an exquisite interplay of different forms of rhythmic neuronal activity. Sleep and oscillatory activity in DS has been studied from the mid 20th century, but literature has appeared sporadically in the last decades. From the early descriptive studies, few works reported the differential features of sleep in DS individuals, showing a wide diversity but consistency of results. Some studies addressed network oscillatory activity emerging from external brain circuits of DS individuals, with results related to different frequency bands of the electroencephalogram (EEG) signal. Even fewer studies have been published in models, but interestingly, some of them focused in the network mechanisms that may explain intellectual disability and cognitive deficits observed in DS.

In the last 50 years, sleep and oscillatory activity in DS has been a focus of interest due to the evident alterations of sleep patterns in individuals with this chromosomal alteration, consisting on a partial or total trisomy of chromosome 21. The description of oscillatory activity during sleep and its functional mechanisms (Steriade et al., 1991; Contreras et al., 1996; McCormick and Bal, 1997; Chauvette et al., 2011) allowed the exploration of the emerging cortical rhythmic patterns associated with sleep stages, such as slow wave sleep. This type of activity has been proposed to be involved in main physiological and cognitive roles, such as synaptic homeostasis (Cirelli and Tononi, 2015), memory consolidation (Bigum et al., 1970; Buzsáki, 1998; Miyamoto et al., 2016), and even has been described as the default mode activity in the neocortex (Sanchez-Vives and Mattia, 2014). It is well studied that the slow oscillation orchestrates several types of rhythmic activity both during sleep and, interestingly, in some regimes of anesthesia (Steriade, 2006; Chauvette et al., 2011; Ruiz-Mejias et al., 2011), a fact that helped to study oscillatory patterns in animal models, thus allowing the possibility to study its alterations in DS models such as transgenic and trisomic mice.

Oscillatory activity is typically studied in the surface of the head by means of EEG recordings in humans. The first EEG recordings in DS individuals were performed by Gunnarson (1945) in the 40's, concretely in DS children, by Vergani and Aldeghi (1958); Beley et al. (1959); Gibbs et al. (1965) in the 50's, and later by Hirai and Izawa (1964) in the 60's. Altered sleep patterns in DS humans were firstly observed by Petre-Quadens and Jouvet (1967); Castaldo (1969); Feinberg et al. (1969) in the 60's by means of polysomnographic recordings. Another posterior study in the 70's by Clausen et al. (1977) combined EEG recording with power spectrum analysis in different frequency bands to study oscillatory activity in DS subjects in different sleep stages. These studies showed consistently increased waking periods as well as reduced REM sleep in DS individuals. In later studies EEG signals were combined with localization of current sources, measures of cognitive performance, among others. Initial studies described as well how DS brain responded to external stimuli in form of evoked potentials in the 60's and early 70's (Barnet and Lodge, 1967; Bigum et al., 1970), while work in the 60's with EEG recordings described as well the increased occurrence of epilepsy in DS (Borselli and Sferlazzo, 1963; Seppäläinen and Kivalo, 1967). Sleep patterns have found to be altered in DS children, with increased wakefulness, increased percentage of stage 1 of NREM sleep, and decreased REM sleep (Miano et al., 2008). The works in this field mostly characterized EEG patterns focusing in alterations in the alpha rhythms. The following are studies which relay on the interictal altered EEG sleep and oscillatory and patterns in DS individuals. Later, we review the work performed in mouse models of DS to foster explanations for the impairment observed in humans.

## OSCILLATIONS IN DS ADULTS

The studies reviewed here commonly report alterations in the emergent oscillatory activity related to alpha oscillations. A first evidence of this in sleeping/awake humans with DS performed

power spectrum analysis –a transformation of the amplitude of the signal in the frequency domain– in EEG recordings to evaluate oscillatory activity across waking and sleep stages (Clausen et al., 1977). Clausen et al. (1977) reported increased power of frequencies between 1 and 20 Hz in all stages of sleep with exception to a narrow band of 8–13 Hz corresponding to alpha waves. The differences in power were more prominent in stages 1, REM sleep and the waking state. In addition, in the waking state, a peak in the alpha band in controls was not present in DS humans.

The alpha alteration background seems to correlate with cognitive performance and age. Reporting evidence on this statement, a study performed EEGs, psychometric testing, quantitative computed tomography, and positron emission tomography with fludeoxyglucose –that provides a measure of brain metabolism– in 19 young DS subjects, 9 older DS subjects and 13 healthy control subjects. When comparing the age-matched patients with normal alpha background subjects, older DS patients with decreased alpha background had dementia, fewer visuospatial skills, decreased attention span, larger third ventricles, as well as a global decrease in cerebral glucose utilization with parietal hypometabolism. Despite this, in the young patients with DS the EEG background did not correlate with psychometric or metabolic findings, but, however, in the neuroanatomical analysis the third ventricles were significantly larger in those with abnormal EEG background (Devinsky et al., 1990). Alpha alterations have been also measured as ratios between eyes-closed and eyes-open brain states. A study showed that alpha band power ratio between eyes-closed and eyes-open recording in the occipital area of waking EEG of DS humans was substantially increased compared to controls, thus showing the power of alpha waves was significantly lower in DS patients during the open-eyes state, with the degree of alterations correlating with neuropsychological test scores of DS subjects (Partanen et al., 1996). In DS adolescents, a lower power in the alpha waves was also reported, as well as in the beta and lower gamma EEG bands. As a novelty, DTF (directionality) values globally prevailed from right to left occipital areas in normal young subjects and in the opposite direction in DS patients (Babiloni et al., 2009).

In an effort to provide an improved spatial resolution of the previously observed alterations, further work aimed to identify the current sources of the alterations in wave activity. A second study in adults, Babiloni et al., 2009 combined EEG recording with LORETA technique (Low Resolution Electromagnetic Tomography), which allowed the correlation of EEG signals to the Current Source Density (CSD) representations in three dimensions. Central, parietal, occipital, and temporal cortical sources of resting alpha and beta rhythms were reported to be lower in amplitude in DS, whereas the opposite was true for occipital delta cortical sources. The authors argued that this result in alpha cortical source is consistent to what is typically found in Alzheimer's disease, which is often associated to older DS humans (Babiloni et al., 2010). In this direction, further work combined also EEG with e/sLORETA technique (exact/standard LORETA) and provided correlation with cognitive performance. In similar frequency bands than the studies of Babiloni et al. (2010), they reported increased CSD in theta, alpha1 and beta1 classical

**TABLE 1** | Summary of findings of oscillatory activity in down syndrome (DS) adults.

First author of the work (year of publication)	DS subjects (recording technique)	Main findings in DS
Clausen et al., 1977	DS adolescents and young adults (EEG-polysomnography).	Increased oscillatory activity from 1 to 20 Hz during waking and sleep states with exception to alpha oscillations in parietal areas. Peak in alpha oscillations in waking state was not present.
Devinsky et al., 1990	Waking young and older DS adults (EEG).	Dementia, fewer visuospatial skills, decreased attention span, larger third ventricles, and a global decrease in cerebral glucose utilization with parietal hypometabolism in older DS adults with decreased alpha oscillations. Larger third ventricles were in young DS adults with decreased alpha oscillations.
Partanen et al., 1996	Waking DS adults with eyes closed and eyes open (EEG).	Increased alpha band power ratio between eyes-closed and eyes-open in occipital area in DS (decreased alpha oscillations in the eyes-open state). Correlation between the degree of changes and neuropsychological test scores.
Babiloni et al., 2009	Resting DS adolescents with eyes closed (EEG).	Reduced alpha, beta and gamma oscillations in occipital area. Directionality in occipital area prevailed left to right in DS (opposite in controls).
Babiloni et al., 2010	Resting DS adults with eyes closed (EEG-LORETA).	Lower current sources of alpha and beta current sources in central, parietal, occipital, and temporal cortical areas. Increased delta cortical sources in occipital area.
Velikova et al., 2011	Waking DS adults (EEG-e/s LORETA).	Increased current sources in theta, alpha1 and beta1 absolute bands (averaged electrodes). Reduced current sources in relative alpha2 band that positively correlated with cognitive tests (averaged electrodes). Increased current sources in alpha peak frequency-adjusted bands (alpha1, alpha2 and theta) in occipital areas. Negative correlation between cognitive performance and theta/alpha current source in right frontal lobe and right posterior cingulate cortex.
Virji-Babul et al., 2008, 2010	DS adults (MEG).	Reduced mu (alpha) rhythms in sensorimotor areas. Disruption of the action observation cortical network.

bands and in individual alpha peak frequency-adjusted bands, while relative alpha2 was decreased. A negative correlation between cognitive performance and theta/alpha CSD in the right frontal lobe and right posterior cingulate cortex was also found (Velikova et al., 2011).

Despite EEG studies revealed valuable information, spatial and tissue filtering limitations of the technique motivated the exploration of magnetoencephalographic (MEG) signals in DS adults, which consistently found a decrease in alpha oscillatory activity. Data shows an attenuation of the “mu” rhythm in DS subjects –which refers to alpha oscillations related to sensorimotor areas– with a distinct laterality in the pattern of mu suppression, as well as distinct cortical pattern activation in the action observation network and asynchronous firing of activated areas. These results supported a dysfunction of the mirror system, where actions that are performed by others are recognized by activation of the same neural substrates that are involved in the actual execution of the observed action (Virji-Babul et al., 2008, 2010).

In summary, while waking EEG general oscillatory activity is increased in adolescent and young DS adults in parietal areas and beta and gamma is reduced in occipital areas, the studies show consistent absence of peaks and reductions in alpha rhythms in the neocortex of DS subjects. Sleeping DS subjects also show increased oscillatory activity in a wide range of frequencies across sleep stages in DS individuals, again with the exception of alpha oscillations. Despite these results, is worth to mention the difficulty of extracting high frequency components such as gamma oscillations from the EEG signals due to tissue filtering between neocortex and the recording electrodes, and the possible limitations in sampling frequency of acquiring signals from multiple channels, as well as that gamma frequencies are especially sensitive to normalization and noise artifacts. **Table 1** summarizes the findings exposed in this section.

## OSCILLATIONS DURING DEVELOPMENT AND AGING IN DS

One of the issues in the study of oscillatory activity in DS humans is its different consolidation during development, which found as well consistent alterations in alpha waves. An EEG study by Schmid et al. (1985) in the first years of life of DS children, they reported an increase in absolute power in theta oscillations in the fronto-central area, whereas alpha absolute power was slightly increased or even decreased. They found as well that the most prominent differences were when calculating the relative power of alpha oscillations, which was reduced starting at the age of 6 months, and that the reduction was more prominent with growth (Schmid et al., 1985). DS children have been reported as well to have some degree of lag of maturation of EEG-sleep patterns. In this direction, Hamaguchi et al., 1989 reported that, in addition to a low incidence of quiet sleep (non-REM) in the first 3 months, DS children had less spindles and later disappearance of frontal sharp waves, which are common during the first stages of development (Hamaguchi et al., 1989). When looking at the activity between area pairs in subjects ranging from 9 to 26 years old, it seems that coherence strongly and consistently discriminates between DS and control groups (McAlaster, 1992). A study in DS children provided evidence of decreased alpha waves during REM sleep in the occipital area, supporting differences during development in oscillatory activity of DS individuals, while beta, theta, and delta bands did not differ significantly between the groups (Śmigielska-Kuzia et al., 2005).

Parallel work in DS individuals explored the impairment of oscillatory activity in aging. When comparing DS individuals with non-DS with mental retardation and control subjects, occipital alpha-peak frequency was decreased with age in DS individuals. Furthermore, in older subjects this slowing was accompanied with an increase of activity in 6–8 Hz –theta

**TABLE 2 |** Summary of findings oscillatory activity in DS during development and aging in DS.

First author of the work (year of publication)	DS subjects (recording technique)	Main findings
Schmid et al., 1985	DS children (EEG).	Increase in absolute power of theta oscillations in fronto-central area. Slight increase or even decrease in absolute power of alpha oscillations. Decrease in relative power of alpha oscillations.
Hamaguchi et al., 1989	DS children (EEG-polysomnography).	Lag of maturation of EEG-sleep patterns. Less spindles and later disappearance of frontal sharp waves.
McAlaster, 1992	9 month to 26 year-old DS humans (EEG).	Coherence between pairs of areas strongly and consistently discriminated between DS and control groups.
Śmigielska-Kuzia et al., 2005	Sleeping DS children (EEG-polysomnography).	Decreased alpha waves during REM sleep in occipital area.
Ono et al., 1992; Ono, 1993	Waking DS adults across aging (EEG).	Occipital alpha-peak waves decreased in frequency with age in DS individuals. In older subjects slowing was accompanied with increase of activity in 6–8 Hz, but didn't correlate with cognitive performance.
Soininen et al., 1993	Waking DS adults (EEG).	Slowing in EEG oscillatory (slow, delta, theta) activity in temporal area that correlates with cognitive performance and an age-related decline of cognitive functions in DS humans.
Murata et al., 1994	Waking DS adults (EEG).	Slowing in mean EEG oscillatory activity with aging.
Visser et al., 1996	Waking DS adults, follow-up study (EEG).	Slowing in occipital alpha oscillations oscillatory activity related to cognitive decline with aging and to incidence of Alzheimer's disease.
Katada et al., 2000	Large population of waking DS adults including a follow-up study (EEG).	Earlier steep slowing of EEG alpha-peak in DS in frontal to occipital areas.
Salem et al., 2015	DS-AD adults (EEG).	Decrease in theta oscillations in several areas that negative correlate with cognitive performance that can detect cognitive deterioration.

oscillations, notice that 8 Hz is the limit between theta and alpha oscillations-, but didn't correlate with cognitive performance in the daily life of individuals (Ono et al., 1992; Ono, 1993). Further evidence on this also supported a slowing in EEG oscillatory activity in the temporal area in DS aging adults aged 20 to 60. Concretely, they work showed that DS and Alzheimer's disease (AD) patients had more prominent slow, delta and theta activity than control subjects. However, in this case those findings significantly correlated with and an age-related decline of cognitive functions in DS humans (Soininen et al., 1993). A slowing in frequency of oscillatory activity in EEG recordings was also reported in DS aging subjects, where the authors found significant differences between mean frequency of EEG power in delta, theta, alpha, and beta bands was 9.37 at 20's in DS individuals, while in controls differences were non-significant between individuals in their 20's and 60's (Murata et al., 1994). The slowing in alpha rhythmic activity may be related also to the presence of Alzheimer-type dementia in DS individuals, as a follow-up study describes. The authors showed that the slowing of occipital alpha oscillations was related to a cognitive decline in a set of patients that were diagnosed post-mortem as presenting a severe form of AD (Visser et al., 1996). Further evidence supported the slowing in the dominant alpha peak in aging, also in a follow-up study in large population of DS subjects. Here, oscillations in resting EEGs from the frontal, central and occipital regions were examined in groups of subjects in intervals of 5 years. The number of subjects with DS who showed a dominant component within 8 Hz band of the basic rhythm reached maximum in its appearance rate at 40 to 44 years of age in the occipital area, while in healthy subjects the presence of this 8 Hz rhythm at this age was in a minority of cases. Confirming previous results, a slowing in this basic rhythm

progressed already at 30 to 34 years in DS subjects. Opposite to this, in non-DS mental retardation, the number of subjects who showed dominant component at 8 Hz reached maximum at 45 to 49 years of age, and this slowing of the basic rhythm was not as clear as in DS. In a parallel follow-up study published in the same article, EEGs were recorded repeatedly once a year during 8 or 9 years in persons with DS and with non-DS mental retardation. The result was that although the lowering in EEG frequency to 8 Hz took place in various years of age individually, earlier distinct decrease of the frequency was commonly noticed in DS subjects (Katada et al., 2000). As stated, in this last two studies the authors hypothesized that the earlier slowing could be a senile sign and be related to the decline of brain function referring to AD. Supporting this idea, decreases in theta oscillations in several regions of the neocortex negatively correlated with cognitive impairment, and could discriminate cognitive decline in DS adults with AD (Salem et al., 2015). A summary of the findings in this section is available in **Table 2**.

## SLEEP AND OSCILLATIONS IN DS MOUSE MODELS

In parallel to the previously mentioned work in humans, EEG signals have also been obtained in several mouse models of DS. The analysis of neuronal oscillatory activity in DS mouse models started of course in the 2000's with the irruption of transgenesis and related technologies. Since then, EEG-sleep patterns have been studied in a few of those models. In a first study in two transgenic mice, hSOD and hAPP -overexpressing the human genes SOD1 and APP, respectively, Colas et al., 2004 found that the number of REM sleep episodes were decreased and



REM latency increased after lights off in hSOD mice, similarly to some sleep features in DS humans. In contrast, hAPP mice exhibited no change in REM sleep but an increase in waking and a decrease in slow wave sleep before light transition as well as an increase in theta power in REM and waking, also found in DS humans. In addition, sleep deprivation affected differently both models: while hSOD mice did not experience slow wave sleep or REM sleep rebounds after sleep deprivation but EEG activity in the delta-SWS activity was enhanced, hAPP mice exhibited slow wave sleep and REM sleep rebounds as well as enhancement of delta-slow wave activity (Colas et al., 2004). Similar features were studied in the same lab in trisomic models – Ts65Dn and TsjCe trisomic mice carrying different lengths, respectively, of extra chromosomal fragments homologous to regions in human chromosome 21. They found increased waking periods at the expense of non-REM sleep, increased power in theta waves during sleep and a delayed sleep rebound after sleep deprivation in Ts65Dn mice. In contrast, Ts1Cje had limited sleep and EEG abnormalities, showing only a delayed sleep rebound after sleep deprivation and no difference in the power of theta oscillations (Colas et al., 2008). As a hypothesis extracted from those studies, where theta oscillations were altered in hAPP and Ts65Dn mice, the authors suggested a possible correlation between APP over-expression and changes in hippocampal theta oscillations. A recent work provided evidence of altered sleep patterns in the trisomic Dp16 model, where increased waking time at the expense of NREM sleep was found. Despite alterations in oscillatory activity did not recapitulate the alpha phenotype observed in DS humans, EEG recordings showed theta, beta, and alpha abnormalities across different brain states, suggesting also impairment in the circuits (Levenga et al., 2018).

Undoubtedly, an interesting feature of mouse models is the possibility to underpin the mechanisms underlying the observed alterations. In this direction, slow wave activity was recorded in slices of the somatosensory cortex of Ts65Dn mice to elucidate the mechanisms behind sensory deficits in DS (Sutton, 2005; Bruni et al., 2010). In a characterization of the intrinsic and network properties of layer 4 regular spiking pyramidal neurons in the

somatosensory cortex, spontaneous excitatory and inhibitory synaptic inputs to this cell type were found to be reduced, as well as the duration of neuronal UP states -i.e., active states in the slow oscillation that alternate with silent or DOWN states at ~1 Hz. Importantly, the authors provide results of intracellular recordings that show a decreased intrinsic excitability of the layer 4 regular spiking cells, which explains the reduced synaptic activity and the shorter duration of UP states (Cramer et al., 2015). These findings contribute to the idea that cortical network in DS may be unbalanced, or at least less excitable due to intrinsic properties of neurons, as this work shows.

We performed a multidisciplinary work that studied oscillatory activity in the transgenic DS model TgDyrk1A. This transgenic mouse overexpresses DYRK1A, a double-kinase protein implicated in neurodevelopmental and synaptic processes and a candidate to generate cognitive deficits in DS. Alterations of oscillatory activity were studied in the prefrontal cortex *in vivo* with Local Field Potential (LFP, an EEG in depth) recordings. Neurons' firing rate as well as gamma –30 to 90 Hz- oscillations were found to be reduced during the UP states of slow oscillations in the transgenic mice, which was proven both in waking and anesthetized mice, and providing further evidence in direction of a decreased cortical excitability. Propagation of UP states along the motor cortex was also slowed as they moved away from their site of generation in prefrontal cortex. These results made us hypothesize that the prefrontal cortex of TgDyrk1A is over-inhibited, which was tested by anatomical experiments. Excitatory and inhibitory synaptic contacts to both pyramidal –excitatory- and parvalbumin –inhibitory- neurons were counted, showing that inhibitory contacts were reduced specifically over inhibitory cells. This supported the idea of an effective “super inhibition” of the network, which was tested in a computational model that reported the structural changes were sufficient to explain the observed functional alterations. The findings pointed out to a reduction in recurrent inhibition as a mechanism that may explain cognitive deficits in DS (Ruiz-Mejias et al., 2016). A summary of the findings in DS mouse models is presented in **Table 3**.

**TABLE 3 |** Summary of findings of oscillatory activity in DS mouse models.

First author of the work	Year of publication	DS subjects (recording technique)	Main findings
Colas et al., 2004	2004	SOD1 and hAPP mice (EEG).	Increase in waking and a decrease in slow wave sleep before light transition as well as an increase in theta power in REM and waking.
Colas et al., 2008	2008	Ts1Cje and Ts65Dn mice (EEG).	Increased waking periods at the expense of non-REM sleep in Ts65Dn mice. Increased power in theta oscillations during sleep and a delayed sleep rebound after sleep deprivation in Ts65Dn mice. Limited sleep and EEG abnormalities, delayed sleep rebound after sleep deprivation in TsCje mice.
Cramer et al., 2015	2015	Ts65Dn mice slices (intracellular).	Reduced spontaneous excitatory and inhibitory synaptic inputs to somatosensory cortex L4 regular spiking pyramidal neurons. Reduced UP state duration in these cells. Decreased intrinsic excitability in these cells.
Ruiz-Mejias et al., 2016	2016	TgDyrk1A mice (LFP).	Decreased firing rates and gamma oscillations in prefrontal cortex of anesthetized and waking mice. Slower propagation of UP states in frontal cortex of anesthetized mice. Reduced inhibitory contacts to inhibitory cells.
Levenga et al., 2018	2018	Dp(16) 1Yey/+ mice (EEG).	Theta decreased –within the light phase- and beta increased in waking. Alpha and beta waves increased and delta decreased in NREM sleep. Beta increased in REM sleep. Increased time awake at the expense of NREM sleep.

## WHERE DO WE COME FROM?

This paper reviews the literature on interictal EEG sleep and oscillatory activity in DS individuals, gathering as well the findings in this field on DS mouse models. Previous studies characterized well the sleep activity in DS, including those episodes related with epileptic activity. As well, EEG studies provided first insights of oscillatory activity in DS humans both during sleep and wake periods, and also associated to the development and aging life periods. These works, conversely, were constraint by limitations in spatial and temporal resolution, due to the fact of being non-invasive techniques and the technical development in those years, even the LORETA techniques approach provided spatial localization of EEG current sources in outer brain areas.

The field of sleep and oscillations in DS is diverse, and as stated above, literature has appeared dispersedly during the last decades, showing widespread results in humans and mouse models. Since the early polysomnographic studies a remarkable leap has been carried out from the main observational studies in humans, and this review shows there is a wide and open field of study for research in this topic. The works presented here point out to unbalanced network excitation and inhibition in underlying cortical areas in DS, which generate aberrant rhythmic activity at different levels.

Brain architecture alterations have been also characterized DS humans (Crome et al., 1966; Weis et al., 1991; Pinter et al., 2001) as well as in trisomic DS models (Aldridge et al., 2007). Given that the origin of EEG/LFP signals comes from both structural and synaptic components (Buzsáki et al., 2012), the observed phenotypes in oscillatory patterns both in DS humans and models may have also a structural component comprising the quantity and packing of all neuronal compartments in the local circuit, including synapses.

## WHERE DO WE GO?

Our lab work demonstrates that oscillatory activity, understood as an emergent property from the outer networks of the brain, is useful to infer alterations in the underlying circuits. Moreover, novel biomarkers may be obtained from the study of this activity, that might be used to track the effect of new therapeutic approaches, both at basic and clinical level. At this level, an alpha rhythm decrease in DS individuals as well as a slowing of a dominant alpha peak in aging DS subjects seem to be consistent findings across studies that could be used with this purpose. Further works can also take a closer look at the communication between cortical areas with means of cross-correlation and coherence analysis of network signals, or increase the spatial resolution and improve tissue filtering effects of brain surface recordings with means of MEG techniques to extract gamma rhythms related to cognition (Rojas et al., 2008; Buzsáki and Wang, 2012; Siegel et al., 2012). EEG recordings –and video-EEG– are critical to diagnosis and monitor alterations in neurological processes such as epilepsy (Aldrich and Jahnke, 1991; Cascino, 2002; Staba et al., 2014) or useful to early characterize dementia

or AD (Al-Qazzaz et al., 2014; Houmani et al., 2018). A further use may be the inclusion of EEG/MEG recordings in clinical trials to track the effect of novel therapeutic strategies in DS, such as the green tea polyphenol and specific DYRK1A inhibitor epigallocatechin-3-gallate (De la Torre et al., 2014; Stagni et al., 2017). Importantly, EEG and LFP recordings in mice will provide valuable information of oscillatory activity regarding aneuploidy mimicking DS, as well as the underlying circuit impairment that explain the observed emerging alterations. The fostering of new data in DS brains will improve the knowledge of how brain works in intellectual disability to eventually enable a better quality of life for DS patients.

## ALPHA AND GAMMA RHYTHMS AS POTENTIAL BIOMARKERS OF CORTICAL INHIBITION IMPAIRMENT

It has been shown that event-related synchronization (ERS) in the alpha band has a functional correlate of inhibition in cognitive, motor or auditory tasks (Klimesch et al., 2007; Jensen and Mazaheri, 2010; Strauß et al., 2014). Despite this, there is few knowledge of which would be the underlying circuit mechanisms supporting this process. In this review we showed how changes in oscillatory activity could serve as a biomarker of alterations in the underlying circuit. To the date, the neurophysiological mechanisms responsible of alpha rhythms are still unclear, although GABAergic feedback from interneurons has been proposed as a participating mechanism in the generation of these oscillations (Jones et al., 2000; Lőrincz et al., 2009). Here we show a consistent decrease and slowing in alpha waves in DS. The pending questions are whether these alterations in oscillatory activity are related to changes in the effective functional inhibition, and also which is the alteration in the cortical circuitry that is supporting impairment in oscillatory activity. A monoaminergic unbalance – i.e., noradrenaline and serotonin – has been also described in aging DS humans (Dekker et al., 2017), which may be related to the observed slowing of the EEG oscillatory activity. In turn, there are consistent evidences that the GABAergic system is responsible of generating gamma oscillations (Cardin et al., 2009; Buzsáki and Wang, 2012; Siegel et al., 2012). The precise timing of the synaptic activity of this system, the structural synaptic dispositions and the interplay with resonance generators at cellular level in generating either alpha and gamma oscillatory activity remains elusive, and further research will be needed to elucidate the putative dynamical capacity of inhibitory circuitry to participate in both kinds of activity, with emphasis in alpha oscillations.

## CONCLUSION

This article reviews the literature on the relationship of sleep and oscillatory activity in DS, coming from activity from the outer areas of the brain. A consistent phenotype has been observed regarding alterations that down-regulate alpha oscillations in



different brain states, with a general up-regulation of EEG oscillatory activity in sleep. During brain circuits' development across childhood, DS children show a lag of maturation of rhythmic patterns with an uneven settlement and expression of alpha waves. In turn, the dominant alpha rhythm of the EEG experiences a slowing from adulthood and across lifespan, with may be related in aging DS humans to dementia. This review also claims mouse models are valuable tools for characterizing the network mechanisms behind intellectual disability observed in DS humans. Results are presented from studies that try to go a step beyond the observation and provide an explanation for the alterations observed at network level, thus demonstrating a decrease in neuronal activity in DS neocortex. These alterations are present either at intrinsic or synaptic levels, which are not exclusive. In addition to the early work focused on alpha oscillations, a deeper understanding of further oscillatory patterns is claimed here, such as gamma oscillations, which are directly related to cognitive function. Research work also provides evidence that functional and anatomical synaptic

studies in transgenic/trisomic models may help to determine the underlying mechanisms of cognitive dysfunction in DS. Future work might cover the gaps to determine the relationship between intrinsic and synaptic neuronal properties and microcircuit contributions in the generation of altered network oscillatory patterns implicated in cognition in DS models, and clarify whether alterations in the neocortex are ubiquitous in the whole cortex or rather area specific.

## AUTHOR CONTRIBUTIONS

MR-M wrote the manuscript.

## ACKNOWLEDGMENTS

I would like to thank Thomas Gener and Silvia Milazzo for scientific and language review of the manuscript, respectively.

## REFERENCES

- Aldrich, M. S., and Jahnke, B. (1991). Diagnostic value of video-EEG polysomnography. *Neurology* 41:1060.
- Aldridge, K., Reeves, R. H., Olson, L. E., and Richtsmeier, J. T. (2007). Differential effects of trisomy on brain shape and volume in related aneuploid mouse models. *Am. J. Med. Genet. Part A* 143, 1060–1070.
- Al-Qazzaz, N. K., Ali, S. H. B., Ahmad, S. A., Chellappan, K., Islam, M., and Escudero, J. (2014). Role of EEG as biomarker in the early detection and classification of dementia. *Sci. World J.* 2014:906038. doi: 10.1155/2014/906038
- Babiloni, C., Albertini, G., Onorati, P., Muratori, C., Buffo, P., Condoluci, C., et al. (2010). Cortical sources of EEG rhythms are abnormal in down syndrome. *Clin. Neurophysiol.* 121, 1205–1212. doi: 10.1016/j.clinph.2010.02.155
- Babiloni, C., Albertini, G., Onorati, P., Vecchio, F., Buffo, P., Sarà, M., et al. (2009). Inter-hemispheric functional coupling of eyes-closed resting EEG rhythms in adolescents with down syndrome. *Clin. Neurophysiol.* 120, 1619–1627. doi: 10.1016/j.clinph.2009.06.017
- Barnet, A. B., and Lodge, A. (1967). Click evoked EEG responses in normal and developmentally retarded infants. *Nature* 214, 252–255.
- Beley, A., Sevestre, P., Lecuyer, R. J., and Leroy, C. (1959). Contribution to the EEG of mongoloids. *Rev. Neurol.* 101, 457–459.
- Bigum, H. B., Dustman, R. E., and Beck, E. C. (1970). Visual and somato-sensory evoked responses from mongoloid and normal children. *Electroencephalogr. Clin. Neurophysiol.* 28, 576–585.
- Borselli, L., and Sferlazzo, R. (1963). Considerations on the problem of epilepsy in mongolism. *Riv. Clin. Pediatr.* 72, 45–53.
- Bruni, M., Cameron, D., Dua, S., and Noy, S. (2010). Reported sensory processing of children with Down syndrome. *Phys. Occup. Ther. Pediatr.* 30, 280–293. doi: 10.3109/01942638.2010.486962
- Buzsáki, G. (1998). Memory consolidation during sleep: a neurophysiological perspective. *J. Sleep Res.* 7, 17–23.
- Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* 13, 407. doi: 10.1038/nrn3241
- Buzsáki, G., and Wang, X. J. (2012). Mechanisms of gamma oscillations. *Annu. Rev. Neurosci.* 35, 203–225. doi: 10.1146/annurev-neuro-062111-150444
- Cardin, J. A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., et al. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* 459:663. doi: 10.1038/nature08002
- Cascino, G. D. (2002). Video-EEG monitoring in adults. *Epilepsia* 43, 80–93.
- Castaldo, V. (1969). Down's syndrome: a study of sleep patterns related to level of mental retardation. *Am. J. Mental Defic.* 74, 187–190.
- Chauvette, S., Crochet, S., Volgushev, M., and Timofeev, I. (2011). Properties of slow oscillation during slow-wave sleep and anesthesia in cats. *J. Neurosci.* 31, 14998–15008. doi: 10.1523/JNEUROSCI.2339-11.2011
- Cirelli, C., and Tononi, G. (2015). Sleep and synaptic homeostasis. *Sleep* 38, 161–162.
- Clausen, J., Sersen, E. A., and Lidsky, A. (1977). Sleep patterns in mental retardation: down's syndrome. *Electroencephalogr. Clin. Neurophysiol.* 43, 183–191.
- Colas, D., London, J., Gharib, A., Cesputio, R., and Sarda, N. (2004). Sleep-wake architecture in mouse models for down syndrome. *Neurobiol. Dis.* 16, 291–299.
- Colas, D., Valletta, J. S., Takimoto-Kimura, R., Nishino, S., Fujiki, N., Mobley, W. C., et al. (2008). Sleep and EEG features in genetic models of Down syndrome. *Neurobiol. Dis.* 30, 1–7. doi: 10.1016/j.nbd.2007.07.014
- Contreras, D., Timofeev, I., and Steriade, M. (1996). Mechanisms of long-lasting hyperpolarizations underlying slow sleep oscillations in cat corticothalamic networks. *J. Physiol.* 494, 251–264.
- Cramer, N. P., Xu, X., Haydar, T. F., and Galdzicki, Z. (2015). Altered intrinsic and network properties of neocortical neurons in the Ts65Dn mouse model of down syndrome. *Physiol. Rep.* 3:e12655. doi: 10.14814/phy2.12655
- Crome, L., Cowie, V., and Slater, E. (1966). A statistical note on cerebellar and brain-stem weight in mongolism. *J. Intellect. Disabil. Res.* 10, 69–72. doi: 10.1111/j.1365-2788.1966.tb00173.x
- De la Torre, R., De Sola, S., Pons, M., Duchon, A., de Lagran, M. M., Farré, M., et al. (2014). Epigallocatechin-3-gallate, a DYRK1A inhibitor, rescues cognitive deficits in Down syndrome mouse models and in humans. *Mol. Nutr. Food Res.* 58, 278–288.
- Dekker, A. D., Vermeiren, Y., Albac, C., Lana-Elola, E., Watson-Scales, S., Gibbins, D., et al. (2017). Aging rather than aneuploidy affects monoamine neurotransmitters in brain regions of down syndrome mouse models. *Neurobiol. Dis.* 105, 235–244. doi: 10.1016/j.nbd.2017.06.007
- Devinsky, O., Sato, S., Conwit, R. A., and Schapiro, M. B. (1990). Relation of EEG alpha background to cognitive function, brain atrophy, and cerebral metabolism in Down's syndrome: age-specific changes. *Arch. Neurol.* 47, 58–62.
- Feinberg, I., Braun, M., and Shulman, E. (1969). EEG sleep patterns in mental retardation. *Electroencephalogr. Clin. Neurophysiol.* 27, 128–141.
- Gibbs, F., Gibbs, E., and Hirsch, W. (1965). Unusual EEG findings in mongoloids. *Elektro Med. Biomed.* 10, 73–75.
- Gunnarson, S. (1945). Electro-encephalographic examinations of imbeciles. Regarding the alpha frequency in spastics and mongoloid idiots. *Acta Paediatr.* 32, 426–434.
- Hamaguchi, H., Hashimoto, T., Mori, K., and Tayama, M. (1989). Sleep in the down syndrome. *Brain Dev.* 11, 399–406.

- Hirai, T., and Izawa, S. (1964). An electroencephalographic study of mongolism—with special reference to its eeg development and intermediate fast wave. *Seishin shinkeigaku zasshi* 66, 166–177.
- Houmani, N., Vialatte, F., Gallego-Jutglà, E., Dreyfus, G., Nguyen-Michel, V. H., Mariani, J., et al. (2018). Diagnosis of Alzheimer's disease with electroencephalography in a differential framework. *PloS one* 13:e0193607. doi: 10.1371/journal.pone.0193607
- Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186
- Jones, S. R., Pinto, D. J., Kaper, T. J., and Kopell, N. (2000). Alpha-frequency rhythms desynchronize over long cortical distances: a modeling study. *J. Comput. Neurosci.* 9, 271–291.
- Katada, A., Hasegawa, S., Ohira, D., Kumagai, T., Harashima, T., Ozaki, H., et al. (2000). On chronological changes in the basic EEG rhythm in persons with down syndrome—with special reference to slowing of alpha waves. *Brain Dev.* 22, 224–229.
- Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition—timing hypothesis. *Brain Res. Rev.* 53, 63–88.
- Levenga, J., Peterson, D. J., Cain, P., and Hoeffer, C. A. (2018). Sleep behavior and EEG oscillations in aged Dp (16) 1Yey/+ mice: a down syndrome model. *Neuroscience* 376, 117–126. doi: 10.1016/j.neuroscience.2018.02.009
- Lőrincz, M. L., Kékesi, K. A., Juhász, G., Crunelli, V., and Hughes, S. W. (2009). Temporal framing of thalamic relay-mode firing by phasic inhibition during the alpha rhythm. *Neuron* 63, 683–696. doi: 10.1016/j.neuron.2009.08.012
- McAlaster, R. (1992). Postnatal cerebral maturation in down's syndrome children: a developmental EEG coherence study. *Int. J. Neurosci.* 65, 221–237.
- McCormick, D. A., and Bal, T. (1997). Sleep and arousal: thalamocortical mechanisms. *Annu. Rev. Neurosci.* 20, 185–215.
- Miano, S., Bruni, O., Elia, M., Scifo, L., Smerieri, A., Trovato, A., et al. (2008). Sleep phenotypes of intellectual disability: a polysomnographic evaluation in subjects with Down syndrome and Fragile-X syndrome. *Clin. Neurophysiol.* 119, 1242–1247. doi: 10.1016/j.clinph.2008.03.004
- Miyamoto, D., Hirai, D., Fung, C. C. A., Inutsuka, A., Odagawa, M., Suzuki, T., et al. (2016). Top-down cortical input during NREM sleep consolidates perceptual memory. *Science* 352, 1315–1318. doi: 10.1126/science.aaf0902
- Murata, T., Koshino, Y., Omori, M., Murata, I., Nishio, M., Horie, T., et al. (1994). Quantitative EEG study on premature aging in adult down's syndrome. *Biol. Psychiatry* 35, 422–425.
- Ono, Y. (1993). EEG changes with aging in adults with down syndrome. *Psychiatry Clin. Neurosci.* 47, 75–84.
- Ono, Y., Yoshida, H., Momotani, Y., Yoshimasu, F., and Higashi, Y. (1992). Age-related changes in occipital alpha rhythm of adults with down syndrome. *Psychiatry Clin. Neurosci.* 46, 659–664.
- Partanen, J., Soininen, H., Könönen, M., Kilpeläinen, R., Helkala, E. L., and Sr, P. R. (1996). EEG reactivity correlates with neuropsychological test scores in down's syndrome. *Acta Neurol. Scand.* 94, 242–246.
- Petre-Quadens, O., and Jouvet, M. (1967). Sleep in the mentally retarded: material and methods. *J. Neurol. Sci.* 4, 354–357.
- Pinter, J. D., Eliez, S., Schmitt, J. E., Capone, G. T., and Reiss, A. L. (2001). Neuroanatomy of Down's syndrome: a high-resolution MRI study. *Am. J. Psychiatry* 158, 1659–1665. doi: 10.1176/appi.ajp.158.10.1659
- Rojas, D. C., Maharajh, K., Teale, P., and Rogers, S. J. (2008). Reduced neural synchronization of gamma-band MEG oscillations in first-degree relatives of children with autism. *BMC Psychiatry* 8:66. doi: 10.1186/1471-244X-8-66
- Ruiz-Mejias, M., Ciria-Suarez, L., Mattia, M., and Sanchez-Vives, M. V. (2011). Slow and fast rhythms generated in the cerebral cortex of the anesthetized mouse. *J. Neurophysiol.* 106, 2910–2921. doi: 10.1152/jn.00440.2011
- Ruiz-Mejias, M., de Lagran, M. M., Mattia, M., Castano-Prat, P., Perez-Mendez, L., Ciria-Suarez, L., et al. (2016). Overexpression of Dyrk1A, a Down syndrome candidate, decreases excitability and impairs gamma oscillations in the prefrontal cortex. *J. Neurosci.* 36, 3648–3659. doi: 10.1523/JNEUROSCI.2517-15.2016
- Salem, L. C., Sabers, A., Kjaer, T. W., Musaeus, C., Nielsen, M. N., Nielsen, A. G., et al. (2015). Quantitative electroencephalography as a diagnostic tool for Alzheimer's dementia in adults with down syndrome. *Dement. Geriatr. Cogn. Dis. Extra* 5, 404–413. doi: 10.1159/000438857
- Sanchez-Vives, M. V., and Mattia, M. (2014). Slow wave activity as the default mode of the cerebral cortex. *Arch. Ital. Biol.* 152, 147–155. doi: 10.12871/000298292014239
- Schmid, R. G., Sadowsky, K., Weinmann, H. M., Tirsch, W. S., and Pöppel, S. J. (1985). Z-transformed EEG power spectra of children with down syndrome vs a control group. *Neuropediatrics* 16, 218–224.
- Seppäläinen, A. M., and Kivalo, E. (1967). EEG findings and epilepsy in down's syndrome. *J. Intellect. Disabil. Res.* 11, 116–125.
- Siegel, M., Donner, T. H., and Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nat. Rev. Neurosci.* 13:121. doi: 10.1038/nrn3137
- Śmigielska-Kuzia, J., Sobaniec, W., Kulak, W., Boakowski, L., and Sołowiej, E. (2005). Quantitative EEG analysis of REM sleep in children with Down syndrome. *Adv. Med. Sci.* 50, 20–22.
- Soininen, H., Partanen, J., Jousma, V., Helkala, E. L., Vanhanen, M., Majuri, S., et al. (1993). Age-related cognitive decline and electroencephalogram slowing in down's syndrome as a model of Alzheimer's disease. *Neuroscience* 53, 57–63.
- Staba, R. J., Stead, M., and Worrell, G. A. (2014). Electrophysiological biomarkers of epilepsy. *Neurotherapeutics* 11, 334–346. doi: 10.1007/s13311-014-0259-0
- Stagni, F., Giacomini, A., Emili, M., Guidi, S., Ciani, E., and Bartesaghi, R. (2017). Epigallocatechin gallate: a useful therapy for cognitive disability in down syndrome? *Neurogenesis* 4:e1270383. doi: 10.1080/23262133.2016.1270383
- Steriade, M. (2006). Grouping of brain rhythms in corticothalamic systems. *Neuroscience* 137, 1087–1106.
- Steriade, M., Dossi, R. C., and Nunez, A. (1991). Network modulation of a slow intrinsic oscillation of cat thalamocortical neurons implicated in sleep delta waves: cortically induced synchronization and brainstem cholinergic suppression. *J. Neurosci.* 11, 3200–3217.
- Strauß, A., Wöstmann, M., and Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Front. Hum. Neurosci.* 8:350. doi: 10.3389/fnhum.2014.00350
- Sutton, S. (2005). "Sensory integration strategies for skill building," in *Workshop at the 2005 Canadian Down Syndrome Society Conference*, Victoria.
- Velikova, S., Magnani, G., Arcari, C., Falautano, M., Franceschi, M., Comi, G., et al. (2011). Cognitive impairment and EEG background activity in adults with Down's syndrome: a topographic study. *Hum. Brain Mapp.* 32, 716–729. doi: 10.1002/hbm.21061
- Vergani, O., and Aldeghi, E. (1958). EEG findings in mongoloid phrenasthenia. *Minerva Pediatr.* 10, 1378–1383.
- Virji-Babul, N., Moiseev, A., Cheung, T., Weeks, D., Cheyne, D., and Ribary, U. (2008). Changes in mu rhythm during action observation and execution in adults with Down syndrome: Implications for action representation. *Neurosci. Lett.* 436, 177–180. doi: 10.1016/j.neulet.2008.03.022
- Virji-Babul, N., Moiseev, A., Cheung, T., Weeks, D. J., Cheyne, D., and Ribary, U. (2010). Neural mechanisms underlying action observation in adults with down syndrome. *Am. J. Intellect. Dev. Disabil.* 115, 113–127. doi: 10.1352/1944-7588-115.2.113
- Visser, F. E., Kuilman, M., Oosting, J., Overweg, J., Van Wijk, J., and Van Huffelen, A. C. (1996). Use of electroencephalography to detect Alzheimer's disease in down's syndrome. *Acta Neurol. Scand.* 94, 97–103.
- Weis, S., Weber, G., Neuhold, A., and Rett, A. (1991). Down syndrome: MR quantification of brain structures and comparison with normal control subjects. *Am. J. Neuroradiol.* 12, 1207–1211.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ruiz-Mejias. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Coherence of Visual-Evoked Gamma Oscillations Is Disrupted by Propofol but Preserved Under Equipotent Doses of Isoflurane

Adeeti Aggarwal<sup>1</sup>, Connor Brennan<sup>1</sup>, Brenna Shortal<sup>1</sup>, Diego Contreras<sup>1</sup>, Max B. Kelz<sup>2</sup> and Alex Proekt<sup>2\*</sup>

<sup>1</sup> Department of Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States,

<sup>2</sup> Department of Anesthesiology and Critical Care, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

## OPEN ACCESS

### Edited by:

Alain Destexhe,  
FRE3693 Unité de Neuroscience,  
Information et Complexité (UNIC),  
France

### Reviewed by:

Gregor Rainer,  
Université de Fribourg, Switzerland  
Ramón Reig,  
Institute of Neurosciences of  
Alicante (IN), Spain

### \*Correspondence:

Alex Proekt  
Alexander.Proekt@uphs.upenn.edu;  
proekt@gmail.com

**Received:** 01 February 2019

**Accepted:** 18 April 2019

**Published:** 08 May 2019

### Citation:

Aggarwal A, Brennan C,  
Shortal B, Contreras D, Kelz MB and  
Proekt A (2019) Coherence  
of Visual-Evoked Gamma Oscillations  
Is Disrupted by Propofol but  
Preserved Under Equipotent Doses  
of Isoflurane.  
Front. Syst. Neurosci. 13:19.  
doi: 10.3389/fnsys.2019.00019

Previous research demonstrates that the underlying state of the brain influences how sensory stimuli are processed. Canonically, the state of the brain has been defined by quantifying the spectral characteristics of spontaneous fluctuations in local field potentials (LFP). Here, we utilized isoflurane and propofol anesthesia to parametrically alter the spectral state of the murine brain. With either drug, we produce slow wave activity, with low anesthetic doses, or burst suppression, with higher doses. We find that while spontaneous LFP oscillations were similar, the average visual-evoked potential (VEP) was always smaller in amplitude and shorter in duration under propofol than under comparable doses of isoflurane. This diminished average VEP results from increased trial-to-trial variability in VEPs under propofol. One feature of single trial VEPs that was consistent in all animals was visual-evoked gamma band oscillation (20–60 Hz). This gamma band oscillation was coherent between trials in the early phase (<250 ms) of the visual evoked potential under isoflurane. Inter trial phase coherence (ITPC) of gamma oscillations was dramatically attenuated in the same propofol anesthetized mice despite similar spontaneous oscillations in the LFP. This suggests that while both anesthetics lead to loss of consciousness (LOC), elicit slow oscillations and burst suppression, only the isoflurane permits phase resetting of gamma oscillations by visual stimuli. These results demonstrate that accurate characterization of a brain state must include both spontaneous as well as stimulus-induced perturbations of brain activity.

**Keywords:** isoflurane, propofol, gamma, VEP, visual evoked potential, anesthesia, burst suppression, brain state

## INTRODUCTION

Anesthesia is a staple in modern healthcare due to its ability to provide a reversible state of unconsciousness, which is essential for painless surgery and for sedation in intensive care units (ICUs). Anesthetics have also proved indispensable for basic neuroscience. Indeed, much of our knowledge concerning sensory processing is derived from experiments performed in anesthetized animals (Mountcastle, 1957; Hubel and Wiesel, 1962; Destexhe et al., 1999). Despite their widespread use, the mechanisms by which anesthetics produce a reversible loss of consciousness

(LOC) remain unknown. One practical implication of this knowledge gap is that clinical monitoring of the anesthetized state is unable to guarantee that all patients are, in fact, unconscious during surgery. While depth of anesthesia monitors do ensure that majority of anesthetized patients are unconscious, 4–10% of patients under general anesthesia exhibit a covert return of consciousness as evidenced by their ability to follow simple verbal commands (Russell, 1989; Schneider Gerhard et al., 2005; Mashour et al., 2011; Sanders et al., 2012). Current EEG-based “depth of anesthesia” devices do not reliably detect these episodes of awareness (Mashour, 2006; Sanders, 2016). While patients with covert awareness are less likely to form memories (Mashour et al., 2011; Mashour et al., 2012; Sanders et al., 2012), up to 70% of those that do, develop long-lasting psychiatric consequences such as post-traumatic stress disorder (PTSD) (Leslie et al., 2010).

General anesthetic agents are structurally heterogeneous and exhibit promiscuous binding to a wide variety of molecular targets (Eckenhoff and Eckenhoff, 1998; Lydic and Baghdoyan Helen, 2005; Franks, 2008). It is highly unlikely that each anesthetic drug disrupts consciousness using the same molecular effectors. Nevertheless, mechanistically distinct anesthetics are known to generate similar patterns of brain activity. The most prevalent pattern of brain activity observed in the anesthetized brain are the canonical slow oscillations first demonstrated in human EEG in the 1930's (Gibbs et al., 1937). For example, propofol, a positive allosteric modulator at GABA<sub>A</sub> receptors (Jurd et al., 2003), induces low frequency large amplitude EEG oscillations and sleep-like spindles (Ching et al., 2010; Purdon et al., 2013). Likewise, the inhaled anesthetic, isoflurane, also produces slow wave activity with distinct UP-states and DOWN-states in the EEG (Ferron et al., 2009). While isoflurane also acts on the GABA<sub>A</sub> receptor (Hall et al., 1994), its actions on the GABA<sub>A</sub> receptor are distinct from those of propofol (Krasowski et al., 1998). Furthermore, actions of isoflurane on the GABA<sub>A</sub> receptor appear to be less critical for its ability to induce anesthesia than those of propofol (Sonner et al., 2007). Finally, both propofol and isoflurane interact with a number of other receptors in the nervous system (Eckenhoff, 2002; Weiser et al., 2015).

It is thought that the slow EEG oscillations observed with a variety, but notably not all anesthetics (Maksimow et al., 2006; Akeju et al., 2016), are a consequence of a switch in the activity patterns of thalamic neurons. These neurons shift from tonic firing, which denote awake desynchronized states, to bursting firing pattern, which synchronizes cortical activity. Thalamic bursting activity is thought to prevent reliable transmission of sensory stimuli from the thalamus to the cortex. This hypothesis suggests that, regardless of the molecular mechanism of action, slow oscillations induced by mechanistically distinct anesthetics should lead to similar disruptions of sensory-evoked responses in the cortex.

There is recent evidence, however, to challenge this hypothesis. Arena et al demonstrate that the amplitude of visual-evoked potentials (VEPs) is attenuated by propofol, but enhanced by increasing concentrations of sevoflurane in rats (Arena et al., 2017). Here, we build upon these observations and characterize the differences in visual-evoked responses in mice

under isoflurane and propofol. We find that although there are similarities in the spontaneous activity elicited by hypnotic doses of isoflurane and propofol, visual-evoked responses to simple visual stimuli are quite different in the primary visual cortex (V1). In the time domain, we find that responses evoked by identical visual stimuli vary dramatically between trials under both anesthetics. However, analysis in the frequency-domain reveals a consistent visual-evoked gamma oscillation (20–60 Hz) present in all mice. This gamma oscillation is coherent across trials in the early phase (<250 ms) of the VEP when mice are under isoflurane anesthesia. Despite similar drug-induced brain states, visual-evoked gamma coherence between trials is greatly attenuated when the same mice are anesthetized with steady-state, target controlled infusions of propofol. This suggests that while both anesthetics disrupt consciousness and elicit slow oscillations, only the isoflurane-induced state of unconsciousness permits phase resetting of gamma oscillations by visual stimuli.

## RESULTS

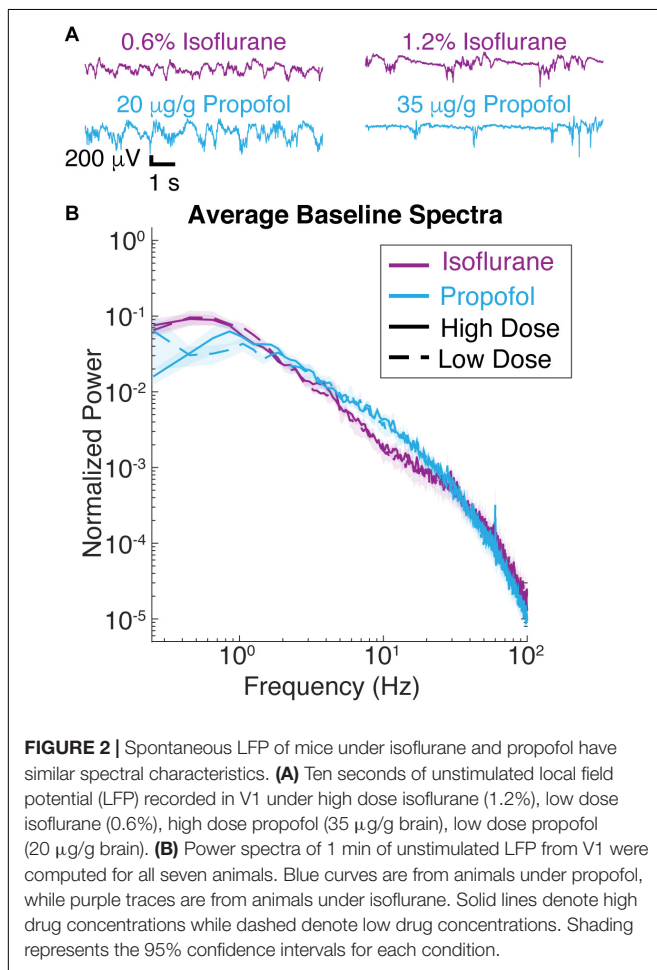
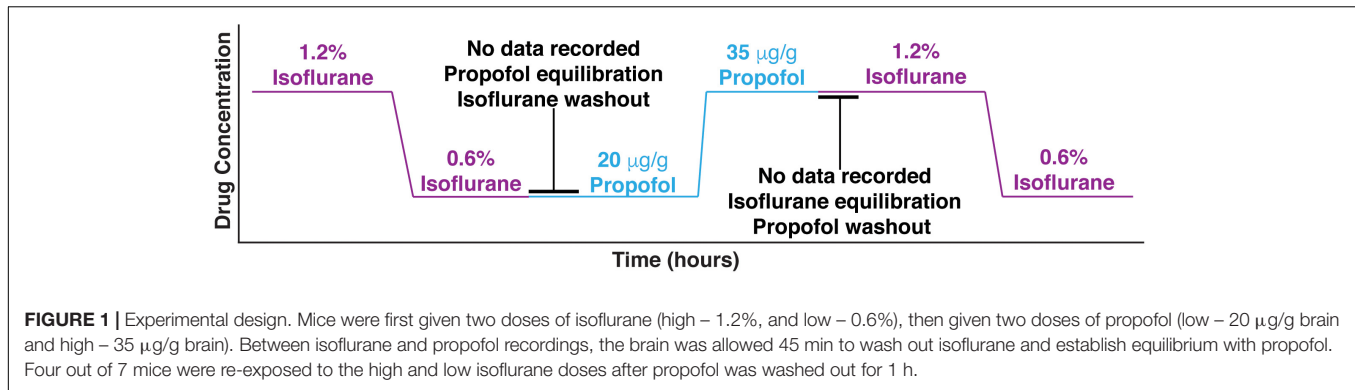
To elucidate the effect of the anesthetic state on visual-evoked brain activity, we performed *in vivo* electrophysiological recordings in mice head-fixed with ear bars ( $n = 7$ ) presented with simple visual stimuli using equipotent doses of two different anesthetic drugs. Each visual stimulus consisted of a short, 10 ms, flash of a green LED light that covered 100% of the right visual field. Spontaneous and evoked local field potentials (LFP) were collected using an electrocorticography (ECoG) electrode placed on top of the dura over the left hemisphere, including primary visual cortex.

### Survey of Spectrally Defined Brain States Under Isoflurane and Propofol

In order to modulate the spontaneous cortical activity, we delivered two different anesthetics: isoflurane and propofol. We were able to maintain steady state concentration of isoflurane via a nose cone delivery due to its relatively fast pharmacokinetics. Propofol was administered intravenously (IV) through a jugular venous catheter with a target controlled infusion (TCI) to ensure that the propofol brain concentration remained constant (Shortal et al., 2018).

Three out of seven mice were first given two doses of isoflurane (high – 1.2%, and low – 0.6%), then given two doses of propofol (low – 20 µg/g brain and high – 35 µg/g brain). The remaining four mice had the aforementioned exposure and were subsequently re-exposed to the same two doses of isoflurane 1 h after propofol was washed out. Re-exposure served as a control for the potential brain desiccation, which might occur during long recording sessions. Re-exposure experiments also established the consistency of specimen preparation (Figure 1). At every anesthetic dose, 1 min of spontaneous activity was collected before visual stimuli were presented. Ten seconds of spontaneous data is shown in Figure 2A, illustrating that burst suppression occurs with high doses of both propofol [median suppression ratio (SR) = 6.65%, interquartile range (IQR) = 25.09%] and isoflurane (median SR = 9.55%,





IQR = 14.74%); and that large amplitude slow waves arise with low doses of propofol (median SR = 0.80%, IQR = 4.23%) and isoflurane (median SR = 1.15%, IQR = 4.23%).

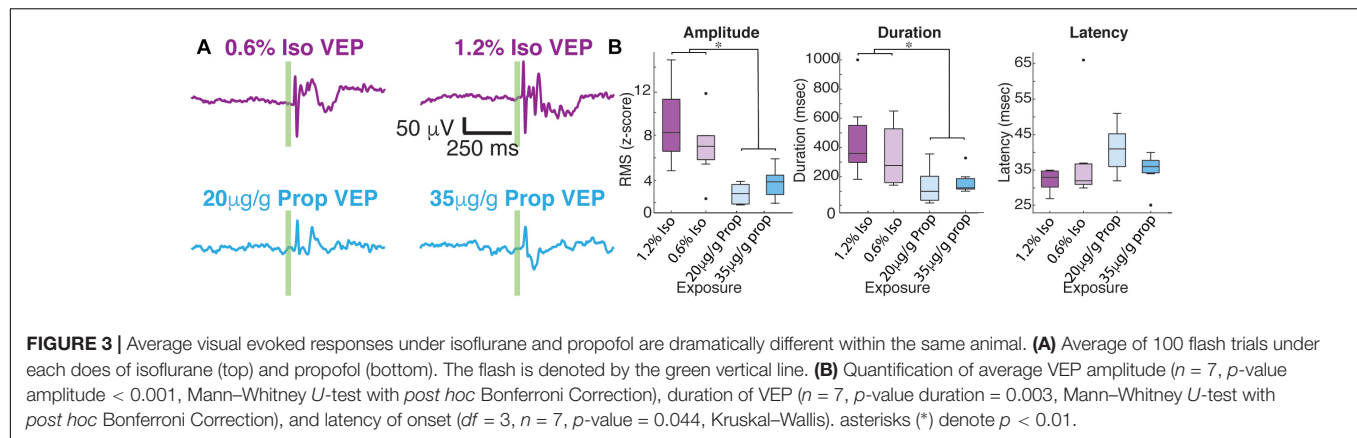
Consistent with these observations, spectra of the ECoG signals under isoflurane and propofol overlap over frequencies ranging from 1 to 4 Hz (**Figure 2B**). While with respect to slow wave activity ECoG spectra under propofol and isoflurane anesthesia were highly similar, there was slightly more power at frequencies between 6 and 10 Hz under propofol. There was

also an increase 0.3–1 Hz under isoflurane compared to propofol ( $df = 3$ ,  $n = 7$ ,  $p = 0.011$ , Kruskal–Wallis, pooled isoflurane vs. pooled propofol,  $n = 7$ ,  $p = 0.005$ , Mann–Whitney  $U$ -test with *post hoc* Bonferroni Correction). Thus, by administering the same animal with these two chemically distinct anesthetics, we can determine how similar slow oscillations induced with two distinct anesthetics affect the characteristics of visual-evoked responses.

## Isoflurane and Propofol Have Dramatically Different Average Visual-Evoked Responses

After 1 min of baseline recording, 100 visual-evoked responses were elicited using a green LED. Averaged VEPs under each anesthetic condition are shown in **Figure 3**. The shape of average VEP has historically been described by its latency to onset, amplitude, and response duration. We defined the latency to onset of the VEP as statistical deviations from the pre-stimulus data (Materials and Methods). Similarly, we define the duration of VEPs by the number of timepoints for the post-stimulus data to recapitulate the pre-stimulus statistics (Materials and Methods). We defined the amplitude of the VEP as the root mean square (RMS) of the first 350 ms of post-stimulus data. No changes in the overall amplitude of the VEP were found for different concentrations of the same anesthetic (low dose isoflurane vs. high dose isoflurane:  $n = 7$ ,  $p = 0.999$ , low dose propofol vs. high dose propofol,  $n = 7$ ,  $p = 0.902$ , Mann–Whitney  $U$ -test). When we normalize the evoked RMS to the RMS calculated from baseline, we still observe that the overall amplitude of the VEP was similar under the two concentrations of isoflurane (low dose isoflurane vs. high dose isoflurane:  $n = 7$ ,  $p = 0.383$ ). Low dose propofol was associated with small but statistically significant decrease in the VEPs relative to high dose propofol ( $n = 7$ ,  $p < 0.001$ , Mann–Whitney  $U$ -test). Furthermore, we were not able to detect differences in duration of VEP at different anesthetic concentrations (low dose isoflurane vs. high dose isoflurane,  $n = 7$ ,  $p = 0.383$ ; low dose propofol vs. high dose propofol,  $n = 7$ ,  $p = 0.209$ , Mann–Whitney  $U$ -test, with *post hoc* Bonferroni Correction). In contrast, differences in VEP characteristics were strongly dependent on the identity of the anesthetic agent. The high and low doses of drugs were combined since there were no dose dependent differences. VEPs under propofol



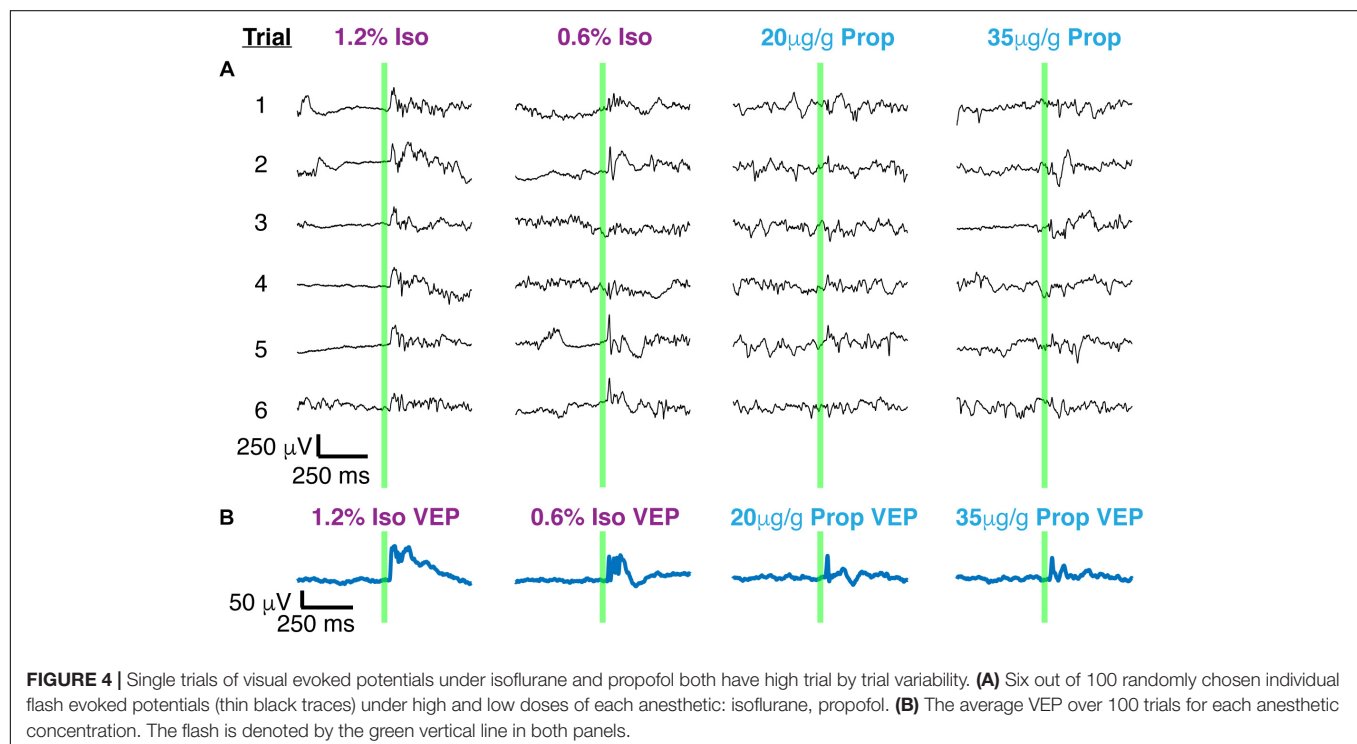


were smaller in amplitude ( $n = 7$ ,  $p$ -value amplitude  $< 0.001$ , Mann–Whitney  $U$ -test with *post hoc* Bonferroni Correction) and shorter in duration ( $n = 7$ ,  $p$ -value duration = 0.003, Mann–Whitney  $U$ -test with *post hoc* Bonferroni Correction) than under isoflurane. Kruskal–Wallis  $U$ -test of the latencies for all four drug conditions was borderline statistically significant ( $df = 3$ ,  $n = 7$ ,  $p$ -value = 0.044). None of the *post hoc* Mann–Whitney  $U$ -tests for pairwise comparisons between drug conditions reached statistical significance.

## Large Trial-to-Trial Variability Under Both Anesthetics

Two distinct scenarios can potentially give rise to the observed differences in the amplitude and duration of the average VEPs:

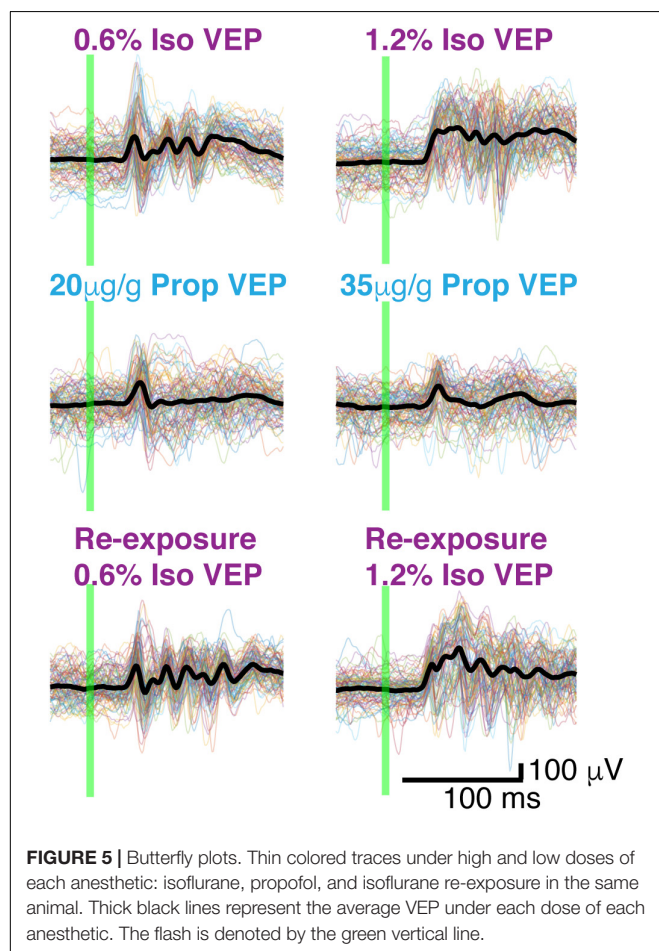
(1) VEPs could be larger in individual trials under isoflurane than under propofol, (2) VEPs could be more consistent among trials under isoflurane. To differentiate between these possibilities, we first surveyed the single trial visual-evoked responses. We found that single trials exhibit large trial-by-trial variability and are strongly dominated by ongoing spontaneous brain activity under both isoflurane and under propofol (**Figure 4A**). We were unable to unequivocally determine the latency of onset or duration of the VEP on a single trial basis since the post-stimulus signal did not deviate significantly from pre-stimulus ECoG. Moreover, we could not find a difference in the amplitude of the single trial responses under isoflurane and propofol ( $\chi^2 = 1.05$ ,  $df = 3$ ,  $n = 7$ ,  $p = 0.197$ , Kruskal–Wallis). Despite this inter-trial variability, averaging across all 100 trials reveals a clear, VEP. The shape of the average



evoked potential, however, rarely resembles any individual trial (**Figure 4B**). Moreover, individual visual evoked trials do not have the same waveform in the time domain. We measured the average pairwise correlation between single trials as an indicator of reliability under each dose of each anesthetic (Tiesinga et al., 2003; Kumbhani et al., 2007). In aggregate, in the time domain, visual evoked single trials are weakly correlated with each other under both anesthetics. Yet, there is a significant increase in the reliability under isoflurane (pooled mean reliability under isoflurane = 0.205) than under propofol (pooled mean reliability under isoflurane = 0.063) ( $n = 7$ ,  $p$ -value pooled reliability  $< 0.001$ , Mann–Whitney  $U$ -test with *post hoc* Bonferroni Correction). This suggests, that the observed differences in the average VEPs under isoflurane and propofol are likely due to differences in the inter-trial consistency of responses rather than to differences in the shape of the VEP on individual trials.

### Visual-Evoked Gamma Power Under Isoflurane and Propofol

While VEPs vary dramatically between trials in many respects, one aspect of the VEP – an oscillation around 40 Hz – was highly consistent between trials and was present in all mice (**Figure 5**).



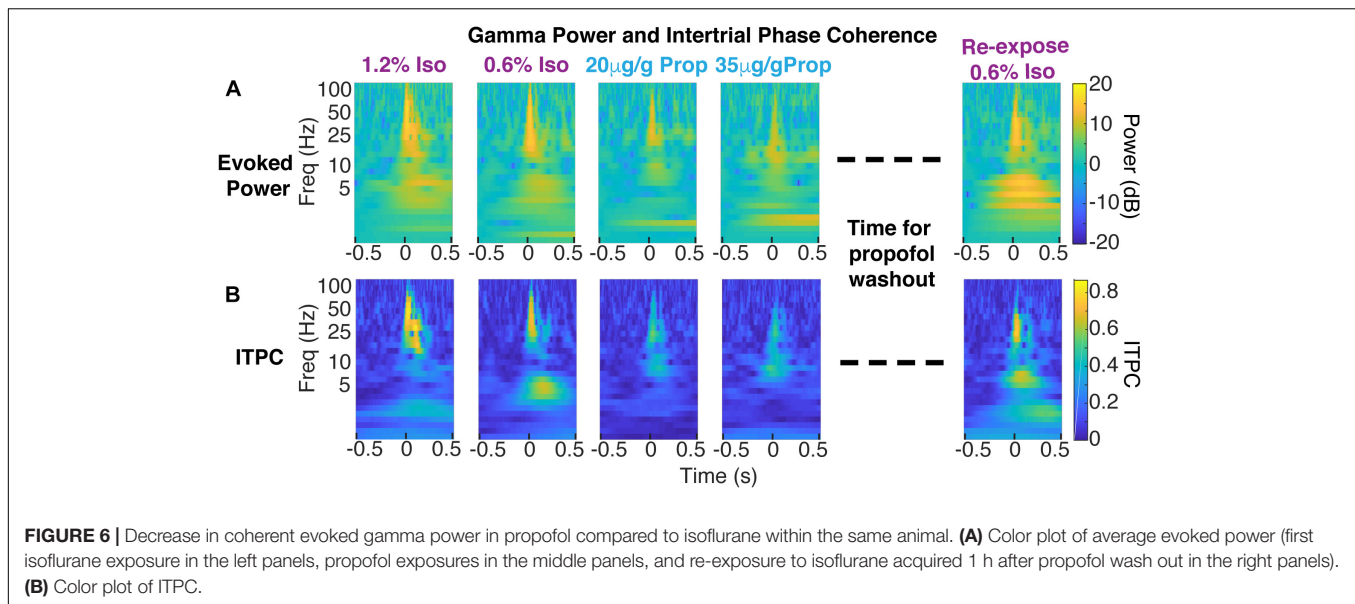
This gamma oscillation can be clearly visualized in the frequency domain (**Figure 6**). To extract the frequency, power and phase characteristics of oscillations present in visual-evoked responses, we convolved single trials with a series of Morlet wavelets. Spectra averaged across trials were then normalized to the pre-stimulus interval (**Figure 6A**). On average, over the first 250 ms, higher gamma power (20–60 Hz), was evoked by the visual stimulus under isoflurane than under propofol (timepoints = 900,  $p < 0.000001$ , Mann–Whitney  $U$ -test).

To determine the variability of the phase of the visual-evoked gamma oscillation, we computed inter-trial phase coherence (ITPC) at each point in the time-frequency plane. Consistent with the observations in the time domain (**Figure 5**), ITPC was significantly increased in the gamma range following the visual stimulus. This increase in the ITPC was most prominent between 50 and 250 ms after stimulus between 20 and 60 Hz (**Figure 6B**). Moreover, the increase in ITPC was larger during anesthesia with isoflurane than propofol. Propofol's reduced ITPC recovered following washout of propofol and re-exposure to isoflurane. Note, that coherence is normalized to signal power. Thus, differences in the ITPC cannot be attributed to higher power of gamma oscillations under isoflurane.

**Figure 7** shows the difference between ITPC evoked under isoflurane and propofol. Here, both high and low concentrations for each individual anesthetic were combined, the anesthetic agent effects are larger than the concentration-dependent effects (**Table 1**). Yellow colors represent higher ITPC under isoflurane while dark blue colors represent higher ITPC under propofol. The maximum difference in evoked coherence occurred 80–130 ms after stimulus onset and was centered at 36 Hz. Indeed, we found a significant increase in the ITPC under isoflurane compared to propofol (timepoints = 900,  $p < 0.000001$ , Mann–Whitney  $U$ -test). Significant decrease in ITPC in the gamma range were present in each mouse (**Table 2**). In 6 out of 7 mice, the visual evoked gamma coherence is statistically greater under isoflurane than under propofol anesthesia with *post hoc* Bonferroni correction. In contrast to the consistent increase in gamma coherence following the visual stimulus, the increase in coherence at lower frequencies (1–5 Hz, centered at 3 Hz) was not consistent among animals. Moreover, none of the average VEPs exhibited a clear oscillation in the 1–5 Hz range that lasted for one or more cycles.

### DISCUSSION

Maintenance of a stable perceptual world is a fundamental requirement of consciousness. In order to create such stable representation of the sensory stimuli, sensory information must be faithfully relayed and integrated with ongoing spontaneous brain activity. The mechanisms through which general anesthetics disrupt perception remain a mystery. Furthermore, it is unknown whether mechanistically distinct classes of anesthetics disrupt sensory processing in a similar manner. Here, we show that although two chemically distinct anesthetics, isoflurane and propofol, produce similar spontaneous ECoG



activity, visual-evoked responses recorded in primary visual cortex obtained during each anesthetic state are quite different. When mice are anesthetized with isoflurane, there is a consistent visual-evoked gamma band oscillation (20–60Hz), which is synchronous across trials. However, when the same mice are anesthetized with propofol, visual-evoked gamma coherence between trials is greatly attenuated. This decrease in consistency of visual responses to identical stimuli likely contributes to the decrease in the size and duration of the visual-evoked responses under propofol. Curiously, both anesthetics elicit similar oscillations in the spontaneous LFP. For instance, under high concentrations of both propofol and isoflurane, the LFP was characterized by burst suppression. Yet, the consistency of elicited responses varied dramatically depending on whether propofol or isoflurane was used to elicit burst suppression. These observations complicate analysis of “brain state” under anesthesia on the basis of spontaneous oscillations in the LFP.

## Sensory Neurophysiology Research Under Anesthesia

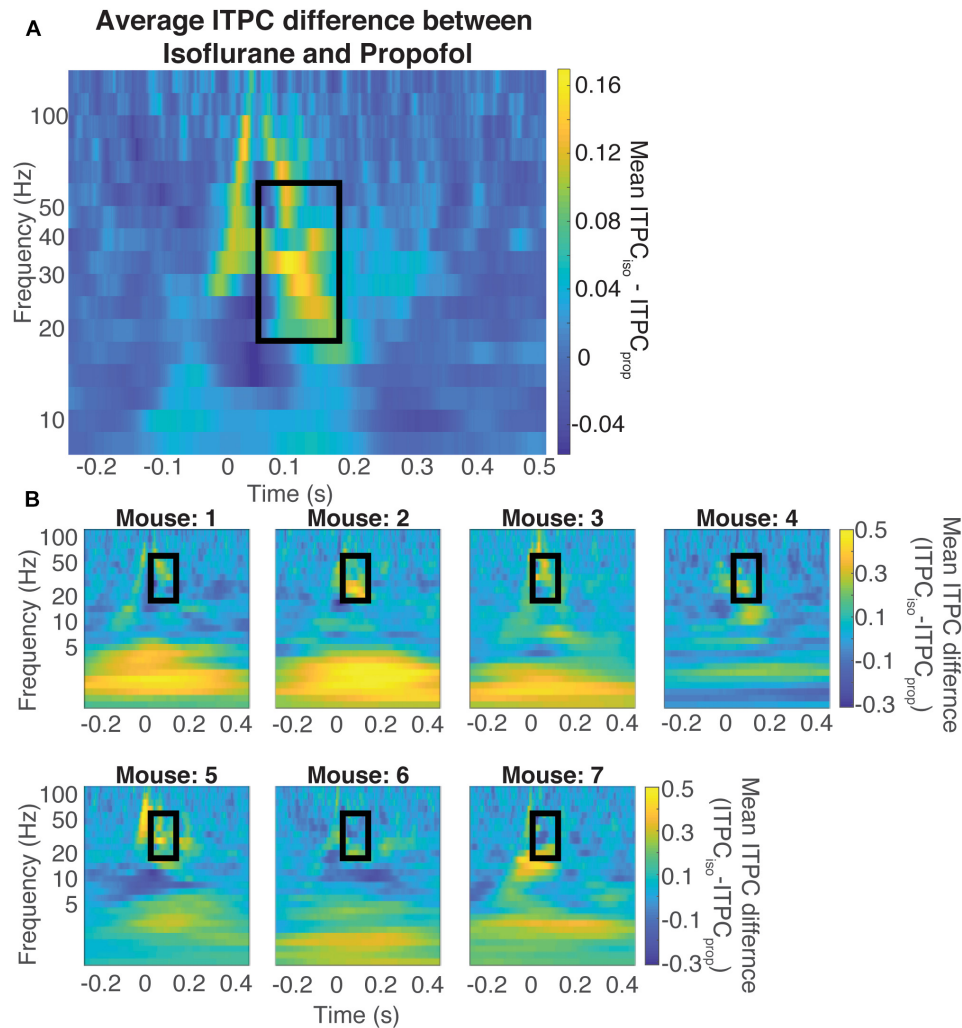
For decades, much of sensory neurophysiology research has been performed in anesthetized preparations. There is increasing evidence that anesthetized and awake sensory responses differ greatly (Imas et al., 2005; Reinhold et al., 2015; Storch et al., 2017). For example, visual cortical responses quickly adapt to a train of high frequency visual stimuli when mice are under isoflurane compared to the awake state (Reinhold et al., 2015). This adaptation is thought to occur because under isoflurane, there is synaptic depression at the level of the lateral geniculate thalamic cells (Reinhold et al., 2015). Furthermore, responses in V1 depend strongly on the behavioral state such as resting vs. running (Niell and Stryker, 2010). It is less obvious, however, that responses to simple visual stimuli in V1 should depend strongly on the anesthetic agent. The

fact that responses to a simple flash in V1 depend strongly upon whether propofol or isoflurane was used to maintain anesthesia is especially surprising because the spontaneous fluctuations in brain activity produced by these anesthetics are very similar.

It is often difficult to determine how animals were anesthetized in existing literature. Materials and Methods sections sometimes note the fluctuations in the spontaneous EEG as a proxy for defining the brain state at the time of recording. However, given the results presented here, along with findings by others (Arena et al., 2017), the type of anesthetic can dramatically alter sensory responses even when the spontaneous oscillations in the ECoG signals are similar. Thus, the traditional characterization of the oscillations of spontaneous brain activity does not appear to unequivocally specify the characteristics of responses evoked by the visual stimulus.

## Possible Mechanisms of Gamma Coherence Breakdown Under Propofol

Variability of evoked responses to identical sensory stimuli limits the ability of these response to reliably convey information about stimulus attributes. Currently, it is not understood which parameters of the visual evoked response encode sensory information. We find that the animal-to-animal and trial-to-trial variability of the visual evoked response decreases in the phase of the visual-evoked gamma oscillations. This makes the phase of the gamma oscillations an appealing candidate for encoding the visual stimulus. Gamma activity can be elicited by one of two prevailing mechanisms. The first arises from strongly activating interneuron-interneuron networks (I-I) (Wang and Buzsáki, 1996; Kopell et al., 2000; Ermentrout and Kleinfeld, 2001; Vinck et al., 2015). The second comes from reciprocally activating interneurons and pyramidal neurons (E-I) (Wilson and Cowan, 1972; Traub et al., 1996; Whittington et al., 2000;



**FIGURE 7 |** Difference in coherence is in the evoked gamma band. Average difference between the ITPC under both doses of isoflurane and propofol (**A**) Yellow colors represent higher ITPC under isoflurane while dark blue colors represent higher ITPC under propofol. The maximum difference in evoked coherence occurs within the black rectangle, at 80 ms after stimulus onset and is centered at 36 Hz. The Quantification of the ITPC in the gamma range (20–60 Hz) within the black rectangle yields a significant difference between the gamma coherence of visual evoked responses (timepoints = 900,  $p < 0.000001$ , Mann–Whitney  $U$ -test). Individual difference between the ITPC under both doses of isoflurane and propofol (**B**). Yellow colors represent higher ITPC under isoflurane while dark blue colors represent higher ITPC under propofol. The Quantification of the ITPC in the gamma range (20–60 Hz) within the black rectangle yields a significant difference between the gamma coherence of visual evoked responses.

Cardin, 2016; Friston and Buzsáki, 2016; Sohal, 2016; Traub et al., 2016). Critical for both of these mechanisms is the shape of the IPSPs produced by fast spiking GABAergic interneurons. Propofol allosterically potentiates GABA signaling through the GABA<sub>A</sub> receptor (Hales and Lambert, 1991; Krasowski et al., 1998; Yip et al., 2013; Weiser et al., 2015). Therefore, under propofol, the duration of the IPSPs may be prolonged (Bai et al., 1999). This may be why we see a decrease in total gamma power under propofol, and in some animals (for example, shown in Figure 6), we see a shift to lower evoked gamma power. However, one caveat to this hypothesis is that isoflurane also is a positive allosteric modulator at synaptic and extrasynaptic GABA<sub>A</sub> receptors (Krasowski et al., 1998; Wang, 2009; Garcia et al., 2010). Isoflurane suppresses GABAergic

IPSPs at lower concentrations (Jones and Harrison, 1993; Pearce, 1996; Banks and Pearce, 1999) but, in a hippocampal slice preparation concentrations similar to that used in this study increased the amplitude and duration of GABAergic IPSPs (Miu and Puil, 1989). Yet, other studies in amygdalar slices suggest that isoflurane prolongs GABA<sub>A</sub> mediated currents without effectively increasing their amplitude (Ranft et al., 2004). Both isoflurane and propofol are highly promiscuous drugs that have significant interactions with a host of membrane proteins (Eckenhoff, 2002; Weiser et al., 2015; Tang and Eckenhoff, 2018). Thus, in the absence of detailed biophysical model, which would include actions of both propofol and isoflurane at the plurality of their molecular targets, it is difficult to attribute the differences in the visually induced gamma



**TABLE 1 |** Difference in evoked gamma coherence across mice (all *p*-values are Bonferroni corrected unless otherwise specified).

Comparison	Mann-Whitney <i>U</i> -test <i>p</i> -value for ITPC difference
Pooled isoflurane – pooled propofol	<0.000001
Low dose isoflurane – high dose isoflurane	Uncorrected <i>p</i> -value = 0.0611
Low dose propofol – high dose propofol	<0.000001
Low dose isoflurane – low dose propofol	<0.000001
High dose isoflurane – high dose propofol	<0.000001

**TABLE 2 |** Individual mouse differences in evoked gamma coherence from isoflurane – propofol (all *p*-values are Bonferroni corrected unless otherwise specified).

Mouse	Mann-Whitney <i>U</i> -test <i>p</i> -value for ITPC difference
1	0.336
2	<0.000001
3	<0.000001
4	<0.000001
5	<0.000001
6	<0.000001
7	<0.000001

oscillations to specific differences in the molecular mechanisms of action of either anesthetic. Perhaps, most surprising is the difference in the visual-evoked responses under deep anesthesia characterized by burst suppression. *In vivo* extracellular and intracellular recordings in cats suggest that burst suppression induced with either propofol (Kroeger and Amzica, 2007) or with isoflurane (Kroeger and Amzica, 2007; Amzica, 2009; Ferron et al., 2009) is associated with hyper-excitability of the cortex. Our observations concur with that of Amzica and colleagues – visual stimuli presented under both isoflurane-induced and propofol-induced burst suppression during the suppression period of the ECoG could on occasion trigger bursts (Kroeger and Amzica, 2007; Amzica, 2009; Ferron et al., 2009). Yet, our results suggest that burst suppression induced with isoflurane allows visual stimuli to entrain the phase of the gamma oscillations while burst suppression induced with propofol does not.

Exactly how the laminar structure of the primary visual cortex generates spontaneous and induced gamma rhythms in V1 is also unknown. There is some evidence that the granular layer of V1 is more resistant spectral composition changes under a mixture of isoflurane and xylazine as compared to higher order brain regions in ferrets (Sellers et al., 2013). Moreover, mice under isoflurane/xylazine, tend to have spontaneous gamma waves that begin in all layers simultaneously, however, visual evoked gamma oscillations begin in granular and supra-granular layers (Welle and Contreras, 2015). This may be because isoflurane increases the number of excitatory and inhibitory cells that are recruited to participate in synchronous responses, as seen in Noda and Takahashi (2015). This would also correspond well to findings seen in ferrets which show that visual evoked multiunit activity have a longer duration when

animals are anesthetized with isoflurane/xylazine compared to in the awake state. Interestingly, just because gamma oscillations are seen in the superficial layers does not necessarily mean that there is in fact gamma oscillations in the deeper layers, or that there is strong synchrony between spiking activity and phase of the gamma oscillation (Sellers et al., 2015; Welle and Contreras, 2015).

Visual-evoked gamma power and frequency has also been shown to increase with arousal and locomotion in mice (Niell and Stryker, 2010; Polack et al., 2013; Fu et al., 2014). While we can rule out locomotor effects given our anesthetized preps, the activation of arousal circuitry may be different under these two anesthetics. Indeed, when mice have larger pupillary diameters, indicating increased arousal (Aston-Jones and Cohen, 2005; Gilzenrat et al., 2010; Eldar et al., 2013), visual evoked responses have large signal to noise ratio, and exhibit two peaks of visual evoked gamma power centered around 75 and 30 Hz (Vinck et al., 2015). These effects may be caused by neuromodulation from sleep and arousal systems. For example, the mix of cholinergic tone and noradrenergic input has been shown to maintain high signal to noise seen in V1 neurons during arousal and locomotion in mice (Polack et al., 2013). Moreover, cholinergic projections from the basal forebrain have been shown to increase visual evoked gamma oscillations in mice (Pinto et al., 2013; Lee et al., 2014). Our results thus may indicate that propofol may depress arousal circuitry involving cholinergic and noradrenergic input more than isoflurane. Thus, understanding how anesthetics alter sensory responses may help us formulate hypotheses about the mechanisms by which anesthetics affect the sleep and arousal system to differentially produce unconsciousness.

## Functional Implications

The precise functional implications of visual-evoked gamma oscillations are currently unknown. Some evidence suggests that the sensory evoked gamma oscillations modulate the firing of sensory neurons and increase the efficiency of sensory encoding (Womelsdorf et al., 2006; Cardin et al., 2009; Womelsdorf et al., 2012). This may occur through decreasing noise by increasing inhibitory drive, or increasing signal by entraining sensory evoked firing to a specific phase of the gamma oscillation (Fries et al., 2001; Cardin et al., 2009; Sohal, 2016). It is important to note that we not only observe visual evoked gamma oscillations in the single trial data, and increase in gamma power, but also that this increase in gamma power is consistent in phase from trial to trial under isoflurane. This implies that the neural processes leading to visual evoked responses occur in a stereotyped fashion under isoflurane compared to propofol anesthesia. A possible mechanism leading to such a phase resetting effect would be strong synchrony in visual cortex neuronal firing under isoflurane compared to propofol. Indeed, in ferrets given isoflurane and xylazine, visual stimuli increase ITPC in V1, however, when the same animals are awake there is an increase in ITPC both in V1 and in the PFC, thereby indicating that such phase coherence may be important for functional connectivity between different regions of the brain (Sellers et al., 2015). However, it is not yet clear if this increase



in phase synchrony across the brain will necessarily be able to provide more information for encoding stimulus attributes. For example, there is an increase in auditory evoked gamma coherence between different areas of rat auditory cortex and belt in rats under isoflurane anesthesia as compared to awake rats (Noda and Takahashi, 2015).

## Limitations

While visual-evoked gamma oscillations have been shown in both awake animals and in anesthetized animals, how these oscillations are associated with perception is beyond the scope of this presented research. To understand how the integration of brain state and the visual-evoked gamma oscillations affects perception, one must create a behavioral paradigm in which animals report their response to visual stimuli.

Another limitation is that all mice were induced with isoflurane and measured under the two doses of isoflurane before they were given propofol. Therefore, under the propofol delivery, there may be a slight mixing effect with isoflurane. This paradigm was chosen because in the acute setting, our induction and insertion of the jugular venous catheter is best done with isoflurane. This is because isoflurane has fast on – off kinetics and is much easier to titrate during the long, invasive surgeries of jugular cannulation and craniotomy. To induce and maintain an animal with only propofol, one would need to chronically insert the jugular catheter, allow the animal to recover, and then induce the animal with propofol for neurosurgery and beginning the experiment. Moreover, the animal may need a relatively high dose of propofol for induction, which may take a long time to wash out to maintain a steady state low dose for visual stimulation. We attempted to correct for the amount of isoflurane present within the brain of animals by starting propofol delivery 45 min before recording visual evoked potentials with isoflurane off. Previous results by Yanagisawa et al. (2008) and Friedman et al. (2010) show that the brain has at most trace levels of isoflurane after 15 min after isoflurane is turned off. Therefore, after 45 min of isoflurane off and propofol infusion, there should be virtually no isoflurane within the mouse brain. We also monitored propofol washout kinetics using the model that was fit to reflect the elimination time constant from the brain (Shortal et al., 2018). Thus, while there is a trace amount of propofol in the brain during the re-exposure to isoflurane, the albeit incomplete, recovery of responses first observed under isoflurane suggest that the order of drug administration is not likely to be a major contributor to our findings.

Finally, through employing high density, multichannel recording methods, including ECoG as we have used here, we can start asking questions about how visual evoked activity that initiates in V1 propagate to other visual areas. Answering such questions requires extensive analysis of the dynamically changing correlation structure of the spontaneous and evoked activity across electrodes. Moreover, concluding how anesthesia affect the propagation of such signals will require further parameterization of the visual stimulus characteristics as well as brain state with different anesthetics.

## CONCLUSION

Here, we show that even when the spontaneous activity of the brain shows similar spectral features, i.e., delta power or burst suppression, visual evoked activity is better correlated with the anesthetic drug rather than with the ensuing spectral state of the brain. Therefore, the canonical methods of defining the brain state with spontaneous spectral activity are not complete. Recently there has been a resurgence in efforts to define the state of the brain as induced by different anesthetic agents. One promising approach uses phase based functional connectivity measures to determine the flow of information from one brain region to the other (Lee et al., 2013; Kim et al., 2016). Another interesting set of methods include observing changes in the dynamics of the correlations in the signals from multiple electrodes (Solovey et al., 2015). Critical to the success of these methods is the understanding of how these models cope with sensory perturbations. Therefore, in addition to studying spontaneous activity under anesthesia and wakefulness, it will also be important to observe sensory evoked activity.

## MATERIALS AND METHODS

### Animals

All experiments in this study were approved by Institutional Animal Care and Use Committee at the University of Pennsylvania and were conducted in accordance with the National Institutes of Health guidelines. All experiments were performed using adult (12–28 weeks old, 20–30 g) male and female C57BL/6 mice (Jackson Laboratories). Mice were housed under a reverse 12:12 h, light: dark cycle. Mice were provided with food and water *ad libitum*. A total of 11 mice were recorded from in this study. Inclusion criteria for mice included the following: (1) presence of spontaneous activity that was not characterized as burst suppression at lower drug concentrations (2) presence of VEPs at each dose of each anesthetic (as defined by the absolute value of the average LFP response exceeding five standard deviations of pre-stimulus data within 200 ms after stimulus presentation). With this inclusion criteria, we present data from 7 mice in this study.

### Surgery

All surgery was performed under aseptic conditions. Each animal was weighed (20–30 g) immediately prior to surgery to adequately dose propofol delivery. Prior to surgery, 2 mg/kg of dexamethasone was given subcutaneously (SQ) to reduce brain edema. Animals were induced with 2.5% isoflurane in oxygen (flow rate 500 ml/min), and maintained at 1.5% isoflurane for the remainder of the surgery. Core-body temperature was maintained at  $37 (\pm 0.5)^{\circ}\text{C}$  using a temperature controller with core-body temperature monitoring (TC- 1000 Temperature Controller, CWE, Incorporated, Ardmore, PA, United States). First, a jugular venous catheter was placed prior to neurosurgery to allow for Targeted Controlled Infusion (TCI) propofol infusion as described in Shortal et al. (2018). Animals were then secured into a Kopf stereotaxic frame. 0.25 ml of 2% lidocaine gel was

applied to the scalp to provide a local nerve blockade during surgery. The scalp was then incised and retracted, permitting maximum exposure of the surface of the skull. The bone was cleaned and dried before a craniotomy was performed using a dental drill. One large craniotomy was drilled over the left hemisphere (+1 mm to -5 mm AP, +0.25 mm to +6 mm ML of bregma), and a small reference screw was secured in the right skull bone (+1 mm AP, +1 mm ML of bregma). A 64-electrode surface grid (Neuronexus: E64-500-20-60) was positioned over the dura to obtain ECoG signal. Mineral oil was applied on top of the ECoG grid every 20 min to preserve the health of the underlying dura and brain. Animals were scarified the same day immediately after the final visual recording session.

## Visual Stimulation

Visual stimuli consisted of a brief 10 ms flash of a bright green LED light (0.43 mW/cm<sup>2</sup>), placed 2 cm away from the mouse's right eye. The flash covered 100% of the mouse's visual field. Hundred flashes were given under each anesthetic dose step.

## Anesthetic Delivery Protocol

After the jugular catheter was placed and the craniotomy was completed, the doses and anesthetics were parametrically altered. The animal was first given two doses of isoflurane (Terrell Isoflurane, Novaplus): 1.2% (high dose isoflurane) and 0.6% (low dose isoflurane). Animals always received the higher isoflurane concentration before receiving the lower dose. The brain was allowed 15 min to equilibrate after the amount of isoflurane was changed (Friedman et al., 2010). After the two isoflurane doses, the TCI propofol (1000 mg per 100 ml Diprivan, Fresenius Kabi, United States) was administered using the model described in Shortal et al. (2018). The two target concentrations were 20 µg/g brain (low dose propofol), and 35 µg/g brain (high dose propofol). One out of seven of the mice was producing spontaneous movement at the 20 µg/g brain dose of propofol. Therefore, this mouse received 30 µg/g brain for low dose propofol, and 40 µg/g brain for high dose propofol to maintain slow waves and burst suppression activity, respectively. Due to the slow rate of excretion of propofol, the lower concentration of propofol was administered before the higher dose. Due to the fast rate of onset of propofol, the equilibration time between propofol changes was 8 min. Animals were not intubated, nor was an arterial catheter placed for pCO<sub>2</sub> or blood pressure measurement.

In four of the seven animals, re-exposure doses of isoflurane were given in order to control for possible brain desiccation effects or impedance changes from keeping a large craniotomy open for a long period of time. In these animals, after achieving the 35 µg/g brain concentration of propofol, the propofol infusion was turned off and 1.2% isoflurane was administered for 1 h in order to allow propofol to wash out. Propofol washout was monitored using the same TCI model used for propofol delivery. This model estimates the amount of propofol remaining in the brain parenchyma. Following this wash out period, visual stimuli were again given under 1.2 and 0.6% re-exposure isoflurane.

## Electrophysiology and Preprocessing

In six mice, signals were amplified via a Neuralynx headstage (HS36), digitized through Cheetah 64 acquisition system (Neuralynx, ERP-27, Lynx-8), and collected at a rate of 3030.3 samples/s. In one mouse, signals were amplified via an Intan headstage (Intan, RHD2132), digitized through Omniplex acquisition system (Plexon, Omniplex), and collected at a rate of 30,000 samples/s.

LFP data collected with Neuralynx was filtered online using a proprietary FIR filter between 0.1 and 325 Hz. LFP data collected with Plexon was filtered offline using a custom-built FIR filter between 0.1 Hz and 325 Hz, with the MATLAB functions, *firls.m* and *filtfilt.m*. Offline, both data sets were decimated to 1000 samples/s, noise channels were manually removed and trials with excess motion artifact or saturated data was rejected before the mean was subtracted from the data. All data analysis was completed using custom built Matlab (Mathworks) code unless otherwise stated.

## Selection of Electrode Over Primary Visual Cortex (V1)

The latency of onset of the VEP was calculated for each electrode in the array as the time point at which if their post-stimulus average exceeds three standard deviations above the pre-stimulus baseline for three consecutive time points. The electrode which had the lowest latency of onset was denoted as V1. The amplitude of the VEP was calculated by determining the RMS of the first 350 ms of post-stimulus data. This was determined for both raw voltage signals and for voltage signals normalized to 500 ms of pre-stimulus data and expressed as a z-score.

$$RMS = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \dots + x_n^2)} \quad (1)$$

Where  $x$  is the voltage in the post-stimulus average, and  $n$  is the time point in ms. Duration of the VEP was defined as the first time point in which the post-stimulus data returns to within two standard deviations of the pre-stimulus data for 20 consecutive time points.

## Quantification of Reliability in Time Domain

To assess reliability of the LFP evoked response to the visual stimulus in time domain, a similar method was used to that of Tiesinga et al. (2003) and Kumbhani et al. (2007). First, pairwise correlations of single trial evoked potentials were computed over the first 350 ms of post-stimulus activity for each dose of each anesthetic. These correlations were then averaged together to compute reliability. Correlation was computed using the Matlab *corr.m* function.

$$Reliability = \frac{\sum_{i=1}^{trials} \sum_{j=1+i}^{trials} (LFP_i - \mu_i)(LFP_j - \mu_j) / \sigma_i * \sigma_j}{(trials^2 - trials) / 2} \quad (2)$$

Where  $LFP_i$  is the evoked response during the  $i$ -th trial,  $\mu_i$  and  $\sigma_i$  are the mean and standard deviation of the single trial response.

### Quantification of Suppression Ratio (SR)

To assess determine which epochs of the LFP were suppressed, both a frequency based metric and an amplitude based metric was applied to the data. First, spectrograms were calculated using multi-taper spectral analysis by applying the MATLAB function, *swTfSpecAnalog.m*, written by Dr. Andrew Hudson. Spectral analysis was performed from 2 to 500 Hz with a set of 5 Slepian tapers, over a window size of 500 ms, with 80% overlap. The total power was then calculated for each window for frequencies between 2 and 100 Hz. The resultant total power time series of a burst suppression data set (high dose isoflurane or high dose propofol) was then subjected to k-means clustering to find 2 centroids- one that would correspond to bursts, and the other to suppressions. From this, the maximum threshold for classifying suppression based on total power was calculated for each mouse. Concurrently, RMS was also calculated over the LFP data in 500 ms windows with 80% overlap. A manual maximum RMS threshold by eye was selected for each mouse. Time windows were classified as suppression as long as the total power and RMS were below their respective thresholds. The suppression ratio (SR) was calculated by number of time windows with suppression divided by the total number of time windows.

### Spectral Analysis of Spontaneous LFP

One minute of unstimulated LFP was extracted from each mouse under each concentration of isoflurane and propofol. Re-exposure isoflurane baseline data was excluded since excess propofol may have remained within the brain given the propofol's slow excretion rate. Power spectral density of each segment was calculated using multi-taper spectral analysis by applying the MATLAB function, *mtpsd.m*, written by Dr. Andrew Hudson. Spectral analysis was performed from 0.05 to 100 Hz with a set of 20 Slepian tapers. All power spectra were normalized to total power. The average power spectra and 95% confidence intervals for each concentration of isoflurane and propofol was calculated using the ensuing normalized power spectra. Spectra in **Figure 2B** are shown on a log-log scale.

### Wavelet Analysis

Power, phase, and frequency information was extracted by convolving single trial data with a set of Morlet wavelets (0.1–150 Hz, with a step-width 0.25 Hz and normalized amplitude), generated with using continuous wavelet transform, *contwt.m*, in MATLAB, written by Christopher Torrence and Gilbert Compo<sup>1</sup> (Torrence and Compo, 1998). The ensuing power spectrograms of single trial data were averaged within each mouse, under each concentration of isoflurane and propofol. The average spectrograms were then normalized to 300 ms of baseline data. Spectrograms shown in **Figure 6** are shown with frequency in log space.

<sup>1</sup><http://paos.colorado.edu/research/wavelets/>

### Inter-trial Phase Coherence (ITPC) Analysis

Inter-trial phase coherence (ITPC) was used to quantify the amount of phase synchrony between trials at each frequency. ITPC over V1 in each mouse under each concentration of anesthetics. First, angle vectors were extracted from the wavelet coefficients at each time point and each frequency by applying Euler's formula and setting the single trial vector length to 1. ITPC was then calculated by taking the mean length of the angle vector across trials. ITPC at each timepoint and frequency of all mice were averaged separately under isoflurane and propofol. **Figure 7** shows the difference in the average visual-evoked ITPC between mice under isoflurane and propofol.

### Statistical Analysis

Statistics presented in **Figure 3** were presented using Kruskal–Wallis–Mann  $U$ -test for group comparison and a Mann–Whitney  $U$ -test for concentration specific effects.  $P$ -values were Bonferroni corrected for multiple comparisons among four groups. **Figure 7** were performed using Mann–Whitney  $U$ -tests.  $P$ -values were made more stringent using a Bonferroni correction for multiple comparisons among 900 time points.

### ETHICS STATEMENT

All experiments in this study were approved by Institutional Animal Care and Use Committee at the University of Pennsylvania and were conducted in accordance with the National Institutes of Health guidelines. All experiments were performed using adult (12–28 weeks old, 20–30 g) male and female C57BL/6 mice (Jackson Laboratories). Mice were housed under a reverse 12:12 h, light: dark cycle. Mice were provided with food and water *ad libitum*.

### AUTHOR CONTRIBUTIONS

AA, CB, DC, MK, and AP: conceptualization. AA, BS, DC, and AP: data curation. AA, CB, and AP: formal analysis. DC, MK, and AP: funding acquisition. AA and AP: writing – original draft. AA, CB, BS, MK, DC, and AP: writing – review and editing.

### FUNDING

This research was supported through the Translational Neuroscience Initiative from the Penn Medicine Translational Neuroscience Center (PMTNC), RO1 GM088156 (MK), RO1 GM124023 (AP), T32 EY007035 (AA).

### ACKNOWLEDGMENTS

We wish to thank Dr. Andrew Hudson for sharing spectral analysis MATLAB code. We also thank Dr. Andrew McKinstry-Wu, Sarah Reitz, Andrzej Z. Wasilczuk for helpful discussions.



## REFERENCES

- Akeju, O., Song, A. H., Hamilos, A. E., Pavone, K. J., Flores, F. J., Brown, E. N., et al. (2016). Electroencephalogram signatures of ketamine anesthesia-induced unconsciousness. *Clin. Neurophysiol.* 127, 2414–2422. doi: 10.1016/j.clinph.2016.03.005
- Amzica, F. (2009). Basic physiology of burst-suppression. *Epilepsia* 50(Suppl. 12), 38–39. doi: 10.1111/j.1528-1167.2009.02345.x
- Arena, A., Lamanna, J., Gemma, M., Ripamonti, M., Ravasio, G., Zimarino, V., et al. (2017). Linear transformation of the encoding mechanism for light intensity underlies the paradoxical enhancement of cortical visual responses by sevoflurane. *J. Physiol.* 595, 321–339. doi: 10.1113/JP.272215
- Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709
- Bai, D., Pennefather, P. S., MacDonald, J. F., and Orser, B. A. (1999). The general anesthetic propofol slows deactivation and desensitization of GABA A receptors. *J. Neurosci.* 19, 10635–10646. doi: 10.1523/jneurosci.19-24-10635.1999
- Banks, M. I., and Pearce, R. A. (1999). Dual actions of volatile anesthetics on GABAergic IPSCs: dissociation of blocking and prolonging effects. *Anesthesiology* 90, 120–134. doi: 10.1097/0000542-199901000-00018
- Cardin, J. A. (2016). Snapshots of the brain in action: local circuit operations through the lens of  $\gamma$  oscillations. *J. Neurosci.* 36, 10496–10504. doi: 10.1523/jneurosci.1021-16.2016
- Cardin, J. A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., et al. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* 459, 663–667. doi: 10.1038/nature08002
- Ching, S., Cimenser, A., Purdon, P. L., Brown, E. N., and Kopell, N. J. (2010). Thalamocortical model for a propofol-induced  $\gamma$ -rhythm associated with loss of consciousness. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22665–22670. doi: 10.1073/pnas.1017069108
- Destexhe, A., Contreras, D., and Steriade, M. (1999). Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during. *J. Neurosci.* 19, 4595–4608. doi: 10.1523/jneurosci.19-11-04595.1999
- Eckenhoff, M. F. (2002). Multiple specific binding targets for inhaled anesthetics in the mammalian brain. *J. Pharmacol. Exp. Ther.* 300, 172–179. doi: 10.1124/jpet.300.1.172
- Eckenhoff, M. F., and Eckenhoff, R. G. (1998). Quantitative autoradiography of halothane binding in rat brain. *J. Pharmacol. Exp. Ther.* 285, 371–376.
- Eldar, E., Cohen, J. D., and Niv, Y. (2013). The effects of neural gain on attention and learning. *Nat. Neurosci.* 16, 1146–1153. doi: 10.1038/nn.3428
- Ermentrout, G. B., and Kleinfeld, D. (2001). Traveling electrical waves in cortex: insights from phase dynamics and speculation on a computational role. *Neuron* 29, 33–44. doi: 10.1016/S0896-6273(01)00178-7
- Ferron, J.-F., Kroeger, D., Chever, O., and Amzica, F. (2009). Cortical Inhibition during burst Suppression Induced with isoflurane anesthesia. *J. Neurosci.* 29, 9850–9860. doi: 10.1523/JNEUROSCI.5176-08.2009
- Franks, N. P. (2008). General anaesthesia: from molecular targets to neuronal pathways of sleep and arousal. *Nat. Rev. Neurosci.* 9, 370–386. doi: 10.1038/nrn2372
- Friedman, E. B., Sun, Y., Moore, J. T., Hung, H. T., Meng, Q. C., Perera, P., et al. (2010). A conserved behavioral state barrier impedes transitions between anesthetic-induced unconsciousness and wakefulness: evidence for neural inertia. *PLoS One* 5:e11903. doi: 10.1371/journal.pone.0011903
- Fries, P., Neuenschwander, S., Engel, A. K., Goebel, R., and Singer, W. (2001). Rapid feature selective neuronal synchronization through correlated latency shifting. *Nat. Neurosci.* 4, 194–200. doi: 10.1038/84032
- Friston, K., and Buzsáki, G. (2016). The functional anatomy of time: what and when in the brain. *Trends Cogn. Sci.* 20, 500–511. doi: 10.1016/j.tics.2016.05.001
- Fu, Y., Tucciarone, J. M., Espinosa, J. S., Sheng, N., Darcy, D. P., Nicoll, R. A., et al. (2014). A cortical circuit for gain control by behavioral state. *Cell* 156, 1139–1152. doi: 10.1016/j.cell.2014.01.050
- Garcia, P. S., Kolesky, S. E., and Jenkins, A. (2010). General anesthetic actions on GABA A receptors. *Curr. Neuropharmacol.* 8, 2–9. doi: 10.2174/157015910790909502
- Gibbs, F., Gibbs, E., and Lennox, W. (1937). Effect on the electro-encephalogram of certain drugs which influence nervous activity. *Arch. Intern. Med.* 60, 154–166.
- Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., and Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn. Affect. Behav. Neurosci.* 10, 252–269. doi: 10.3758/CABN.10.2.252
- Hales, T. G., and Lambert, J. J. (1991). The actions of propofol on inhibitory amino acid receptors of bovine adrenomedullary chromaffin cells and rodent central neurones. *Br. J. Pharmacol.* 104, 619–628. doi: 10.1111/j.1476-5381.1991.tb12479.x
- Hall, A. C., Lieb, W. R., and Franks, N. P. (1994). Stereoselective and non-stereoselective actions of isoflurane on the GABAA receptor. *Br. J. Pharmacol.* 112, 906–910. doi: 10.1111/j.1476-5381.1994.tb13166.x
- Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154. doi: 10.1113/jphysiol.1962.sp006837
- Imas, O. A., Ropella, K. M., Ward, B. D., Wood, J. D., and Hudetz, A. G. (2005). Volatile anesthetics enhance flash-induced gamma oscillations in rat visual cortex. *Anesthesiology* 102, 937–947. doi: 10.1097/0000542-200505000-00012
- Jones, M. V., and Harrison, N. L. (1993). Effects of volatile anesthetics on the kinetics of inhibitory postsynaptic currents in cultured rat hippocampal neurons. *J. Neurophysiol.* 70, 1339–1349. doi: 10.1152/jn.1993.70.4.1339
- Jurd, R., Arras, M., Lambert, S., Drexler, B., Siegwart, R., Crestani, F., et al. (2003). General anesthetic actions in vivo strongly attenuated by a point mutation in the GABA A receptor  $\beta 3$  subunit. *FASEB J.* 17, 250–252. doi: 10.1096/fj.02-0611fj
- Kim, M., Mashour, G. A., Moraes, S., Vanini, G., Tarnal, V., Janke, E., et al. (2016). Functional and topological conditions for explosive synchronization develop in human brain networks with the onset of anesthetic-induced unconsciousness. *Front. Comput. Neurosci.* 10:1. doi: 10.3389/fncom.2016.00001
- Kopell, N., Ermentrout, G. B., Whittington, M. A., and Traub, R. D. (2000). Gamma rhythms and beta rhythms have different synchronization properties. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1867–1872. doi: 10.1073/pnas.97.4.1867
- Krasowski, M. D., Koltchine, V. V., Rick, C. E., Ye, Q., Finn, S. E., and Harrison, N. L. (1998). Propofol and other intravenous anesthetics have sites of action on the gamma-aminobutyric acid type A receptor distinct from that for isoflurane. *Mol. Pharmacol.* 53, 530–538. doi: 10.1124/mol.53.3.530
- Kroeger, D., and Amzica, F. (2007). Hypersensitivity of the anesthesia-induced comatose brain. *J. Neurosci.* 27, 10597–10607. doi: 10.1523/jneurosci.3440-07.2007
- Kumbhani, R. D., Nolt, M. J., and Palmer, L. A. (2007). Precision, reliability, and information-theoretic analysis of visual thalamocortical neurons. *J. Neurophysiol.* 98, 2647–2663. doi: 10.1152/jn.00900.2006
- Lee, A. M., Hoy, J. L., Bonci, A., Wilbrecht, L., Stryker, M. P., and Niell, C. M. (2014). Identification of a brainstem circuit regulating visual cortical state in parallel with locomotion. *Neuron* 83, 455–466. doi: 10.1016/j.neuron.2014.06.031
- Lee, U., Ku, S., Noh, G., Baek, S., Choi, B., and Mashour, G. A. (2013). Disruption of frontal-parietal communication by ketamine, propofol, and sevoflurane. *Anesthesiology* 118, 1264–1275. doi: 10.1097/ALN.0b013e31829103f5
- Leslie, K., Chan, M. T. V., Myles, P. S., Forbes, A., and McCulloch, T. J. (2010). Posttraumatic stress disorder in aware patients from the B-aware trial. *Anesth. Analg.* 110, 823–828. doi: 10.1213/ANE.0b013e3181b8b6ca
- Lydic, R., and Baghdoyan, H. A. (2005). Sleep, anesthesiology, and the neurobiology of arousal state control. *Anesthesiology* 103, 1268–1295. doi: 10.1097/0000542-200512010-00024
- Maksimow, A., Särkelä, M., Långsjö, J. W., Salmi, E., Kaisti, K. K., Yli-Hankala, A., et al. (2006). Increase in high frequency EEG activity explains the poor performance of EEG spectral entropy monitor during S-ketamine anesthesia. *Clin. Neurophysiol.* 117, 1660–1668. doi: 10.1016/j.clinph.2006.05.011
- Mashour, G. A. (2006). Integrating the science of consciousness and anesthesia. *Anesth. Analg.* 10, 975–982. doi: 10.1213/01.ane.0000232442.69757.4a
- Mashour, G. A., Orser, B. A., and Avidan, M. S. (2011). Intraoperative awareness: from neurobiology to clinical practice. *Anesthesiology* 114, 1218–1233. doi: 10.1097/ALN.0b013e31820fc9b6
- Mashour, G. A., Shanks, A., Tremper, K. K., Kheterpal, S., Turner, C. R., Ramachandran, S. K., et al. (2012). Prevention of intraoperative awareness with explicit recall in an unselected surgical population. *Anesthesiology* 117, 717–725. doi: 10.1097/aln.0b013e31826904a6

- Miu, P., and Puil, E. (1989). Isoflurane-induced impairment of synaptic transmission in hippocampal neurons. *Exp. Brain Res.* 75, 354–360.
- Mountcastle, V. B. (1957). Modality and topographic properties neurons of cat's somatic sensory cortex. *J. Physiol.* 20, 408–434. doi: 10.1152/jn.1957.20.4.408
- Niell, C. M., and Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* 65, 472–479. doi: 10.1016/j.neuron.2010.01.033
- Noda, T., and Takahashi, H. (2015). Anesthetic effects of isoflurane on the tonotopic map and neuronal population activity in the rat auditory cortex. *Eur. J. Neurosci.* 42, 2298–2311. doi: 10.1111/ejn.13007
- Pearce, R. A. (1996). Volatile anaesthetic enhancement of paired-pulse depression investigated in the rat hippocampus in vitro. *J. Physiol.* 492(Pt 3), 823–840. doi: 10.1113/jphysiol.1996.sp021349
- Pinto, L., Goard, M. J., Estandian, D., Xu, M., Kwan, A. C., Lee, S. H., et al. (2013). Fast modulation of visual perception by basal forebrain cholinergic neurons. *Nat. Neurosci.* 16, 1857–1863. doi: 10.1038/nn.3552
- Polack, P.-O. O., Friedman, J., and Golshani, P. (2013). Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *Nat. Neurosci.* 16, 1331–1339. doi: 10.1038/nn.3464
- Purdon, P. L., Pierce, E. T., Mukamel, E. A., Prerau, M. J., Walsh, J. L., Foon, K., et al. (2013). Electroencephalogram signatures of loss and recovery of consciousness from propofol. *Proc. Natl. Acad. Sci. U.S.A.* 110, E1142–E1151. doi: 10.1073/pnas.1221180110
- Ranft, A., Kurz, J., Deuringer, M., Haseneder, R., Dodt, H. U., Zieglgänsberger, W., et al. (2004). Isoflurane modulates glutamatergic and GABAergic neurotransmission in the amygdala. *Eur. J. Neurosci.* 20, 1276–1280. doi: 10.1111/j.1460-9568.2004.03603.x
- Reinhold, K., Lien, A. D., and Scanziani, M. (2015). Distinct recurrent versus afferent dynamics in cortical visual processing. *Nat. Neurosci.* 18, 1789–1797. doi: 10.1038/nn.4153
- Russell, I. F. (1989). Conscious awareness during general anaesthesia: relevance of autonomic signs and isolated arm movements as guides to depth of anaesthesia. *Baillieres. Clin. Anaesthesiol.* 3, 511–532. doi: 10.1016/s0950-3501(89)80016-9
- Sanders, R. (2016). Incidence of connected consciousness after tracheal intubation. *Anesthesiology* 126, 214–222. doi: 10.1097/ALN.0000000000001479
- Sanders, R. D., Tononi, G., Laureys, S., and Sleight, J. W. (2012). Unresponsiveness  $\neq$  unconsciousness. *Anesthesiology* 116, 946–959. doi: 10.1097/ALN.0b013e318249d0a7
- Schneider Gerhard, M. D., Hollweck Regina, M. S., Ningler Michael, M. S., Stockmanns Gudrun, P. D., and Kochs Eberhard, F. (2005). Detection of consciousness by electroencephalogram and auditory evoked potentials. *Anesthesiology* 103, 934–943. doi: 10.1097/00005542-200511000-00006
- Sellers, K. K., Bennett, D. V., Hutt, A., and Fröhlich, F. (2013). Anesthesia differentially modulates spontaneous network dynamics by cortical area and layer. *J. Neurophysiol.* 110, 2739–2751. doi: 10.1152/jn.00404.2013
- Sellers, K. K., Bennett, D. V., Hutt, A., Williams, J. H., and Fröhlich, F. (2015). Awake vs. anesthetized: layer-specific sensory processing in visual cortex and functional connectivity between cortical areas. *J. Neurophysiol.* 113, 3798–3815. doi: 10.1152/jn.00923.2014
- Shortal, B. P., Reitz, S. L., Aggarwal, A., Meng, Q. C., McKinstry-Wu, A. R., Kelz, M. B., et al. (2018). Development and validation of brain target controlled infusion of propofol in mice. *PLoS One* 13:e0194949. doi: 10.1371/journal.pone.0194949
- Sohal, V. S. (2016). How close are we to understanding what (if anything)  $\gamma$  oscillations do in cortical circuits? *J. Neurosci.* 36, 10489–10495. doi: 10.1523/jneurosci.0990-16.2016
- Solovey, G., Alonso, L. M., Yanagawa, T., Fujii, N., Magnasco, M. O., Cecchi, G. A., et al. (2015). Loss of consciousness is associated with stabilization of cortical activity. *J. Neurosci.* 35, 10866–10877. doi: 10.1523/JNEUROSCI.4895-14.2015
- Sonner, J. M., Werner, D. F., Elsen, F. P., Xing, Y., Liao, M., Harris, R. A., et al. (2007). Effect of isoflurane and other potent inhaled anesthetics on minimum alveolar concentration, learning, and the righting reflex in mice engineered to express  $\alpha 1$   $\gamma$ -aminobutyric acid type A receptors unresponsive to isoflurane. *Anesthesiology* 106, 107–113. doi: 10.1097/00005542-200701000-00019
- Storchi, R., Bedford, R. A., Martial, F. P., Allen, A. E., Wynne, J., Montemurro, M. A., et al. (2017). Modulation of fast narrowband oscillations in the mouse retina and dLGN according to background light intensity. *Neuron* 93, 299–307. doi: 10.1016/j.neuron.2016.12.027
- Tang, P., and Eckenhoff, R. (2018). Recent progress on the molecular pharmacology of propofol. *F1000Res.* 7:123. doi: 10.12688/f1000research.12502.1
- Tiesinga, P., Whitmer, D., Sejnowski, T. J., Fellous, J. M., and Schreiber, S. (2003). A new correlation-based measure of spike timing reliability. *Neurocomputing* 5, 925–931. doi: 10.1016/s0925-2312(02)00838-x
- Torrence, C., and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* 79, 61–78.
- Traub, R. D., Contreras, D., Cunningham, M. O., Murray, H., Fiona, E. N., Roopun, A., et al. (2016). Single-column thalamocortical network model exhibiting gamma oscillations, sleep spindles, and epileptogenic bursts single-column thalamocortical network model exhibiting gamma oscillations, sleep spindles, and epileptogenic bursts. *J. Neurophysiol.* 93, 2194–2232. doi: 10.1152/jn.00983.2004
- Traub, R. D., Whittington, M. A., Stanford, I. M., and Jefferys, J. G. R. (1996). A mechanism for generation of long-range synchronous fast oscillations in the cortex. *Nature* 383, 621–624. doi: 10.1038/383621a0
- Vinck, M., Batista-Brito, R., Knoblich, U., and Cardin, J. A. (2015). Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron* 86, 740–754. doi: 10.1016/j.neuron.2015.03.028
- Wang, X. (2009). Propofol and isoflurane enhancement of tonic gamma-aminobutyric acid type A current in cardiac vagal neurons in the nucleus ambiguus. *Anesth. Analg.* 108, 142–148. doi: 10.1213/ane.0b013e31818d8b79
- Wang, X. Y., and Buzsáki, G. (1996). Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. *J. Neurosci.* 16, 6402–6413. doi: 10.1523/jneurosci.16-20-06402.1996
- Weiser, B. P., Hall, M. A., Weinbren, N. L., Woll, K. A., Dailey, W. P., Eckenhoff, M. F., et al. (2015). Macroscopic and macromolecular specificity of alkylphenol anesthetics for neuronal substrates. *Sci. Rep.* 5, 1–6. doi: 10.1038/srep09695
- Welle, C. G., and Contreras, D. (2015). Sensory-driven and spontaneous gamma oscillations engage distinct cortical circuitry. *J. Neurophysiol.* 115, 1821–1835. doi: 10.1152/jn.00137.2015
- Whittington, M. A., Traub, R. D., Kopell, N., Ermentrout, B., and Buhl, E. H. (2000). Inhibition-based rhythms: Experimental and mathematical observations on network dynamics. *Int. J. Psychophysiol.* 38, 315–336. doi: 10.1016/s0167-8760(00)00173-2
- Wilson, H. R., and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* 12, 1–24. doi: 10.1016/s0006-3495(72)86068-5
- Womelsdorf, T., Fries, P., Mitra, P. P., and Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature* 439, 733–736. doi: 10.1038/nature04258
- Womelsdorf, T., Lima, B., Vinck, M., Oostenveld, R., Singer, W., Neuenschwander, S., et al. (2012). Orientation selectivity and noise correlation in awake monkey area V1 are modulated by the gamma cycle. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4302–4307. doi: 10.1073/pnas.1114223109
- Yanagisawa, M., Sun, Y., Kelz, M. B., Thornton, M., Moore, J. T., Chen, J., et al. (2008). An essential role for orexins in emergence from general anesthesia. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1309–1314. doi: 10.1073/pnas.0707146105
- Yip, G. M. S., Chen, Z. W., Edge, C. J., Smith, E. H., Dickinson, R., Hohenester, E., et al. (2013). A propofol binding site on mammalian GABA A receptors identified by photolabeling. *Nat. Chem. Biol.* 9, 715–720. doi: 10.1038/nchembio.1340

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Aggarwal, Brennan, Shortal, Contreras, Kelz and Proekt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Evaluating Complexity of Fetal MEG Signals: A Comparison of Different Metrics and Their Applicability

Julia Moser<sup>1\*</sup>, Siouar Bensaid<sup>2</sup>, Eleni Kroupi<sup>3</sup>, Franziska Schleger<sup>1</sup>, Fabrice Wendling<sup>2</sup>, Giulio Ruffini<sup>3</sup> and Hubert Preißl<sup>1</sup>

<sup>1</sup>fMEG Center/Internal Medicine IV/Institute for Diabetes Research and Metabolic Diseases of the Helmholtz Center Munich at the University of Tübingen, Tübingen, Germany, <sup>2</sup>INSERM, U1099, Rennes 1 University, Rennes, France, <sup>3</sup>Starlab SLU, Barcelona, Spain

## OPEN ACCESS

### Edited by:

James W. Grau,  
Texas A&M University, United States

### Reviewed by:

Daniel Abasolo,  
University of Surrey, United Kingdom  
Rathinaswamy Bhavanandhan  
Govindan,  
Children's National Health System,  
United States

### \*Correspondence:

Julia Moser  
julia.moser@student.uni-  
tuebingen.de

**Received:** 30 January 2019

**Accepted:** 06 May 2019

**Published:** 27 May 2019

### Citation:

Moser J, Bensaid S, Kroupi E,  
Schleger F, Wendling F, Ruffini G  
and Preißl H (2019) Evaluating  
Complexity of Fetal MEG Signals:  
A Comparison of Different Metrics  
and Their Applicability.  
*Front. Syst. Neurosci.* 13:23.  
doi: 10.3389/fnsys.2019.00023

In this work, we aim to investigate whether information based metrics of neural activity are a useful tool for the quantification of consciousness before and shortly after birth. Neural activity is measured using fetal magnetoencephalography (fMEG) in human fetuses and neonates. Based on recent theories on consciousness, information-based metrics are established to measure brain complexity and to assess different levels of consciousness. Different metrics (measures of entropy, compressibility and fractality) are, thus, explored in a reference population and their usability is evaluated. For comparative analysis, two fMEG channels were selected: one where brain activity was previously detected and one at least 15 cm away, that represented a control channel. The usability of each metric was evaluated and results from the brain and control channel were compared. Concerning the ease of use with fMEG data, Lempel-Ziv-Complexity (LZC) was evaluated as best, as it is unequivocal and needs low computational effort. The fractality measures have a high number of parameters that need to be adjusted prior to analysis and therefore forfeit comparability, while entropy measures require a higher computational effort and more parameters to adjust compared to LZC. Comparison of a channel with brain activity and a control channel in neonatal recordings showed significant differences in most complexity metrics. This clear difference can be seen as proof of concept for the usability of complexity metrics in fMEG. For fetal data, this comparison produced less clear results which can be related to leftover maternal signals included in the control channel. Further work is necessary to conclusively interpret results from the analysis of fetal recordings. Yet this study shows that complexity metrics can be used for fMEG data on early consciousness and the evaluation gives a guidance for future work. The inconsistency of results from different metrics highlights the challenges of working with complexity metrics as neural correlates of consciousness, as well as the caution one should apply to interpret them.

**Keywords:** fMEG, complexity, LZC, entropy, fractality, neurodevelopment

## INTRODUCTION

Consciousness is known to be one of the characteristics that make humans unique. But when does this aspect of the human mind arise? Is it possible that consciousness already exists before birth? From the 24th week of gestation, a fetus can process sensory stimuli at a cortical level, as thalamocortical connections are already established (Kostović and Judoš, 2010). Long range pyramidal neurons—which are known to be important for conscious processing (Dehaene et al., 1998)—are developed around week 26 (Lagercrantz and Changeux, 2009). Yet, it is difficult to assess the conscious state of a fetus in the mother's womb. Fetal magnetoencephalography (fMEG) is a tool to noninvasively investigate fetal brain activity in the last trimester of pregnancy and in neonates shortly after birth (Preissl et al., 2004). This measurement of fetal/neonatal brain activity makes it possible to investigate neural correlates of consciousness and pursue the question of the debut of consciousness in human life.

During the last decades, work in the field of disorders of consciousness led to an increased interest in neural correlates of consciousness. Neural processes that allow conscious experience need to be highly integrated and differentiated, which are properties that can be measured by complexity of neurological data (Tononi and Edelman, 1998). In nature, complexity of physiology is related to the adaptive capacity of an organism (Costa et al., 2002). This is translated into physiological signals with long-range correlations across various spatio-temporal scales—a behavior that is named self-organization—that indicate the presence of self-invariant and self-similar structures (Pritchard and Duke, 1995). The self-organizational properties of a complex system can be quantified by estimating its dimension (Theiler, 1990), or its ability to compress information (Cover and Thomas, 2012; Ruffini, 2017a).

For a system to be complex, it has to operate on several scales and also show an interplay between those scales (Lutzenberger et al., 1995). This is a property of a so-called chaotic system and can be measured in space and time (Elbert et al., 1994). Typically chaos in space is estimated with the fractal dimension, which is defined as the dimension of a strange attractor towards which a complex system evolves in phase-space (Grassberger and Procaccia, 1983b). Thus, the fractal dimension, namely  $D$ , describes the overall complexity of an object, which can be the geometrical complexity, the space filling property, the roughness of a surface, or the variation of a time series. The fractal dimension  $D$  is defined by the logarithmic ratio of change in detail with change in scale (Di Ieva, 2016). This relates to the distinct characteristic of a fractal, namely the property of self-similarity: i.e., pieces of an object are similar to larger pieces of it as well as to the whole object (Eke et al., 2002). In nature, fractals are usually only statistically self-similar which means that smaller excerpts are not necessarily exact copies of the larger ones, but they are the same on average (Pritchard and Duke, 1995). In contrast to fractal dimension, entropy gives information about the dynamical properties of an attractor and not about its geometrical shape (Rodríguez-Bermúdez and García-Laencina, 2015). Chaos in time, therefore, relates to this stability and

sensitivity to initial conditions (Elbert et al., 1994; Baranger, 2000). Related to those properties of complex systems, several measures were developed to quantify their complexity. The sensitivity to initial conditions can be quantified in terms of the Lyapunov exponent and the Kolmogorov entropy, also known as information dimension (Theiler, 1990). Different entropy measures, as well as the measure of compressibility, can be employed for this. Ruffini (2017a) recently proposed a theory of consciousness that considers the brain an engine that strives to model the world with simplicity, while learning it is a result of exchanging information with it. According to this theory, the ability of the brain to compress information is an indicator of consciousness.

In consciousness neuroscience, this quantification of complexity is used in different scenarios. The main areas of application are research with patients with disorders of consciousness, anesthesia monitoring and sleep studies (e.g., Casarotto et al., 2016). For instance, Burioka et al. (2005) showed that approximate entropy calculated on small segments of electroencephalography (EEG) data decreases with depth of sleep. In particular, there is a linear decrease from wakefulness over sleep stage 1 until 4, while rapid eye movement sleep showed values similar to wakefulness. Analysis of data from EEG, MEG and intracranial EEG recordings confirmed this drop of complexity with a drop in wakefulness (Mateos et al., 2018). Complexity was calculated with entropy and compressibility measures. Zhang et al. (2009) could differentiate between active sleep and quiet sleep in newborns by means of sample entropy (SE). Similarly, the dimensional complexity of the EEG pattern measured by correlation dimension (CD) was found to be higher in active sleep compared to quiet sleep for infants in their 1st months of life (Janjarsjitt et al., 2008b; Scher et al., 2009). Furthermore, scale-free properties caused by the self-similarity of fractals can be used to differentiate between sleep stages whereas an increase or decrease of values depends on the scale-free parameter estimated (Weiss et al., 2009).

In general, the measurement of this scale-free behavior appears promising in the investigation of state transitions (Weiss et al., 2009). Studies with Propofol anesthesia showed a change in scale-free behavior before and after loss of consciousness (Eagleman et al., 2018) and a difference between wakefulness and loss of consciousness as well as recovery from anesthesia (Tagliazucchi et al., 2016). Also with the help of entropy measures, Eagleman et al. (2018) could show a change in complexity of scalp EEG data before and after loss of responsiveness in anesthesia patients. Similarly, Schartner et al. (2015) could distinguish between loss of consciousness during Propofol-induced anesthesia and wakeful rest by means of entropy measures as well as compressibility measures. Furthermore, higher entropy values were shown in the EEG data of healthy control participants compared to unresponsive wakefulness patients, matched for sex and age (Sarà and Pistoia, 2010).

Compressibility measures in combination with transcranial magnetic stimulation are widely used to distinguish between patients with different disorders of consciousness (e.g.,

unresponsive wakefulness state, minimally conscious state, locked in syndrome) and healthy subjects in different sleep stages. In addition, sedation with different anesthetics can be differentiated (e.g., Propofol and Ketamine), whereby in both cases patients are behaviorally unresponsive but in case of Ketamine they report vivid dreams (e.g., Casali et al., 2013; Sarasso et al., 2015; Casarotto et al., 2016; Bodart et al., 2017; Rosanova et al., 2018). Moreover, a recent study on MEG revealed increased lempel-ziv-complexity (LZC) complexity during a psychedelic state of consciousness induced using Ketamine, LSD, and Psilocybin compared to a placebo effect (Schartner et al., 2017). Regarding other, related, patient populations, schizophrenic patients have higher LZC compared to healthy controls (Fernández et al., 2011), and depressed patients have higher MEG pre-treatment complexity that decreases after 6 months of pharmacological treatment (Méndez et al., 2012).

These findings show that there are many valid approaches to use complexity metrics to quantify consciousness, although to our knowledge there are no studies that investigated complexity on early consciousness. Yet, the exact relation of the different metrics is hard to grasp and it is difficult to define a metric that is most suitable for a specific purpose. Therefore, in the current study, as we use a different type of data than previous studies, we follow an explorative approach and apply numerous different metrics to our fMEG data (Table 1). The aim of this approach is to capture different aspects of complexity and compare the metrics regarding their behavior towards this special type of data and their usability when employing them to pursue a novel question. In particular, data from fetal MEG recordings with different gestational ages and additional neonatal recordings are used in this study. The data included in the analysis previously showed auditory event-related responses, which allowed for identification of channels with high brain activity within the sensor space. Such data are used, as it is otherwise difficult to localize clusters of brain activity. The goal of the analysis is to evaluate if complexity metrics are a useful tool for fMEG analysis in the search for fetal consciousness.

## MATERIALS AND METHODS

### Fetal Magnetoencephalography

fMEG is a non-invasive tool to measure heart and brain activity in fetuses in the last trimester of pregnancy and in neonates shortly after birth (Preissl et al., 2004). For the recording of fetal- and neonatal data, the SARA (SQUID Array for Reproductive Assessment, VSM MedTech Ltd., Port

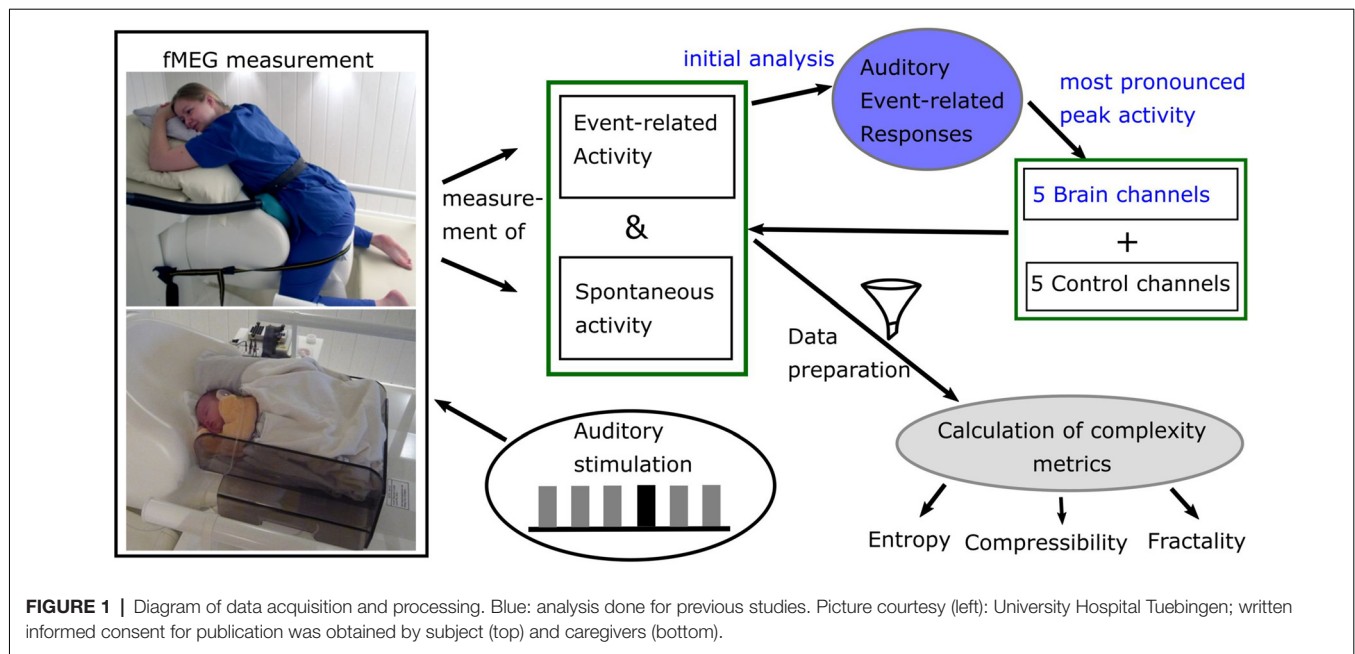
Coquitlam, BC, Canada) system installed at the fMEG Center at the University of Tuebingen was used (Figure 1). To attenuate magnetic activity from the environment, the device is installed in a magnetically shielded room (Vakuumschmelze, Hanau, Germany). The system includes 156 primary magnetic sensors and 29 reference sensors. The magnetic sensors are distributed over a concave array whose shape is designed to match the maternal abdomen. Based on an ultrasound measurement (Ultrasound Logiq 500MD, GE, UK) prior to the fMEG recording, the position of the fetal head is determined and is marked by a localization coil placed on the maternal abdomen. Three additional localization coils are placed on the spine, left and right side of the subject, to track position changes in relation to the sensor array. An ultrasound directly after the measurement is used to confirm the fetal head position. In case of a change in position, datasets are excluded. Auditory stimulation can be presented *via* a balloon placed between the maternal abdomen and the sensor array. Neonates get a small, child-appropriate earphone (Ear Muffins from Natus, Biologic, San Carlos, CA, USA), placed on one ear. For neonatal recordings, a cradle is attached to the fMEG device. The neonate is attended by one parent inside the measurement room and is measured asleep or quiet awake.

### Dataset

For the current analysis, fetal and neonatal data from previously analyzed studies were used (Linder et al., 2014; Morin et al., 2015). Both studies were approved by the local Ethical Committee of the Medical Faculty of the University of Tuebingen (No. 476/2008MPG1 and 339/2010BO1). All participants gave written informed consent in accordance with the Declaration of Helsinki and agreed on reuse of data for additional studies. The auditory stimulation paradigm used in these studies consists of an auditory oddball paradigm with a 500 Hz tone as standard and a 750 Hz tone as deviant. The standard occurs in 80% of times in a pseudorandomized order. Each tone is 500 ms long with an inter-trial interval of 1,500 ms. Forty-five fetal recordings were selected from subjects where auditory event-related responses were detected. They have a gestational age range from 29 to 39 weeks—15 of them in an early phase of the third trimester (29–32 weeks), 15 in the middle (33–36 weeks) and 15 in a later phase (37–39 weeks)—approximately uniformly distributed over the whole age range. Fifteen neonatal recordings were included with an age range from 4 to 46 days (mean = 17.47; SD = 12.68). For all subjects, data with auditory stimulation (“audio”) and data without stimulation (“spont”) is available. As the length of the fetal datasets varies from 6 to 15 min, for all

TABLE 1 | Overview used metrics.

Concept investigated	Method	Objective
Entropy	Multiscale entropy	Measurement of self-similarity of time series by looking at repeating sequences on multiple scales
Entropy	Multiscale permutation entropy	Measurement of self-similarity of time series by looking at probability of patterns in data on multiple scales
Compressibility	Lempel-Ziv-Welch compression	Quantification of compressibility of time series
Fractality	Correlation dimension	Measurement of strangeness of attractor, towards which complex system evolves
Fractality	Scale free approaches	Detection of power-law exponent that describes scale free behavior and additionally description of multifractal properties



of them only the first 6 min were used. The neonatal datasets all have a length of 10 min.

## Data Analysis

### Preprocessing

As first step for all datasets, the maternal magnetocardiogram and fetal magnetocardiogram were detected by template matching or using the Hilbert transform algorithm and were subtracted from the relevant signal, through signal space projection (Vrba et al., 2004; McCubbin et al., 2006; Wilson et al., 2008). One of the two methods was selected depending on which method detected the magnetocardiogram more accurately, which is the established procedure for fetal brain analysis (e.g., Linder et al., 2014). Matlab R2016b was used (The MathWorks, Natick, MA, USA) for all further processing steps, except for the calculation of the compressibility measure, where Python 2.7 was used (Python Software Foundation<sup>1</sup>). Fieldtrip (Oostenveld et al., 2011) was used to filtering and downsampling of data. All fetal data were filtered between 0.5 and 10 Hz and all neonatal data between 0.5 and 15 Hz which is the usual filtering range for the analysis of event-related brain responses (Linder et al., 2014; Schleger et al., 2014). Data were downsampled from 610.35 Hz to 256 Hz. An example dataset can be seen in **Figure 2**.

In the previous analysis of these datasets, five channels were selected for the analysis of auditory event-related responses (example for analysis procedure in Schleger et al., 2014). As these were the five channels with the highest evoked brain activity, they were selected as brain channels for this study ("brain"). Additional to those five channels, five control channels ("control") were selected which were all more than 15 cm away from the previously selected brain channels. By default, the five channels were situated in the upper middle part of the sensor

field. This was suitable for all neonatal recordings due to the similar positioning of the neonate on the sensor array. In fetal recordings, if the fetus was positioned in a way that was too close to the default control channels, resulting in selected brain channels with a distance less than 15 cm, five channels in the lower middle part were selected as control channels. As the five brain channels showed similar behavior, for simplicity of the later analysis only the brain channel with the highest amplitude, and one control channel, were used. Distribution of selected channels within sensor array can be seen in **Figure 3**. The usability of a single channel for complexity analysis was previously demonstrated (e.g., Scher et al., 2005). For later analysis the signal was cut into windows of 1 min (Scher et al., 2009; Kaffashi et al., 2013), to generate a more stationary signal. This results in six time-windows for fetal data and ten time-windows for neonatal data. If a time window included a signal that was higher than 1pT, it was classified as containing an artifact and the concerning time window was removed from the analysis.

### Power Spectral Analysis

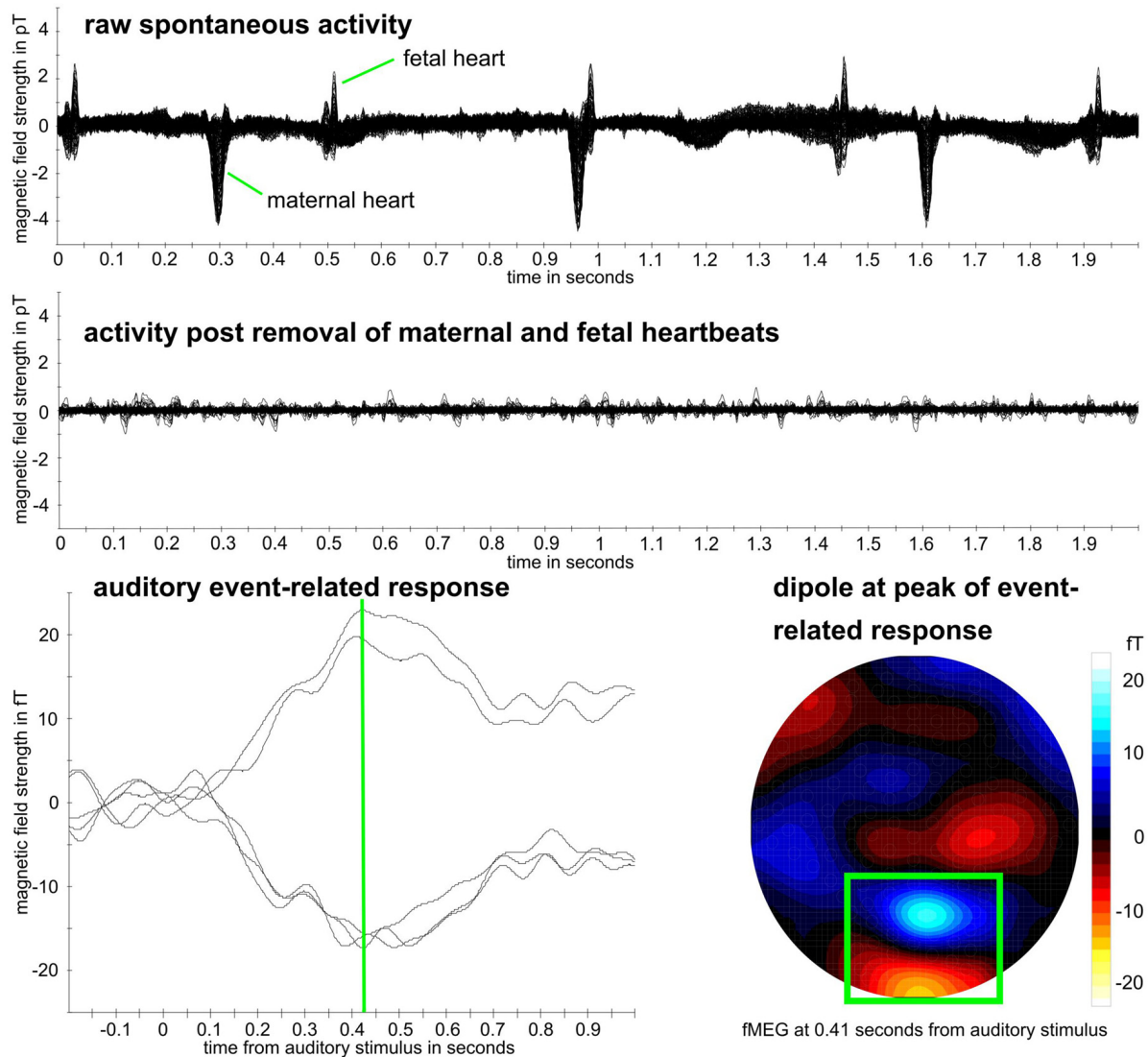
To provide a more traditional view on the data, we also calculated power spectral density, using the Fast Fourier Transform algorithm implemented in Matlab. Unlike described in the preprocessing section, data were pre-filtered from 1 Hz to 35 Hz to allow a better display of power in relation to frequency. All other steps were the same. Mean power in the delta range (1.5–3.5 Hz) and theta range (4–8 Hz) was calculated for further comparisons.

### Complexity Metrics

To measure the informational complexity of the fMEG signal, multiscale entropy (MSE) and multiscale permutation entropy (MPE) were used, as well as LZC. Additionally, we included the geometrical properties of the signal and measured its amount of

<sup>1</sup><https://www.python.org/>





**FIGURE 2 |** Example fetal magnetoencephalography (fMEG) dataset. Recording from subject in 38th week of gestation. Top: raw fMEG data trace, overlay of all channels. Middle: trace after removal of maternal and fetal heart activity. Bottom: auditory event-related response. Green rectangle marks dipole that represents fetal brain activity. Event-related peak from five strongest channels within this dipole is shown, whereas peak time is marked in green.

fractality. As an approximation for this, the CD, also known as dimensional complexity (Janjarasjitt et al., 2008b), is calculated. In addition, scale-free behavior, which is a basic property of a fractal, was taken into account.

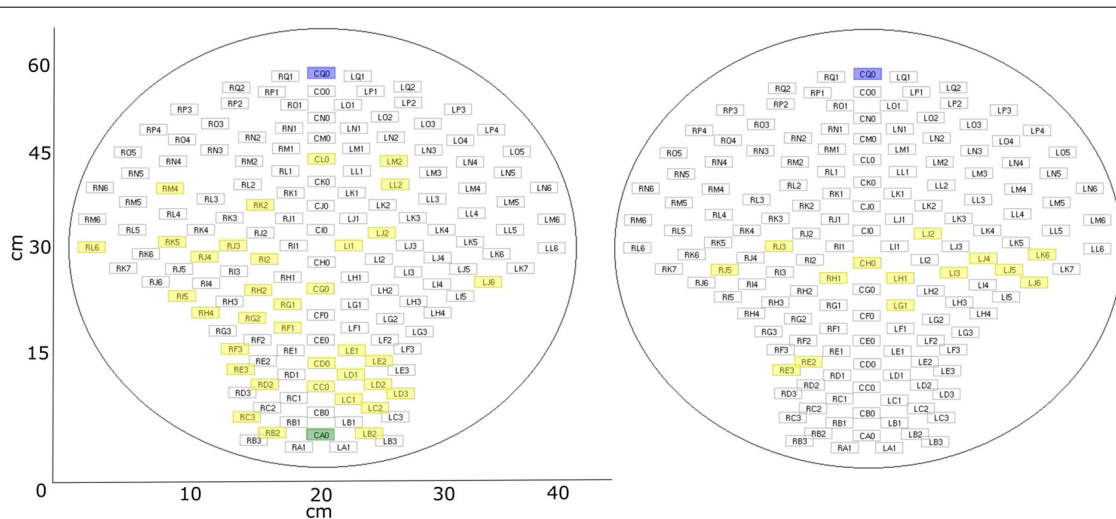
### Multiscale Entropy

MSE calculates SE for different time scales. If SE is calculated, lower values indicate more self-similarity in time series and the calculation is largely independent of recording length and relatively consistent (Richman and Moorman, 2000). This traditional entropy measure takes only one scale into account, therefore MSE uses a coarse-grained time series with different scaling factors which takes long range correlations into account (Costa et al., 2002). By considering multiple scales, both

highly deterministic and completely random signals result in low values, only complex signals can reach a high value (McIntosh et al., 2008).

For the coarse graining step, the dataset is divided into non-overlapping windows and the data points in each window are averaged. SE is based on the definition of Kolmogorov entropy and is defined as the negative logarithm of the probability of two sequences that are similar for  $m$  points, to be similar at the point  $m + 1$  as well (Richman and Moorman, 2000). Equation 1 describes the calculation of SE for one scale whereas  $B$  is the number of template matches of length  $m$  and  $A$  the total number of matches of length  $m + 1$ .  $r$  describes the tolerance within which two points are accounted as similar (Richman and Moorman, 2000).





**FIGURE 3 |** fMEG sensor array with selected channels. Blue: default control channel; green: alternative control channel (if default is closer than 15 cm to brain channel); yellow: selected brain channels. Left: fetal recordings; right: neonatal recordings.

$$SE(m, r, N) = -\ln \frac{A_m(r)}{B_m(r)} \quad (1)$$

For the calculation of MSE, the “msentropy” function out of the WFTB Toolbox (Goldberger et al., 2000; Silva and Moody, 2014) was used. The parameters used in the current analysis are guided by the default parameters described by Costa et al. (2005). We used  $m = 2$ ,  $r = 0.15$  (15% of the standard deviation of the time series) and  $N = 15360$  which is a bit lower than the recommended  $N = 2 \cdot 10^4$  but was selected in terms of comparability of the different methods and did not show any disadvantage compared to an example analysis with a larger  $N$ . We chose a window size between 1 and 20 to calculate the SE for 20 different scales whereas scale one equals the original time series. A time series is considered as more complex than another if a majority of scales show higher entropy values (Costa et al., 2005). For that reason, we used the average MSE over all scales for all further comparisons. To get an impression on the behavior of MSE over different scales, values are displayed in **Figures 4A,B**.

### Multiscale Permutation Entropy

MPE calculates permutation entropy (PE) for multiple time scales. In comparison to other complexity measures, PE is very robust towards noise (Bandt and Pompe, 2002; Zanin et al., 2012). PE looks at different patterns within a time series with the idea that those patterns do not have the same probability of occurrence and that this probability can be informative regarding the underlying dynamics of the system (Zanin et al., 2012). PE uses short samples of a time series to look at their permutation patterns and their frequency of occurrence in relation to all possible permutation patterns (Bandt and Pompe, 2002). PE can be used to quantify complexity of a dynamical time series as it refers to its local order structure (Ouyang et al., 2013). A large value of PE indicates that all permutations are equally

likely, a value close to zero signifies a very regular time series (Ouyang et al., 2013).

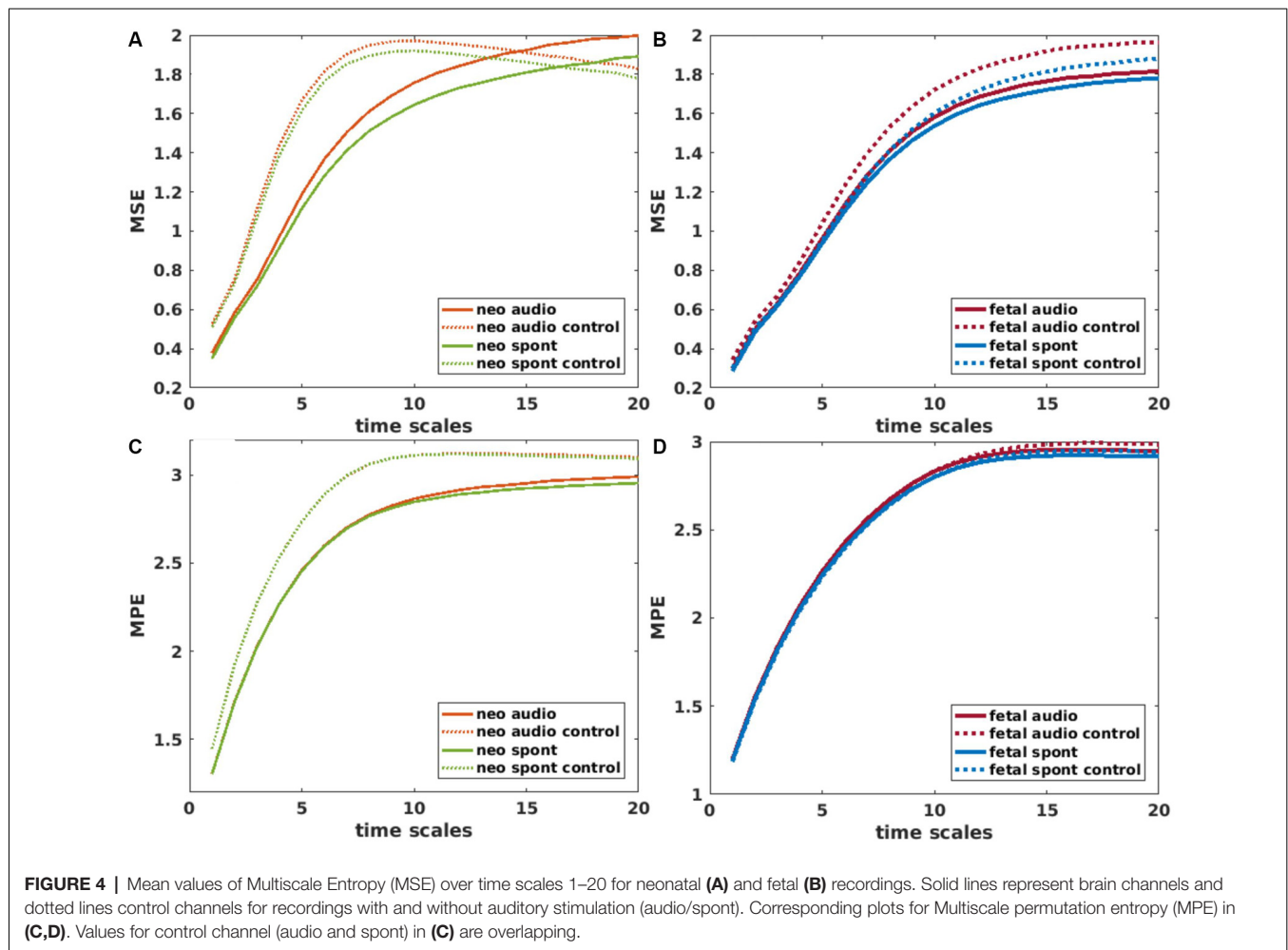
MPE was calculated using the “MPerm” function by Ouyang (2012, November 21)<sup>2</sup>. Like for MSE, the first step is a coarse graining where we selected the same time windows as in the MSE calculation. In this study, the short samples had a length of 4 and a time delay of 1 which corresponds to the default values. Like for the MSE, the average MPE over all scales was used for further comparisons. Values of MPE over different scales are displayed in **Figures 4C,D**.

### Lempel-Ziv-Welch Compressibility

LZC is a measure closely related to Kolmogorov complexity and Shannon entropy (Gao et al., 2011), and is originally described by Ziv and Lempel (1978). For LZC, a dictionary that starts with the shortest new sequence in a time series is built and then adds longer sequences until it captures all non-repetitive sequences (Ruffini, 2017b). The length of this dictionary defines the amount of compressibility of a time series (Ziv and Lempel, 1978). LZC values increase with increasing frequency but not with increasing amplitude as well as with increasing power of noise and increasing signal bandwidth (Aboy et al., 2006).

To calculate LZC, a signal of length  $n$  has to be binarized (with an alphabet with  $A = 2$  symbols), which in our case is done by a median split as a threshold. In particular, values below the median are indicated as zero, whereas values above the median as one. The median split is relatively robust to outliers compared to other methods (Aboy et al., 2006). After the binarization process, the data can be compressed to a set of “words,”  $c(n)$ , and the description length of the dictionary ( $lLZC$ ) is defined as the number of included words times the bits needed to encode those words plus the bits

<sup>2</sup>Multiscale permutation entropy (MPE), version 1.3. Retrieved from <https://ch.mathworks.com/matlabcentral/fileexchange/37288-multiscale-permutation-entropy-mpe->



needed to define a new symbol in the dictionary (Equation 2; Ruffini, 2017a).

$$lLZC = c(n) \log_2 [c(n) + \log_2 A] \quad (2)$$

The complexity counter,  $c(n)$ , is then normalized by the length of the data string (Ruffini, 2017b). In the current analysis,  $\rho\theta$ , which represents the  $c(n)$  normalized by the original string length, is the value used to indicate LZC. A higher  $\rho\theta$  value indicates higher complexity, thus, less ability to compress. For a more detailed description of the process of compression, the reader is referred to Aboy et al. (2006), and for the algorithm used in the current analysis to Ruffini (2017b).

### Correlation Dimension

The CD is a measure for the strangeness of an attractor, which is closely related to the fractal dimension  $D$ . In addition to the geometrical properties of the attractor it takes the dynamics of coverage of the attractor into account (Grassberger and Procaccia, 1983b). The CD uses the statistics of pairwise distances to estimate dimension and is based on the scaling of mass with size (Theiler, 1990). Correlations between points of

long-time series on the attractor are used for that (Grassberger and Procaccia, 1983b). Stationarity of this time series is a requirement to obtain reliable results (Theiler, 1986). Therefore each time window was tested for stationarity with the Augmented Dickey-Fuller-Test (implemented in the Matlab Econometrics Toolbox) and non-stationary time windows excluded from analysis (2% of time windows). For a complete description of a higher dimensional nonlinear system, a time series—which is an observation in one dimension—has to be unfolded into a higher dimensional space, the so called “embedding space” (Janjarsjitt et al., 2008a). The process of embedding a time series in a higher dimension is described by Takens (1981).

The CD is calculated using the correlation integral, defined as the ratio of the distances between any two points that are smaller than a certain radius and all possible distances (Theiler, 1990). To determine the distance in the time series of the points to be correlated, the time delay  $\tau$  is introduced.  $\tau$  can be set with the help of the autocorrelation function (Janjarsjitt et al., 2008a).

For the current analysis, the “gencorint” function was used to determine the CD (Grassberger and Procaccia, 1983a; Albano et al., 1988; Theiler, 1986; out of the Chaotic systems Toolbox,

Leontitis, 2004<sup>3</sup>). For setting the embedding parameter, we used the false nearest neighbor algorithm (Kizilkaya, 2012<sup>4</sup>; Kennel et al., 1992), and determined an embedding dimension of four for the current dataset. To determine the right time delay  $\tau$ , an autocorrelation function was calculated for each subject and time window and the zero point of it was used as an individual value of  $\tau$ . If a time window showed an autocorrelation function with a zero point  $<10$ , it was excluded from further analysis as such a low value does not fit physiological data and is most likely caused by an artifact in this time window. Except for those two parameters all other parameters were set to the default values suggested in the function. For algorithmic efficiency, the slope of the CD can be used as an approximation of the CD (Theiler, 1990). In the current analysis, a linear fit over all points was performed to determine this slope. The slope value was used for all further comparisons.

### Scale-Free Approaches

The self-similarity of a fractal can be expressed by a mathematical power-law with a distinct exponent (Eke et al., 2002). One of these exponents is the Hurst exponent ( $H$ ), which is also referred to as “self-similarity parameter” (Zilber, 2014). It expresses the probability that an event is followed by a similar event and is related to the fractal dimension ( $D$ ) by  $D = 2 - H$ . A value of  $H = 0.5$  shows that a time series is uncorrelated (e.g., white noise).  $0.5 < H$  is an indication for long range correlations and  $H < 0.5$  for long-range anti-correlations (Kantelhardt et al., 2002). For some fractal processes one power-law exponent is not enough to characterize them and a multifractal formalism can be used to describe several exponents (Di Ieva, 2016). In case of multifractality, the scaling function of a signal is not linear anymore and can thus not be described by a single scaling exponent  $H$  but by a nonlinear scaling function (Zilber, 2014). Those multiple  $H$  are called Hölder exponents and can be used to span a multifractal spectrum. The width of this spectrum ( $M$ ) is a measurement for the amount of multifractality (Zilber, 2014).

Self-similar processes also described as  $1/f$  or scale free behavior can be mainly observed in the infraslow frequency range of the power spectrum (Zilber, 2014). This knowledge is important for the selection of scale ranges in scale free analysis. For the following analysis, the focus is on self-similar processes in the range 0.5–2 Hz. To assess  $H$  and  $M$ , two methods for multifractal analysis are employed. For both, all data were normalized to avoid the influence of amplitude differences.

### Multifractal Detrended Fluctuation Analysis

The multifractal detrended fluctuation analysis (MFDFA) is a robust analysis for the estimation of the multifractal spectrum of power-law exponents of a natural time series (Ihlen, 2012). Its basis is the detrended fluctuation analysis (DFA), one of the most popular methods to estimate scale-free behavior in physiological signals which follows the idea that fluctuations within a signal

are following a power-law as a function of the number of sample points (Zilber, 2014). The root mean square (RMS) of a signal is calculated over different scales with a certain number of points within each segment. Larger scales are more affected by slower fluctuations, smaller scales more by faster fluctuations (Ihlen, 2012). For each scale, the RMS of the individual segment—local fluctuation—is calculated, and an overall RMS is computed from these values. The slope of the values of this overall RMS over different scales equals  $H$ . In case of MFDFA this calculation is done for multiple orders. As a result, MFDFA obtains the set of weighted overall RMS values whose slopes obtain Hurst exponents for multiple orders. They can be used, to trace back the multifractal spectrum and its width  $M$  (Ihlen, 2012). For a more detailed description of the calculation steps, see **Supplementary Material** and Ihlen (2012).

The MFDFA toolbox (Ihlen, 2012, MFDFA1 algorithm) was used in this study. As the algorithm is built for random-walk like signals—which are integrals of noise-like signals (Zilber, 2014)—first we checked whether the data resembles noise or random-walk signals. This is determined by the value of  $H$ , which was calculated by a simple DFA. As a cutoff  $H = 1.2$  was selected ( $<$  noise like,  $>$  random walk like; after Ihlen, 2012). If the signal is noise like, it is transformed into a random-walk signal, and if it is random walk like, this step is skipped. As a second step, a general linear detrending of the signal is performed. For the current analysis 19 equally spaced scales ranging from 128 to 512 data points, within a scale, were chosen. The orders were selected to range from  $-5$  to  $5$  in steps of  $0.1$ . The value of  $H_q$  at the order 2 equals  $H$  (Ihlen, 2012). The scale-free parameters  $H$  and  $M$  were evaluated in the further analysis.

### Wavelet-Leader Based Multifractal Formalism

Wavelet-leader based multifractal formalism (WLBMF) poses a fast, theoretically efficient and robust analysis method for multifractal properties of real-world data (Zilber, 2014). It uses wavelet leaders to derive the multifractal properties of a signal by the knowledge of the scaling exponents. This can be done because wavelet leaders precisely reproduce the Hölder exponents of a signal (Wendt et al., 2007). Wavelet leaders are defined as the maximum wavelet coefficients within a predefined segment (Ciuciu et al., 2008). The log-cumulants ( $c1$ – $c3$ ) of the scaling exponents give information about the shape of the multifractal spectrum. Whereas  $c1$  equals its maximum,  $c2$  its width and  $c3$  its asymmetry (Wendt et al., 2007; Zilber, 2014). For a more detailed

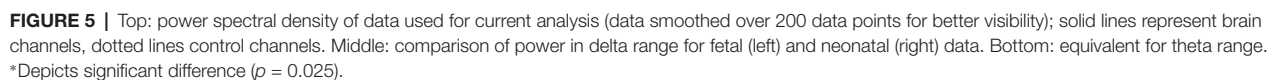
**TABLE 2 |** Comparison of usability of methods.

Method	Computational effort	Parameter space	Comparability	Overall usability
MSE	3.11 s	3	high	OK
MPE	5.69 s	3	moderate	OK
LZC	0.29 s	1	high	good
CD	37.8 min	7	moderate	weak
MFDFA	7.09 s	6	low	weak
WLBMF	2.84 s	7	low	weak

Computational effort reflects the time to compute six time windows of one channel of one dataset. Parameter Space quantifies the number of parameters that can be manipulated within the analysis. Comparability relates to the abundance in related human neuroscience literature.

<sup>3</sup>Chaotic systems toolbox, version 1.0. Retrieved from <https://ch.mathworks.com/matlabcentral/fileexchange/1597-chaotic-systems-toolbox>

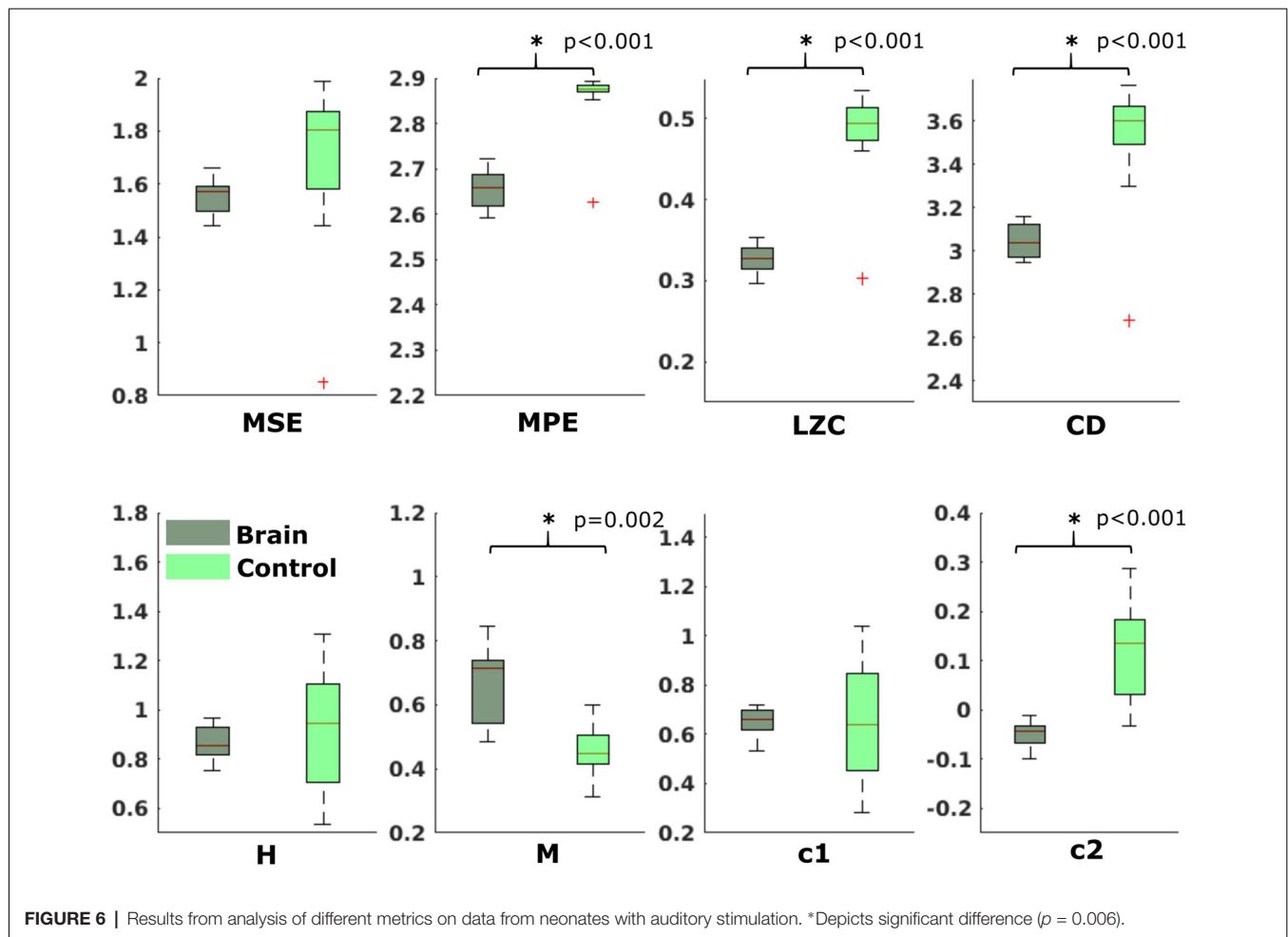
<sup>4</sup>Minimum embedding dimension, version 1.0. Retrieved from <https://ch.mathworks.com/matlabcentral/fileexchange/37239-minimum-embedding-dimension>



For the calculation of the WLBMF, the toolbox described in Wendt et al. (2007) was used. A Daubechies wavelet with three vanishing moments was selected as a mother wavelet. The scales for the WLBMF were chosen in accordance with the scales of the MFDFA. To increase the reliability of the results, Wendt et al. (2007) implemented a bootstrapping process to obtain sets of log-cumulants, which opens up new possibilities for statistical testing. In the present analysis, we used 100 bootstraps and then averaged over the bootstrapped values to retrieve the variables of interest. For a more detailed description see Wendt

For statistical analysis, the results of all time windows of a certain subject and condition were averaged. Those mean values were then tested for normality with a Kolmogorov-Smirnoff test. As they showed to be not normally distributed, groups were compared with a Wilcoxon signed rank test. To determine whether there is a trend over gestational ages, a Pearson correlation was calculated. After Bonferroni correction,





significance levels were set to  $\alpha = 0.006$  for the complexity metrics and  $\alpha = 0.025$  for the additional power spectral analysis.

## RESULTS

### Usability of Methods

After testing several methods for determining complexity of fMEG data, a first step was to evaluate those methods for potential future use. As shown in **Table 2**, we evaluated them in terms of computational costs, parameter space, comparability regarding their prevalence in the literature and summarized this with the term “overall usability.” Calculations were executed on a multiprocessor machine [12 Intel(R) Xeon(R) CPU, X5660, 2.80 GHz; 96 GB RAM] without multicore support. The value for computational cost is the time it took to calculate the results of one subject with one channel and six time windows. LZC was calculated fastest (0.29 s) followed by MSE, MPE and the scale-free metrics (MSE: 3.11 s, WLBMF: 2.84 s, MPE: 5.69 s, MFDFA: 7.09 s). CD took a very long time for calculation (37.8 min). Parameter space is defined as the number of parameters that have to be adjusted within the analysis. Parameter space ranged from 1 (LZC) to 7 (CD and WLBMF). Thereby a low computational effort is favored, as well as a small number of parameters that

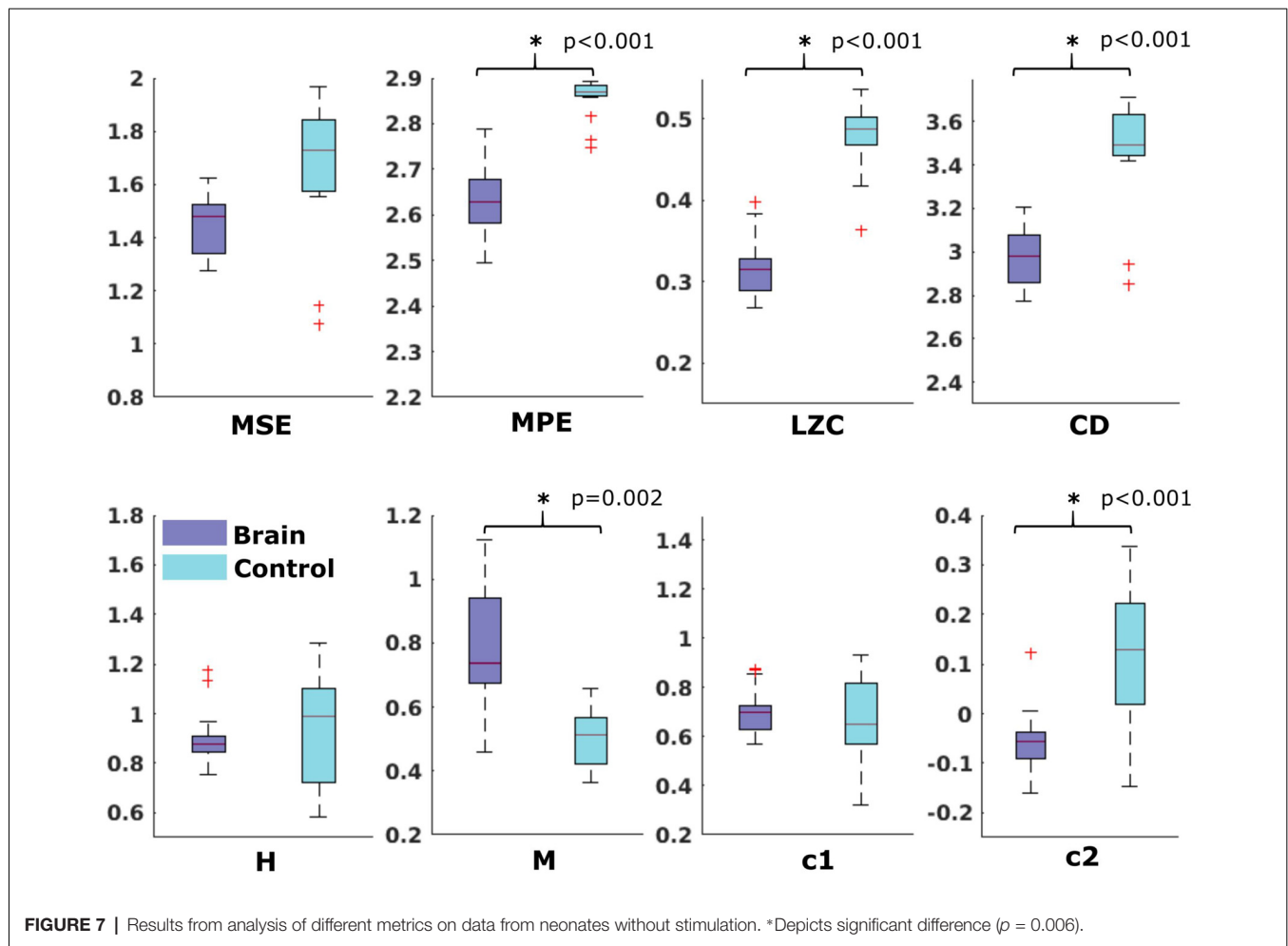
can be adjusted. Abundance of literature is seen as an advantage, to ensure comparability of results. Prevalence in literature is assessed and subjectively estimated within the literature search for this present work and only encounters related human neuroscience literature. No literature search with formal search criteria was performed. Overall LZC had the best overall usability as it is unequivocal and fast and is therefore preferable for future analysis. Nevertheless, this evaluation concerns the present field and has only limited significance for other research questions and other types of data.

### Detection of Brain Activity

#### Power Spectral Analysis

**Figure 5** (top) shows the power spectral density of neonatal, as well as fetal data and the corresponding control channels. For neonatal data, shape of the power spectrum clearly distinguishes between brain and control channel, especially in the range from 0.5 Hz to 15 Hz, which is used for complexity analysis. For fetal datasets, this differentiation is visually less clear, yet, the graph clearly indicates that the major power is in the delta and theta frequency range. Further comparisons in these frequency ranges are shown in **Figure 5**. Comparison of brain and control channels resulted in highly significant differences for neonatal audio and





spont data in both delta and theta range ( $p < 0.001$  in all cases). For fetal data, a significant difference between brain and control was only detected for audio data in the delta range ( $p = 0.001$ ). Interestingly, fetal audio and spont data differed in their power in the delta range ( $p = 0.006$ ).

## Complexity Metrics

### Neonatal Data

In the neonatal datasets, the comparison of brain vs. control resulted in a significant difference for both audio and spontaneous data for LZC, MPE and CD (all  $p < 0.001$ ). Differences for MSE in spontaneous data were marginally significant ( $p = 0.007$ ). In all cases, the control channel showed higher values compared to the brain channel. For the scale-free metrics in both cases, H/c1 could not differentiate between brain and control but M/c2 showed a significant difference ( $p < 0.001$ ). In case of M, the channel with the brain activity appeared more multifractal than the control channel, whereas in case of c2, the opposite was observed. For detailed results, see **Figures 6, 7**.

### Fetal Data

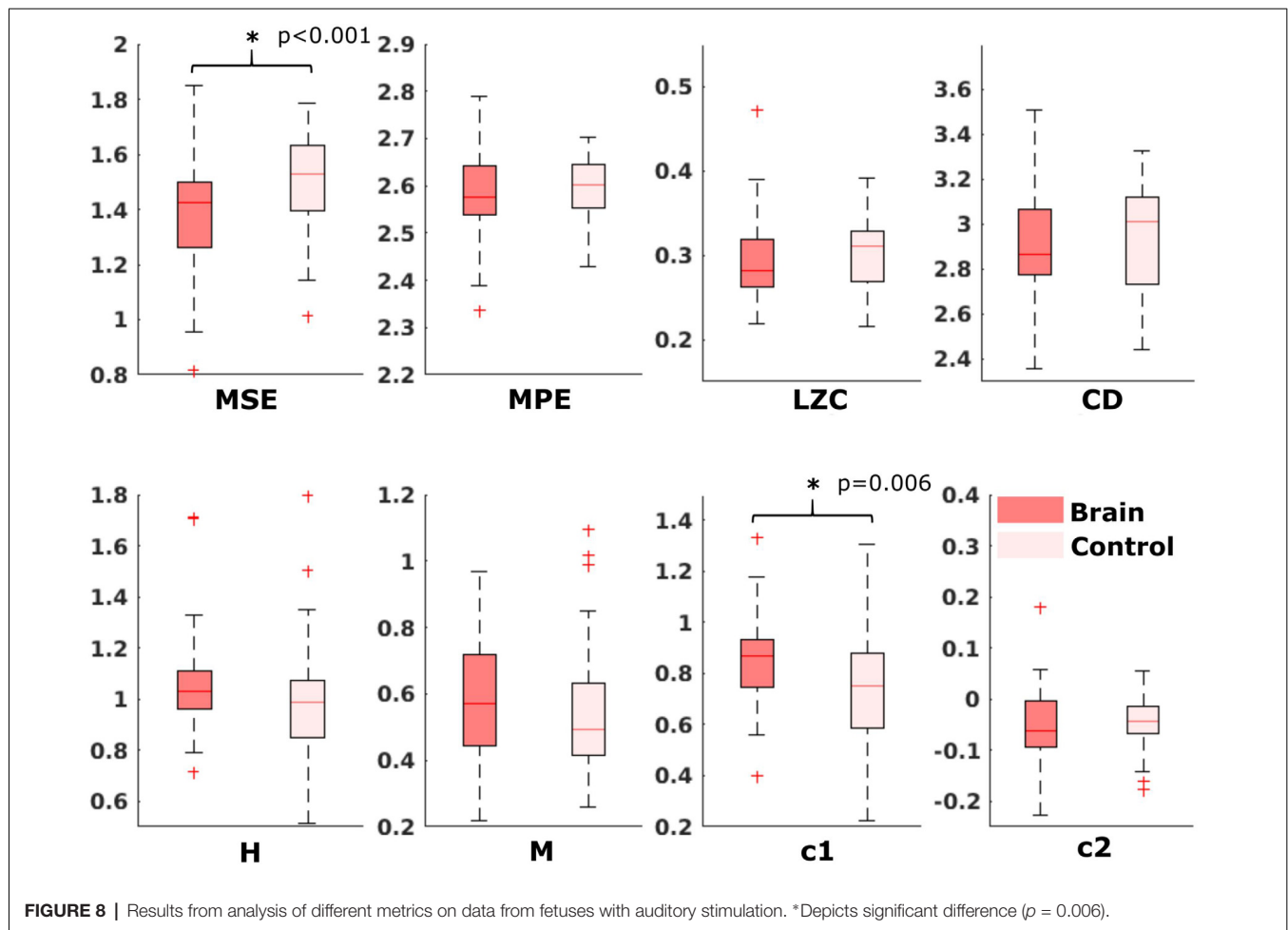
For the comparison of brain vs. control of in the fetal datasets, no clear tendencies could be found. The LZC calculation, as well as

the MPE and CD metric did not reveal a significant difference, neither for audio data nor for spontaneous data. MSE showed a difference for audio data only ( $p > 0.001$ ). Differences between brain and control within the scale-free metrics did not hold after correction for multiple comparisons except c1 for audio data ( $p = 0.006$ ). For detailed results, see **Figures 8, 9**. The correlation analysis did not show a trend over gestational age for the fetal brain data for any of the metrics used.

## DISCUSSION

The usability rating of the different metrics revealed that, concerning the ease of use with fMEG data, LZC was evaluated as best, as it is unequivocal and needs low computational effort. The fractality measures have a high parameter space and therefore forfeit comparability, while entropy measures require a higher computational effort and more parameters to adjust compared to LZC.

In the neonatal population, the channel with brain activity showed lower complexity compared to the control channel, measured by MPE, LZC and CD. As in this scenario, the control channel records environmental noise, it shows us that these

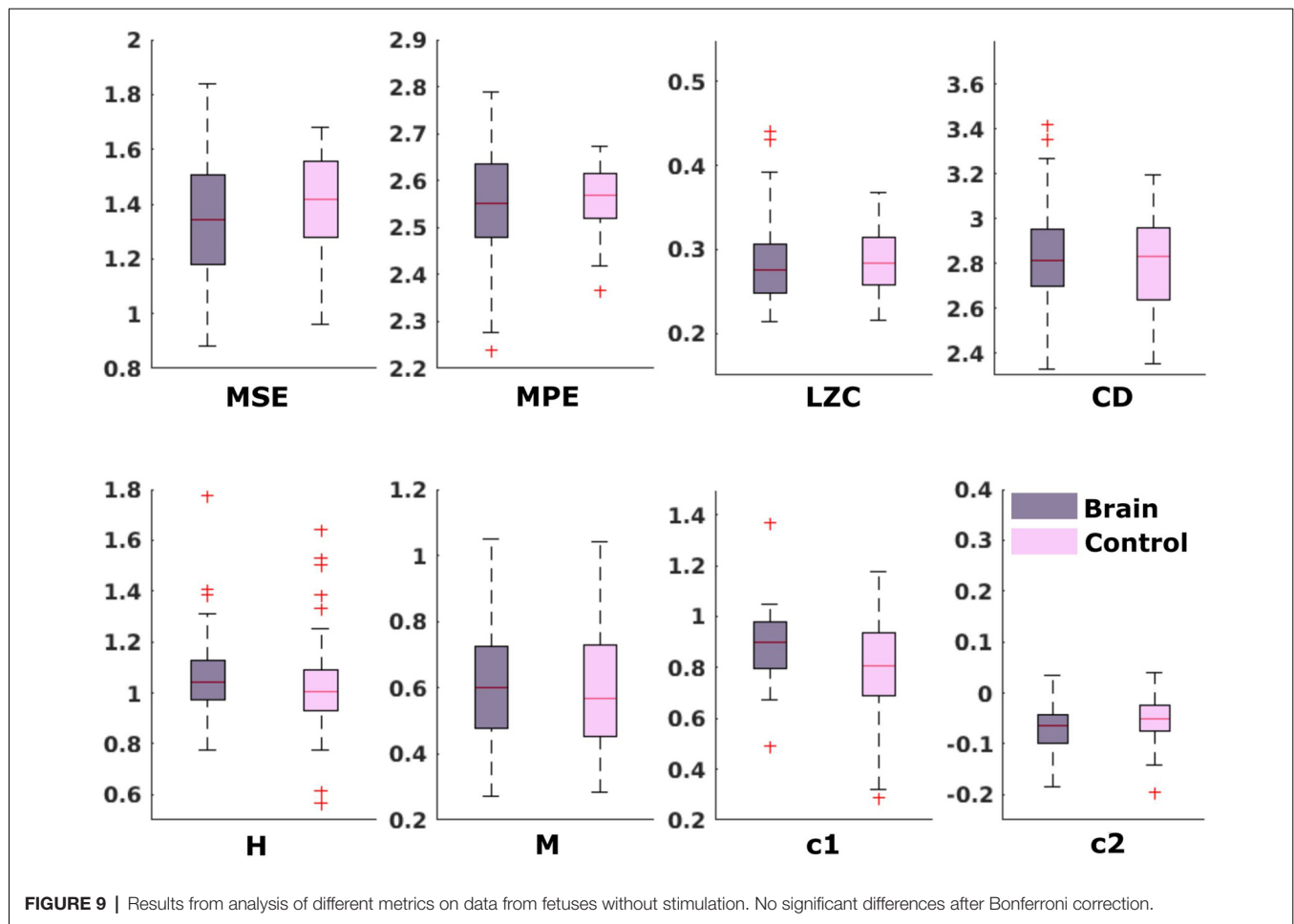


**FIGURE 8** | Results from analysis of different metrics on data from fetuses with auditory stimulation. \*Depicts significant difference ( $p = 0.006$ ).

metrics can clearly differentiate between a physiological signal and noise. This differentiation is supported by the results of the power spectral analysis. These rather clear results can be seen as proof of concept for the general usability of complexity metrics in fMEG. The direction of the difference lies in the nature of the comparison as LZC is highest for Gaussian white noise (Aboy et al., 2006) and entropy maximal for random time series (Costa et al., 2005). For fetal data, this comparison produced less clear results—only MSE and one scale-free metric showed a significant difference. This could be due to the fact that the control channel does not consist of pure environmental noise like in neonatal recordings but can also contain leftovers of physiological signals produced by fetus and mother. Magnetocardiographic activity of both, mother and fetus, should be taken into account as major confounds here. Results from power spectral analysis indicated that there is no clear difference between physiological signals and noise like it is seen in the neonatal recordings. This uncertainty makes analysis with metrics, that are not established for these kinds of signals challenging. The different preprocessing steps that need to be performed, to evaluate fetal compared to neonatal data should be considered likewise, as two heart signals need to be detected and removed instead of one. It is

worth mentioning in this context, that the magnitude of the maternal heart signal is significantly stronger than the fetal brain signal (Figure 2).

The inconsistency of results from different metrics highlights the challenges of working with complexity metrics as neural correlates of consciousness, as well as the caution one should apply to interpret them. Especially, if a dataset consists of as many different aspects as fMEG data does, the choice of the right metric is crucial. Therefore, there is a need for more systematic, comparative studies, to evaluate the relations of different complexity metrics as well as their sensitivity to small changes in analysis parameters. Entropy measures during sleep, for example, can reverse their direction, depending on the time scale used for calculation (Miskovic et al., 2019). Based on the assumption, that when using multiple scales to calculate entropy, random signals result in lower entropy values than complex signals (McIntosh et al., 2008), in the current study we would have expected higher MSE and MPE values for the brain compared to control channel, especially in neonatal data. Yet, we found the opposite in our results; at least when using a scaling range up to 20. Shapes of MSE curves displayed in Figure 4A indicate that there might be a higher complexity for brain signals in a scaling range larger than 20. As the number



of data points left for SE analysis decreases with increasing scale, testing even larger scales is challenging for the present analysis. Because we are looking at activity in a frequency range below 15 Hz, it is nevertheless possible that larger scales reveal more information about patterns in brain activity compared to smaller scales. This challenge—regarding small changes in analysis parameters—especially accounts for metrics with high parameter space like the scale-free metrics that even showed contradicting results in our analysis (M and c2 in neonatal data). The claim of these measures being equivalent (Zilber, 2014) did therefore not hold for the present results. Besides the analysis parameters, the preprocessing steps play a crucial role. We can see this for example in the relatively low LZC values throughout our results, that are related to the previous bandpass filtering of our data, as LZC values decrease with decreasing signal bandwidth (Aboy et al., 2006). This implies that LZC results from different studies, using different filter settings are difficult to compare.

Further work is necessary to conclusively interpret results from this analysis of fetal MEG recordings. Even if largely explorative, this study shows that complexity metrics can be used for fMEG data and the evaluation gives a guidance for future work. Establishing information based metrics of neural activity

for the quantification of consciousness before and shortly after birth still needs additional studies. The broad usage of LZC across a variety of studies, in combination with its use in earlier work on consciousness research, makes this metric especially interesting when pursuing this topic, as results can be compared to other subject populations. Yet, we need a better understanding of each metric and its sensitivity to different aspects of the data as well as their relation to different aspects of complexity, to use these empirical measures of complexity to assess the conscious state of a growing human being. However, as a precise assessment of fetal states is still challenging, the implementation of complexity metrics into fMEG research is a goal, that opens up interesting possibilities.

## ETHICS STATEMENT

Data used in this work was re-used from two studies previously conducted at our center. Both studies were approved by the local Ethical Committee of the Medical Faculty of the University of Tübingen (No. 476/2008MPG1 and 339/2010BO1). All participants gave written informed consent in accordance with the Declaration of Helsinki and agreed on reuse of data for additional studies.

## AUTHOR CONTRIBUTIONS

JM conducted the analysis and wrote the manuscript. SB and EK revised the manuscript. JM, FS and HP conceptualized the study. SB, EK, GR, FW and HP advised the analysis. FS provided preceding analysis. All authors read and approved the final version of the manuscript.

## FUNDING

This work was funded by the FET Open Luminous project (H2020 FETOPEN-2014-2015-RIA under agreement No. 686764) as part of the European Union's Horizon 2020 research and training program 2014–2018. We acknowledge support by

the German Research Foundation and Open Access Publishing Fund of the University of Tuebingen.

## ACKNOWLEDGMENTS

This work was presented as a poster at the HBP International Conference “Understanding Consciousness; A scientific quest for the 21st century,” June 21st–22nd, 2018 in Barcelona.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnsys.2019.00023/full#supplementary-material>

## REFERENCES

- Aboy, M., Hornero, R., Abásolo, D., and Álvarez, D. (2006). Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Trans. Biomed. Eng.* 53, 2282–2288. doi: 10.1109/tbme.2006.883696
- Albano, A. M., Muench, J., Schwartz, C., Mees, A. I., and Rapp, P. E. (1988). Singular-value decomposition and the Grassberger-Procaccia algorithm. *Phys. Rev. A Gen. Phys.* 38, 3017–3026. doi: 10.1103/physreva.38.3017
- Bandt, C., and Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* 88:174102. doi: 10.1103/physrevlett.88.174102
- Baranger, M. (2000). *Chaos, Complexity and Entropy*. Cambridge: New England Complex Systems Institute.
- Bodart, O., Gosseries, O., Wannez, S., Thibaut, A., Annen, J., Boly, M., et al. (2017). Measures of metabolism and complexity in the brain of patients with disorders of consciousness. *Neuroimage Clin.* 14, 354–362. doi: 10.1016/j.nicl.2017.02.002
- Burioka, N., Miyata, M., Cornélissen, G., Halberg, F., Takeshima, T., Kaplan, D. T., et al. (2005). Approximate entropy in the electroencephalogram during wake and sleep. *Clin. EEG Neurosci.* 36, 21–24. doi: 10.1177/155005940503600106
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* 5:198ra105. doi: 10.1126/scitranslmed.3006294
- Casarotto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fecchio, M., Napolitani, M., et al. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Ann. Neurol.* 80, 718–729. doi: 10.1002/ana.24779
- Ciuciu, P., Abry, P., Rabrait, C., and Wendt, H. (2008). Log wavelet leaders cumulant based multifractal analysis of EVI fMRI time series: evidence of scaling in ongoing and evoked brain activity. *IEEE J. Sel. Top. Signal Process.* 2, 929–943. doi: 10.1109/jstsp.2008.2006663
- Costa, M., Goldberger, A. L., and Peng, C.-K. (2002). Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* 89:068102. doi: 10.1103/PhysRevLett.89.068102
- Costa, M., Goldberger, A. L., and Peng, C.-K. (2005). Multiscale entropy analysis of biological signals. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* 71:021906. doi: 10.1103/PhysRevE.71.021906
- Cover, T. M., and Thomas, J. A. (2012). *Elements of Information Theory*. Hoboken, NJ: John Wiley and Sons, Inc.
- Dehaene, S., Kerszberg, M., and Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. U S A* 95, 14529–14534. doi: 10.1073/pnas.95.24.14529
- Di Ieva, A. (2016). *The Fractal Geometry of the Brain*. New York, NY: Springer.
- Eagleman, S. L., Vaughn, D. A., Drover, D. R., Drover, C. M., Cohen, M. S., Ouellette, N. T., et al. (2018). Do complexity measures of frontal EEG distinguish loss of consciousness in geriatric patients under anesthesia? *Front. Neurosci.* 12:645. doi: 10.3389/fnins.2018.00645
- Eke, A., Herman, P., Kocsis, L., and Kozak, L. R. (2002). Fractal characterization of complexity in temporal physiological signals. *Physiol. Meas.* 23, R1–R38. doi: 10.1088/0967-3334/23/1/201
- Elbert, T., Ray, W. J., Kowalik, Z. J., Skinner, J. E., Graf, K. E., and Birbaumer, N. (1994). Chaos and physiology: deterministic chaos in excitable cell assemblies. *Physiol. Rev.* 74, 1–47. doi: 10.1152/physrev.1994.74.1.1
- Fernández, A., López-Ibor, M.-I., Turrero, A., Santos, J.-M., Morón, M.-D., Hornero, R., et al. (2011). Lempel-Ziv complexity in schizophrenia: a MEG study. *Clin. Neurophysiol.* 122, 2227–2235. doi: 10.1016/j.clinph.2011.04.011
- Gao, J., Hu, J., and Tung, W.-W. (2011). Complexity measures of brain wave dynamics. *Cogn. Neurodyn.* 5, 171–182. doi: 10.1007/s11571-011-9151-3
- Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). PhysioBank, PhysioToolkit and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101, e215–e220. doi: 10.1161/01.cir.101.23.e215
- Grassberger, P., and Procaccia, I. (1983a). Characterization of strange attractors. *Phys. Rev. Lett.* 50:346. doi: 10.1103/PhysRevLett.50.346
- Grassberger, P., and Procaccia, I. (1983b). Measuring the strangeness of strange attractors. *Physica D* 9, 189–208. doi: 10.1016/0167-2789(83)90298-1
- Ihlen, E. A. (2012). Introduction to multifractal detrended fluctuation analysis in matlab. *Front. Physiol.* 3:141. doi: 10.3389/fphys.2012.00141
- Janjarsjitt, S., Scher, M. S., and Loparo, K. A. (2008a). Nonlinear dynamical analysis of the neonatal EEG time series: the relationship between neurodevelopment and complexity. *Clin. Neurophysiol.* 119, 822–836. doi: 10.1016/j.clinph.2007.11.012
- Janjarsjitt, S., Scher, M. S., and Loparo, K. A. (2008b). Nonlinear dynamical analysis of the neonatal EEG time series: the relationship between sleep state and complexity. *Clin. Neurophysiol.* 119, 1812–1823. doi: 10.1016/j.clinph.2008.03.024
- Kaffashi, F., Scher, M. S., Ludington-Hoe, S. M., and Loparo, K. A. (2013). An analysis of the kangaroo care intervention using neonatal EEG complexity: a preliminary study. *Clin. Neurophysiol.* 124, 238–246. doi: 10.1016/j.clinph.2012.06.021
- Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., and Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A Stat. Mech. Appl.* 316, 87–114. doi: 10.1016/S0378-4371(02)01383-3
- Kennel, M. B., Brown, R., and Di Barbanel, H. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* 45, 3403–3411. doi: 10.1103/physreva.45.3403
- Kizilkaya, M. (2012). Minimum embedding dimension, version 1.0. Available online at: <https://ch.mathworks.com/matlabcentral/fileexchange/37239-minimum-embedding-dimension>. Accessed May 14, 2019.
- Kostović, I., and Judoš, M. (2010). The development of the subplate and thalamocortical connections in the human foetal brain. *Acta Paediatr.* 99, 1119–1127. doi: 10.1111/j.1651-2227.2010.01811.x
- Lagercrantz, H., and Changeux, J.-P. (2009). The emergence of human consciousness: from fetal to neonatal life. *Pediatr. Res.* 65, 255–260. doi: 10.1203/PDR.0b013e3181973b0d



- Leontitis, A. (2004). Chaotic systems toolbox, version 1.0. Available online at: <https://ch.mathworks.com/matlabcentral/fileexchange/1597-chaotic-systems-toolbox>. Accessed May 14, 2019.
- Linder, K., Schleger, F., Ketterer, C., Fritsche, L., Kiefer-Schmidt, I., Hennige, A., et al. (2014). Maternal insulin sensitivity is associated with oral glucose-induced changes in fetal brain activity. *Diabetologia* 57, 1192–1198. doi: 10.1007/s00125-014-3217-9
- Lutzenberger, W., Preissl, H., and Pulvermüller, F. (1995). Fractal dimension of electroencephalographic time series and underlying brain processes. *Biol. Cybern.* 73, 477–482. doi: 10.1007/s004220050203
- Mateos, D. M., Erra, R. G., Wennberg, R., and Perez Velazquez, J. L. (2018). Measures of entropy and complexity in altered states of consciousness. *Cogn. Neurodyn.* 12, 73–84. doi: 10.1007/s11571-017-9459-8
- McCubbin, J., Robinson, S. E., Cropp, R., Moiseev, A., Vrba, J., Murphy, P., et al. (2006). Optimal reduction of MCG in fetal MEG recordings. *IEEE Trans. Biomed. Eng.* 53, 1720–1724. doi: 10.1109/TBME.2006.876619
- McIntosh, A. R., Kovacevic, N., and Itier, R. J. (2008). Increased brain signal variability accompanies lower behavioral variability in development. *PLoS Comput. Biol.* 4:e1000106. doi: 10.1371/journal.pcbi.1000106
- Méndez, M. A., Zuluaga, P., Hornero, R., Gómez, C., Escudero, J., Rodríguez-Palancas, A., et al. (2012). Complexity analysis of spontaneous brain activity: effects of depression and antidepressant treatment. *J. Psychopharmacol.* 26, 636–643. doi: 10.1177/0269881111408966
- Miskovic, V., MacDonald, K. J., Rhodes, L. J., and Cote, K. A. (2019). Changes in EEG multiscale entropy and power-law frequency scaling during the human sleep cycle. *Hum. Brain Mapp.* 40, 538–551. doi: 10.1002/hbm.24393
- Morin, E. C., Schleger, F., Preissl, H., Braendle, J., Eswaran, H., Abele, H., et al. (2015). Functional brain development in growth-restricted and constitutionally small fetuses: a fetal magnetoencephalography case-control study. *BJOG* 122, 1184–1190. doi: 10.1111/1471-0528.13347
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869
- Ouyang, G., Li, J., Liu, X., and Li, X. (2013). Dynamic characteristics of absence EEG recordings with multiscale permutation entropy analysis. *Epilepsy Res.* 104, 246–252. doi: 10.1016/j.eplepsyres.2012.11.003
- Preissl, H., Lowery, C. L., and Eswaran, H. (2004). Fetal magnetoencephalography: current progress and trends. *Exp. Neurol.* 190, 28–36. doi: 10.1016/j.expneurol.2004.06.016
- Pritchard, W. S., and Duke, D. W. (1995). Measuring chaos in the brain—a tutorial review of EEG dimension estimation. *Brain Cogn.* 27, 353–397. doi: 10.1006/brcg.1995.1027
- Richman, J. S., and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049. doi: 10.1152/ajpheart.2000.278.6.H2039
- Rodríguez-Bermúdez, G., and García-Laencina, P. J. (2015). Analysis of EEG signals using nonlinear dynamics and chaos: a review. *Appl. Math. Inf. Sci.* 9:5, 2309–2321. doi: 10.12785/amis/090512
- Rosanova, M., Fecchio, M., Casarotto, S., Sarasso, S., Casali, A. G., Pigorini, A., et al. (2018). Sleep-like cortical OFF-periods disrupt causality and complexity in the brain of unresponsive wakefulness syndrome patients. *Nat. Commun.* 9:4427. doi: 10.1038/s41467-018-06871-1
- Ruffini, G. (2017a). An algorithmic information theory of consciousness. *Neurosci. Conscious.* 2017:nix019. doi: 10.1093/nc/nix019
- Ruffini, G. (2017b). *Lempel-Ziv Complexity Reference*. Starlab Technical Note, TN00344 (V1.0).
- Sarà, M., and Pistoia, F. (2010). Complexity loss in physiological time series of patients in a vegetative state. *Nonlinear Dynamics Psychol. Life Sci.* 14, 1–13.
- Sarasso, S., Boly, M., Napolitani, M., Gosseries, O., Charland-Verville, V., Casarotto, S., et al. (2015). Consciousness and complexity during unresponsiveness induced by propofol, xenon and ketamine. *Curr. Biol.* 25, 3099–3105. doi: 10.1016/j.cub.2015.10.014
- Schartner, M. M., Carhart-Harris, R. L., Barrett, A. B., Seth, A. K., and Muthukumaraswamy, S. D. (2017). Increased spontaneous MEG signal diversity for psychoactive doses of ketamine, LSD and psilocybin. *Sci. Rep.* 7:46421. doi: 10.1038/srep46421
- Schartner, M., Seth, A., Noirhomme, Q., Boly, M., Bruno, M.-A., Laureys, S., et al. (2015). Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. *PLoS One* 10:e0133532. doi: 10.1371/journal.pone.0133532
- Scher, M. S., Ludington-Hoe, S., Kaffashi, F., Johnson, M. W., Holditch-Davis, D., and Loparo, K. A. (2009). Neurophysiologic assessment of brain maturation after an 8-week trial of skin-to-skin contact on preterm infants. *Clin. Neurophysiol.* 120, 1812–1818. doi: 10.1016/j.clinph.2009.08.004
- Scher, M. S., Waisanen, H., Loparo, K., and Johnson, M. W. (2005). Prediction of neonatal state and maturational change using dimensional analysis. *J. Clin. Neurophysiol.* 22, 159–165. doi: 10.1097/01.WNP.0000161258.92634.D6
- Schleger, F., Landerl, K., Muenssinger, J., Draganova, R., Reinl, M., Kiefer-Schmidt, I., et al. (2014). Magnetoencephalographic signatures of numerosity discrimination in fetuses and neonates. *Dev. Neuropsychol.* 39, 316–329. doi: 10.1080/87565641.2014.914212
- Silva, I., and Moody, G. B. (2014). An open-source toolbox for analysing and processing physionet databases in matlab and octave. *J. Open Res. Softw.* 2:e27. doi: 10.5334/jors.bi
- Tagliazucchi, E., Chialvo, D. R., Siniatchkin, M., Amico, E., Brichant, J.-F., Bonhomme, V., et al. (2016). Large-scale signatures of unconsciousness are consistent with a departure from critical dynamics. *J. R. Soc. Interface* 13:20151027. doi: 10.1098/rsif.2015.1027
- Takens, F. (1981). “Detecting strange attractors in turbulence,” in *Dynamical Systems and Turbulence, Warwick 1980*, eds D. Rand and L. S. Young (Berlin, Heidelberg: Springer), 366–381.
- Theiler, J. (1986). Spurious dimension from correlation algorithms applied to limited time-series data. *Phys. Rev. A Gen. Phys.* 34, 2427–2432. doi: 10.1103/physrev.34.2427
- Theiler, J. (1990). Estimating fractal dimension. *J. Opt. Soc. Am. A* 7, 1055–1073. doi: 10.1364/JOSAA.7.001055
- Tononi, G., and Edelman, G. M. (1998). Consciousness and complexity. *Science* 282, 1846–1851. doi: 10.1126/science.282.5395.1846
- Vrba, J., Robinson, S. E., McCubbin, J., Lowery, C. L., Eswaran, H., Wilson, J. D., et al. (2004). Fetal MEG redistribution by projection operators. *IEEE Trans. Biomed. Eng.* 51, 1207–1218. doi: 10.1109/TBME.2004.827265
- Weiss, B., Clemens, Z., Bódizs, R., Vágó, Z., and Halász, P. (2009). Spatio-temporal analysis of monofractal and multifractal properties of the human sleep EEG. *J. Neurosci. Methods* 185, 116–124. doi: 10.1016/j.jneumeth.2009.07.027
- Wendt, H., Abry, P., and Jaffard, S. (2007). Bootstrap for empirical multifractal analysis. *IEEE Signal Process. Mag.* 24, 38–48. doi: 10.1109/msp.2007.4286563
- Wilson, J. D., Govindan, R. B., Hatton, J. O., Lowery, C. L., and Preissl, H. (2008). Integrated approach for fetal QRS detection. *IEEE Trans. Biomed. Eng.* 55, 2190–2197. doi: 10.1109/TBME.2008.923916
- Zanin, M., Zunino, L., Rosso, O. A., and Papo, D. (2012). Permutation entropy and its main biomedical and econophysics applications: a review. *Entropy* 14, 1553–1577. doi: 10.3390/e14081553
- Zhang, D., Ding, H., Liu, Y., Zhou, C., Ding, H., and Ye, D. (2009). Neurodevelopment in newborns: a sample entropy analysis of electroencephalogram. *Physiol. Meas.* 30, 491–504. doi: 10.1088/0967-3334/30/5/006
- Zilber, N. (2014). *ERF and Scale-Free Analyses of Source-Reconstructed MEG Brain Signals During a Multisensory Learning Paradigm*. (Doctoral dissertation, Université Paris Sud-Paris XI, Paris).
- Ziv, J., and Lempel, A. (1978). Compression of individual sequences via variable-rate coding. *IEEE Trans. Inf. Theory* 24, 530–536. doi: 10.1109/tit.1978.1055934

**Conflict of Interest Statement:** EK is an employee of Starlab SLU, the company that gave birth to Neuroelectrics in 2011, and GR is a co-founder, shareholder, and employee of Neuroelectrics and Starlab.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Moser, Bensaid, Kroupi, Schleger, Wendling, Ruffini and Preißl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Physically Sufficient Neural Mechanisms of Consciousness

Matthew Owen<sup>1\*</sup> and Mihretu P. Guta<sup>2,3</sup>

<sup>1</sup>Department of Philosophy, Gonzaga University, Spokane, WA, United States, <sup>2</sup>Department of Philosophy, Biola University, La Mirada, CA, United States, <sup>3</sup>Department of Philosophy, Azusa Pacific University, Azusa, CA, United States

Neural correlates of consciousness (for brevity NCC) are foundational to the scientific study of consciousness. Chalmers (2000) has provided the most informative and influential definition of NCC, according to which neural correlates are minimally *sufficient* for consciousness. However, the sense of sufficiency needs further clarification since there are several relevant senses with different entailments. In section one of this article, we give an overview of the desiderata for a good definition of NCC and Chalmers's definition. The second section analyses the merit of understanding the sufficiency of neural correlates for corresponding consciousness according to three relevant types of sufficiency: logical, metaphysical, and physical. In section three, a theoretical approach to consciousness studies is suggested in light of the sense in which NCC are sufficient for consciousness. Section four addresses a concern some might have about this approach. By the end, it will become apparent that our conception of NCC has important implications for research methodology, neuroethics, and the vitality of the search for NCC.

**Keywords:** neural correlate of consciousness (NCC), consciousness, theoretical approaches, sufficiency, unresponsive wakefulness syndrome (UWS), Alzheimer's disease, brain organoids

## OPEN ACCESS

### Edited by:

Olivia Gosseries,  
University of Liège, Belgium

### Reviewed by:

Mazyar Fallah,  
York University, Canada  
Marie-Christine Alice Nizzi,  
Harvard University, United States  
Michele Farisco,  
Uppsala University, Sweden

### \*Correspondence:

Matthew Owen  
owen@gonzaga.edu

**Received:** 04 April 2019

**Accepted:** 11 June 2019

**Published:** 04 July 2019

### Citation:

Owen M and Guta MP  
(2019) Physically Sufficient Neural  
Mechanisms of Consciousness.  
*Front. Syst. Neurosci.* 13:24.  
doi: 10.3389/fnsys.2019.00024

## INTRODUCTION

The search for neural correlates of consciousness (NCC) is integral to the science of consciousness (see Metzinger, 2000b; Kandel and Hudspeth, 2013, p. 18; Hohwy and Bayne, 2015; Frith and Rees, 2017, pp. 161–163; Overgaard, 2017; Storm et al., 2017). This search for the neuronal mechanisms at the mind-brain hinge has been marching on steadily for several decades (see Crick and Koch, 1990). Yet these are still relatively early days in an audacious endeavor, and thus further conceptual clarity is still in order.

*Sufficiency* is a key concept in an adequate definition of NCC (Koch, 2004, p. 97). However, there are multiple senses in which something might be “sufficient” since there are multiple types of sufficiency. The aim of this article is to clarify the sense in which NCC are sufficient for consciousness. For this has significant implications pertaining to research methodology, neuroethics, and the vitality of NCC research itself, which will become apparent throughout the article.

Chalmers (2000) has provided the most perspicuous definition of NCC, according to which neural correlates are sufficient for consciousness (see Koch, 2019, ch. 5). In the first section of this article, overviews of Chalmers's definition and the desiderata for a good definition of NCC are provided. The second section analyses the merit of understanding the sufficiency of NCC according to three relevant types of sufficiency: logical sufficiency, metaphysical sufficiency, and physical sufficiency. This analysis demonstrates that the definition of NCC is best off if neural correlates are defined according to physical sufficiency.

However, the identification of NCC so defined will provide a limited amount of information that is insufficient by itself to fully address particular issues consciousness researchers are interested in. For this reason, in section three, a theoretical approach to consciousness studies is suggested. This approach allows for a theoretical framework to provide additional information upon which further inferences regarding consciousness can be based. To elucidate the practical application of such an approach, two examples are provided. The first concerns an unresponsive patient and what the Integrated Information Theory and the Global Neuronal Workspace (GNW) theory would predict about the patient's state of consciousness given the neural activity present. The second example concerns Alzheimer's disease and what the inability to consciously recall memories implies about the persistence of the self in light of two different theoretical frameworks, the psychological continuity view of the self and the Mind-Body Powers model of NCC. Section four addresses the concern that the proposed theoretical approach will philosophically pollute consciousness science.

## DEFINING NCC

To date, the most influential definition of NCC is found in Thomas Metzinger's (2000b) edited volume *NCC: Empirical and Conceptual Questions*<sup>1</sup>. In "What Is a Neural Correlate of Consciousness?" Chalmers (2000) points out that there are various conceptions of NCC within the research literature. So he aims to provide conceptual clarity by giving a theoretically neutral, reasonable and coherent definition that reflects common usage (Chalmers, 2000, pp. 31, 38). This section clarifies the desiderata for the definition of NCC and then briefly explicates Chalmers's definition. This will lay the foundation for analyzing the concept of "sufficiency" central to the definition in the following section "Analyzing Sufficiency".

## Desiderata

The desiderata for a definition of NCC are the standards it needs to meet to be an adequate definition. Chalmers (2000, pp. 31, 38) briefly suggests several desiderata but without much elaboration. As mentioned above, he aimed for a definition that: (1) fits with common usage; (2) is reasonable and coherent; and (3) is theoretically neutral. We assume the same desiderata, but since adequately exploring the merit of different senses of sufficiency depends on these desiderata, a clarification of each is in order.

The first desideratum is that the definition comports with standard usage of the term within the NCC research community. When Chalmers proposed his definition for NCC there were fewer relevant publications. Since then there has been an explosion of literature and studies pertaining to NCC. As Koch (2019, ch. 5) points out: "Over the years, the deceptively simple concept of 'neuronal correlates of consciousness' has been dissected, refined, extended, transmogrified and dismissed." Consequently, it is no small task to verify that this desideratum of cohering with standard usage has been met. In an effort to do so reasonably, we must first consider the use and

understanding of the term by leading researchers. One indicator that Chalmers's definition meets this desideratum is Koch's (2019, ch. 5) acknowledgment that his own preferred operational definition was "Helped along by Chalmers's more rigorous formulation. ..." And although Sascha Benjamin Fink (2016, p. 3) finds Chalmers's definition unsatisfactory and offers an alternative, he acknowledges its wide embrace:

*Most follow Chalmers in his outline: Mormann and Koch (2007) talk about "neural mechanisms or events which are sufficient for consciousness," and similar wording can be found in Hohwy and Frith (2004), Block (2005, p. 46), Hohwy (2007, 2009), Tononi and Koch (2008), Aru et al. (2012), Bayne and Hohwy (2013) and others. Although none of the mentioned authors quotes Chalmers (2000), the similarities are obvious. It is therefore reasonable to assume that the Chalmers-NCC is the default understanding of "NCC" in the neuroscience of consciousness<sup>2</sup>.*

The second desideratum is that the definition be reasonable and coherent. On the one hand, it should be internally coherent. If there is a logical contradiction within the definition itself, then it would be rendered useless. On the other hand, it would also be problematic if the definition is clearly contradictory with external knowledge. For example, suppose Hume's (2007, p. 55) anti-realist<sup>3</sup> view of causation was a proven fact but the definition of NCC fundamentally hinged on a realist view of causation. The implication would be that the referent "NCC" would not refer to anything that actually exists, which would sabotage NCC research. Thus, the definition must be internally and externally coherent.

The third desideratum is that the definition be theoretically neutral. Half a century before the contemporary search for NCC began, in an article entitled "The cerebral cortex in man: the cerebral cortex and consciousness," Penfield (1938, p. 417) wrote:

*It seems quite proper that neurologists should push their investigation into the neurologic mechanism associated with consciousness and should inquire closely into the localization of that mechanism without apology and without undertaking responsibility for the theory of consciousness.*

Penfield is presenting a vision of brain science that studies the neuronal mechanism of consciousness but does not require researchers to commit to a theory of consciousness in order to identify the mechanism (see Penfield, 1975, pp. 4, 47 and 114). Given such neutrality, the legitimacy of such research does not hinge upon the truth or acceptance of a particular view of consciousness.

Today there is a myriad of views about the mind and consciousness embraced by influential philosophers and

<sup>1</sup>See also Chalmers (2010, ch. 3).

<sup>2</sup>Fink's claim that "none of the mentioned authors quotes Chalmers (2000)" seems mistaken, as Mormann and Koch (2007) and Bayne and Hohwy (2013, p. 25) cite Chalmers (2000), as does Block (2005, p. 47) although it's misdated in Block's bibliography.

<sup>3</sup>There is debate about whether this is Hume's view (see Beebe, 2006). Here, we are simply using a common conception of his view as an example to illustrate a point.

neuroscientists alike (see Haldane, 1998; Koons and Bealer, 2010; van Gulick, 2017). Especially in such a context, it is to the benefit of NCC research if the definition of NCC avoids tethering NCC research to a particular view of consciousness<sup>4</sup>. In this regard, Koch's remarks hit home: "Note that the NCC themselves are neutral from the point of view of physicalism/materialism or one of the various shades of dualism. Under any reading, consciousness will have physical correlates."<sup>5</sup> For that to be true, the definition of NCC must be theoretically neutral with respect to metaphysical views about the nature of consciousness.

The definition of NCC should also be theoretically neutral in another sense. It should not entail any particular empirical neurobiological conclusion about NCC, such as whether NCC is predominantly in the front or back of the brain (see Boly et al., 2017). It should leave such empirical issues open for empirical investigations to settle. This will not only assure that empirical matters are left to be decided by empirical research, but also prevent NCC research as a whole from being tethered to a particular empirical hypothesis. Having clarified the desiderata for a good definition of NCC (see Table 1), let us turn to Chalmers's definition.

## Chalmers's Definition

Before providing his own definition (Chalmers 2000, p. 17–18) first considers a basic definition of an NCC: "A neural system N is an NCC if the state of N correlates directly with states of consciousness." This definition elicits three questions that guide Chalmers's (2000, p. 18) definition:

1. What conscious states are relevant?
2. What neural states are relevant?
3. What does it mean for the neural states to correlate directly with conscious states?

Regarding question 1 Chalmers (2000, pp. 18–23) surveys multiple classes of phenomenal consciousness considered in the NCC literature<sup>6</sup>. The first is being conscious vs. not being conscious. The second is a background state of consciousness that is an overall state of consciousness at a particular time. The third is contents of consciousness such as perceiving Theresa May's face on the television screen. The fourth is arbitrary phenomenal properties which can include specific states of any of the above classes and might be useful for a general definition of an NCC that applies to all classes of phenomenal consciousness.

Regarding question 2, the relevant neural states can depend on whether the neural correlate is the so-called "full NCC" or a "content-specific NCC" (Koch et al., 2016a, p. 308). Content-specific NCC are neural correlates of particular conscious states that include content. By contrast Koch et al. (2016a, p. 308) define the full NCC as: "The neural substrate supporting experience in

general, irrespective of its specific content." The full NCC can be the NCC of simply being conscious or the NCC of one's overall conscious experience that may include multiple content specific conscious states.

Regarding question 3 Chalmers (2000, p. 24–28) asks two more fundamental questions about the nature of the direct correlation:

1. Must the neural state be necessary, sufficient, or necessary and sufficient for the conscious state it is correlated with?
2. Must the correlation hold across all cases or only across specific types of cases (i.e., cases with ordinary brain function in an ordinary environment, cases with a normal brain but unusual inputs, cases with varying stimulation, or cases with abnormal brain function due to lesions)?

As for question 3.1 Chalmers (2000, p. 32) reasons that an adequate definition of an NCC should not rule out the possibility of a conscious state having different neural correlates. Prior to empirical investigation, we do not know whether there is multiple realizability within a single human brain. Perhaps someone can have the conscious state of perceiving a face that corresponds with the activity of one coalition of neurons,  $N^1$ , then lose that coalition, after which the individual's perception of a face will correspond to the activity of another coalition of neurons,  $N^2$  (see Koch, 2004, p. 97 footnote 24). Whether this is possible is not the point<sup>7</sup>. Rather, the point is that whether or not it is possible should be decided by empirical neuroscientific investigation, as opposed to a pre-empirical, *a priori* definition. But if NCC are, by definition, *necessary* for the corresponding conscious state then such a scenario would be ruled out by definition. Consequently, the definition would lack theoretical neutrality. Hence Chalmers (2000, p. 32) does not define NCC as "necessary" but rather as "sufficient" (see also Koch, 2004, p. 97 footnote 24).

As for question 3.2 Chalmers (2000, pp. 31) thinks that the conditions most relevant to NCC include normal brain function that can allow for some atypical inputs and brain stimulation as long as there are no changes to brain structure. He points out that lesion studies are repeatedly used to make inferences about neural correlates but reasons that such methodology is flawed since "the identity of an NCC is arguably always relative to specific brain architecture and normal brain functioning, and correlation across abnormal cases should not generally be expected" (2000, pp. 32). Others might agree that caution is necessary, but think that lesion studies can nevertheless provide important information especially when the findings from a lesion study are coupled with and corroborated by other studies, such as stimulation studies in a healthy brain (see Koch et al., 2016a, p. 308)<sup>8</sup>.

Having considered the above questions, and given the aforementioned desiderata Chalmers (2000, p. 31) defines a neural correlate of consciousness as follows<sup>9</sup>.

<sup>4</sup>Metzinger (2000a, pp. 4–5) discusses relevant neutrality.

<sup>5</sup>Personal correspondence 26 December 2016.

<sup>6</sup>Phenomenal consciousness pertains to the felt sensation of what it is like to be in a particular conscious state, whereas access consciousness pertains to the accessibility of a conscious state for active use in reasoning, recall, speech, and the like (see Block, 1995, pp. 230–232). Phenomenal consciousness can be (and some would say always is) access consciousness as well.

<sup>7</sup>See Muñoz-Céspedes et al. (2005), Jensen and Overgaard (2011), Mogensen (2011), Overgaard and Mogensen (2011).

<sup>8</sup>Owen is indebted to Christof Koch for this point.

<sup>9</sup>This general definition can be modified to fit specific classes of phenomenal consciousness (see Chalmers, 2000, p. 31).



**TABLE 1 |** NCC Definition Desiderata.

Reflects standard Use	Reasonable and coherent	Theoretically neutral
The definition comports with how the term is used within the NCC research community.	<p><i>Internal coherence:</i> The definition is internally coherent in that it avoids being internally contradictory.</p> <p><i>External coherence:</i> The definition is coherent with information relevant to NCC research.</p>	<p><i>Metaphysical neutrality:</i> The definition does not entail a particular metaphysical view about the nature of consciousness.</p> <p><i>Empirical neutrality:</i> The definition does not entail a particular empirical theory about the nature of specific NCC.</p>

*An NCC is a minimal neural system N such that there is a mapping from states of N to states of consciousness, where a given state of N is sufficient, under conditions C, for the corresponding state of consciousness.*

An NCC is a minimal neural system that is sufficient under certain conditions for the corresponding state of consciousness (see Koch et al., 2016a, p. 307). While Chalmers does not include the notion of necessity, he does include sufficiency, which is analyzed in section “Analyzing Sufficiency” Yet, before focusing on that particular facet of the definition it will be helpful to briefly clarify other aspects<sup>10</sup>.

The phrase “*minimal neural system*” prevents neural processes that enable the specific neural correlate from being included as part of the NCC (Chalmers, 2000, p. 24). With regards to NCC, we are not concerned with everything taking place in the nervous system when one is in a particular conscious state but rather the minimal neural states that correspond to a particular conscious state. The qualifier “under conditions C” allows for the fact that there are enabling conditions that allow NCC to function properly. The phrase “there is a mapping from states of N to states of consciousness” pertains to subjects across a species, not just an individual member. However, this mapping across a species is not necessarily a correspondence of identical neural states in every subject. The search for NCC is a search for biological regularities, not necessarily identical correspondence relations (for no two brains, not even of twins or even clones, are exactly alike)<sup>11</sup>. Biological regularities of all kinds permit some variation. Furthermore, there are phenomenological variations to consider. If shown an image of the Eiffel Tower, two men could report seeing the image, but the one who met his wife at the tower will likely have a different overall conscious experience as he perceives the image than the one who has never been to Paris. Given such variations, we should not expect the search for NCC to reveal correlations that have zero variation across a species, but rather neurobiological correlations reflecting biological regularities<sup>12</sup>.

To summarize: an NCC, as defined by Chalmers (2000, p. 31), is a minimal neural system that is *sufficient* under certain conditions for the corresponding state of consciousness. Next, the sense in which NCC are sufficient will be considered.

## ANALYZING SUFFICIENCY

Striking a good balance between clarity and complexity, Chalmers’s definition has served NCC research well for nearly two decades. Yet there remains a need for further clarification regarding the sense in which neural correlates are *sufficient* for consciousness. For just as there are various senses of necessity corresponding to different types of necessity, there are various senses of sufficiency corresponding to different types of sufficiency.

In this section, the merit of defining NCC according to three relevant senses of sufficiency is considered in light of the desiderata for a definition of NCC. First logical sufficiency is considered, then metaphysical sufficiency, and finally physically sufficiency. For the purpose of this article we analyze the definition of NCC in light of all three while acknowledging that philosophers disagree on whether metaphysical sufficiency is a distinct type of sufficiency. This issue, however, does not affect our final conclusion.

### Logical Sufficiency

Logical sufficiency is fundamentally due to logical entailments, as opposed to the nature of entities or natural regularities. In order to further grasp the concept of logical sufficiency, let us first consider logical necessity, which logical sufficiency mirrors. *X* is *logically necessary* for *Y* if, and only if, *Y* is *logically impossible* without *X*. The claim that Hilbert is a married bachelor is not logically impossible because Hilbert cannot find a date, but rather because a bachelor is by definition unmarried. Therefore, it is logically necessary that Hilbert be unmarried given that he is a bachelor. And it is logically impossible for Hilbert to be married given he is a bachelor. After all, given the definition of a bachelor, the propositions “Hilbert is a bachelor” and “Hilbert is married” entail Hilbert is not married and married, which is a contradiction (assuming the propositions refer to the same person at the same time).

Relying on the concepts of logical necessity and logical impossibility, let us describe logical sufficiency in the present context as follows:

*X is logically sufficient for Y if, and only if, given X then Y is logically necessary and not-Y is logically impossible.*

Once again, the concept can be elucidated *via* examples. The fact that Hilbert is a bachelor is logically sufficient for the fact that Hilbert is unmarried since a bachelor is by definition unmarried. And given that red is by definition a color, the bench being red is logically sufficient for the bench being colored.

<sup>10</sup>For further clarification, see Owen (2019b, pp. 242–245).

<sup>11</sup>The qualifier regarding clones is an extrapolation from what is known about cloned mice.

<sup>12</sup>Owen is indebted to Koch here too.

Now we can consider what it would mean for a neural correlate of consciousness to be logically sufficient. Neural state  $N^1$  is logically sufficient for the conscious state  $C^1$  if, and only if, given  $N^1$  then  $C^1$  is logically necessary and  $C^1$  not manifesting is logically impossible. While the common example of firing C-fibers as the neural correlate of pain is empirically inaccurate, we can use it to provide elucidation in the present context. Suppose the definition of “firing C-fibers” logically entailed the conscious state of pain. Given that and the presence of firing C-fibers, it would follow out of logical necessity that pain is present. There would be no noncontradictory way around it—if C-fibers are firing, pain is present—assuming that the definition of firing C-fibers entails pain so firing C-fibers are logically sufficient for pain. If NCC were logically sufficient for corresponding consciousness, there would be profound implications.

To begin with, it would significantly affect the methodology in the search for NCC. If NCC were logically sufficient, identifying the conscious state corresponding to a neural correlate could be a matter for *a priori* philosophical investigation devoid of any *a posteriori* empirical investigation. For clarification, “*a priori* knowledge” can be known by reason alone apart from experience. One can know *a priori* that  $2 + 2 = 4$ , or that  $Q$  is true given that “if  $P$  is true then  $Q$  is true” and “ $P$  is true.” By contrast, “*a posteriori* knowledge” is gained through our sense experience of the world around us.

Consciousness research includes both types of reasoning (Guta, 2019); and some things can be known by *a priori* reasoning alone, whereas other things require empirical investigation. For an example of the former, if we know that a subject has a content specific conscious state, such as seeing a face, then the subject must have an overall background state of being conscious. No empirical investigation is needed to know this; it follows as a logical entailment that the subject is conscious from the fact that the subject is conscious of a face. A content-specific conscious state is logically sufficient for being conscious, so we can simply deductively reason to the conclusion that the subject is conscious. By contrast, consider how researchers identify the conscious state that corresponds to a particular NCC. They must know what state the neurons of the brain are in and what conscious state a subject is in. Using functional brain imaging researchers can observe neuronal activity in a subject’s brain, and through reports from a subject, researchers can know a subject’s conscious state. To know what neural state correlates with the conscious state of perceiving a face, researchers try to identify what particular neural state(s) is present when the subject perceives the face. The process of identifying the correspondence relation between the conscious state and its neuronal footprint depends on this empirically gathered information, gained from brain imaging and listening to reports<sup>13</sup>.

However, if NCC were logically sufficient, the methodology could be very different. In principle, researchers could logically deduce what conscious state was present simply by knowing what neural state is present. The corresponding conscious state could be known by deductive reasoning *via* logical entailments if the

neural state were logically sufficient. There would be no need for verbal reports to discern what conscious state the subject is in, or any other empirically identifiable physiological indicator (see Tsuchiya et al., 2015). In other words, what conscious state corresponds could be known *a priori* by following logical entailments and rules of logic, given knowledge of what neural state is present. The correspondence would not need to be empirically discerned but could be known *a priori* by deductive reasoning alone.

But such *a priori* methodology, which would depend on NCC being logically sufficient, does not comport with the *a posteriori* empirical methodology actually used by researchers to identify the correspondence between neuronal and conscious states. The methodology used presupposes that the correspondence cannot be logically deduced, but is rather known through an empirical process. At present, the process involves using brain imaging to identifying neural states present when conscious states are present, which is known from verbal reports or physiological indicators. There is room in this process for inductive, abductive, and deductive reasoning. But it is not a process of purely deductive reasoning along logical entailments from knowledge of the neural states present, which it could be if NCC were logically sufficient for corresponding consciousness. Thus, defining NCC as logically sufficient does not comport with the concept of NCC informing the methodology NCC researchers actually use.

A second problem pertains to the second desideratum. If “sufficient” means “logically sufficient” the definition would not be internally incoherent, but it would suffer from being externally inconsistent with the logical possibility of zombies. Chalmers (1996, p. 94) describes a zombie as something physically the same as a conscious being such as himself, but with no conscious experience. Christof Koch and Crick (2001) applied a rendition of this concept to what they call “zombie agents”; that is, sensorimotor neural systems in the primate brain that can be integral to actions but nevertheless unconscious; and as they clarify: “By “unconscious” we mean any neuronal activity that does not give rise to conscious sensation, thought or memory” (Koch and Crick, 2001, p. 893)<sup>14</sup>. The existence of zombie agents is important to the search for NCC because it raises a key question: what is the difference that makes the difference between neuronal activity that is sufficient for consciousness and that which corresponds to the unconscious (Crick and Koch, 2003, p. 120; Koch, 2012, p. 90)?

What matters for the purported logical sufficiency of NCC is not whether zombies are real or physically possible nor metaphysically possible, but whether they are merely logically possible. They are logically possible if the concept of a zombie is logically consistent. Philosophers often articulate modal claims regarding possibility and necessity in terms of “possible worlds” semantics. Accordingly, zombies are logically possible if a possible world with zombies is logically coherent, devoid of logical contradiction. While such a world might be very different than the actual world and have different laws of nature, it does not seem logically inconsistent. Hence, it is often thought by philosophers that zombies are at least logically possible,

<sup>13</sup>For various reservations about this inquiry, see Hardcastle and Stewart (2009), Shulman (2013) and Tsuchiya et al. (2015).

<sup>14</sup>See also Koch (2004, ch. 12), and Koch (2012, ch. 6).

regardless of whether they are physically possible in the actual world. Here's the implication for a definition of NCC: if neural correlates are logically sufficient for the corresponding consciousness, then zombies are logically incoherent, but it does not seem that they are logically incoherent. Therefore if "sufficient" in the definition of NCC means "logically sufficient," the definition is at risk of failing to meet the desiderata of being both internally and externally coherent. The definition would create an unfortunate conflict: if one thought zombies are logically possible, they would be inclined to think the concept of NCC is not coherent and *vice versa*.

Since talk of zombies might sound trivial, let us consider another example related to neuroethics. In 2011, at the Institute of Molecular Biotechnology in Vienna a postdoctoral researcher, Madeline Lancaster, inadvertently brought about the production of a brain organoid from human embryonic stem cells (Willyard, 2015, p. 520). The brain organoids neuroscientists can now grow consist of several million neurons. In April 2018, *Nature* published an article on the ethics of experimenting with brain organoids. The team of authors, led by Farahany et al. (2018, p. 430), offered the following description of brain organoids:

*Brain organoids can be produced much as other 3D multicellular structures resembling eye, gut, liver, kidney and other human tissues have been built. By adding appropriate signaling factors, aggregates of pluripotent stem cells (which have the ability to develop into any cell type) can differentiate and self-organize into structures that resemble certain regions of the human brain.*

These so-called "mini-brains" resemble human brains in noteworthy ways regarding their constitution, neural activity, and structure. Thus, they prompt a key question with serious ethical implications: are they conscious? The question is natural to ask because it seems possible that despite them being composed of human brain tissue, and having similar structural features, and neural activity, it is possible they are not conscious. To use Chalmers's terminology, it's possible they are "zombie" mini-brains.

Brain organoids are much smaller and far less developed than actual brains, but imagine a scenario in which a full human nervous system is grown. It would still seem reasonable, if not all the more pertinent, to ask: is it conscious? And the question is sensible because it seems at least logically possible that such a nervous system would not be sufficient for consciousness. Yet, it would seem that such would not even be logically possible if NCC are logically sufficient for consciousness. One could appeal to the clause in Chalmers's definition "under conditions C" and claim that the neural correlates in such a nervous system would be logically sufficient for consciousness under the correct conditions. This is a fair point, but it actually highlights the relevance of zombies. A zombie would meet all the appropriate physical conditions and yet it still seems logically possible for consciousness to be lacking despite the presence of sufficient NCC under the correct physical conditions. And this logical possibility suggests that claiming NCC are logically sufficient for corresponding consciousness, given suitable physical conditions, is misguided.

Thus, the mere logical possibility of zombies makes it difficult for the definition of NCC understood according to logical sufficiency to meet the second desiderata of cohering with relevant external information—i.e., zombies are logically possible.

Before concluding this section, it is also worth noting how the definition would fail to be theoretically neutral. As eluded to above Chalmers (2000, p. 32) does not define NCC as *necessary* for corresponding consciousness in order to leave open the possibility of a conscious state having different neural correlates (see Koch, 2004, p. 97 footnote 24). This is a matter to be decided by empirical investigation, not by definition and *a priori* reasoning based on the meaning of the definition. The same is true for whether or not an NCC could be the correlate of multiple conscious states. The former issue pertains to the possibility that a conscious state could correlate with multiple neural states. For example, could pain possibly correlate with C-fibers firing and another neural process involving different neurons rather than C-fibers? This issue we are focusing on now is the reverse, the possibility that a neural state could correlate with multiple conscious states. For example, could C-fibers firing correlate with pain and another conscious state rather than pain? This possibility could be conceptually ruled out with no need for empirical investigation if NCC were by definition logically sufficient for the corresponding conscious state. For if C-fibers firing were the NCC of pain and therefore logically sufficient for pain, then C-fibers firing would necessarily entail the correlated conscious state—pain. As a result, it would not be possible for C-fibers firing to be correlated with some other conscious state instead of pain. It could be conceptually ruled out, but it seems this is an issue that should be empirically settled. However, if NCC are defined as logically sufficient for corresponding consciousness then such issues could be settled pre-empirically, so the definition would lack empirical neutrality.

In sum, if "sufficient" in the definition of NCC means logically sufficient the definition would have problems with each desiderata (see Table 2). With respect to the first desiderata of reflecting standard usage, it would be inconsistent with the concept of NCC that informs the empirical methodology NCC researchers use. Regarding the second desiderata of being reasonable and coherent, it would lack external coherence insofar as the concept of NCC would be inconsistent with the logical possibility of zombies. And as for theoretical neutrality, the definition would logically negate the possibility that a neural correlate might correlate with multiple conscious states which should be left to empirical investigation to confirm or deny. Thankfully, for Chalmers's definition, there are two other relevant senses of "sufficient."

## Metaphysical Sufficiency

The second relevant sense of sufficiency worth considering is metaphysical sufficiency, which pertains to the natures of things. Once again, it will help to first clarify metaphysical necessity, which metaphysical sufficiency mirrors. *X* is *metaphysically necessary* for *Y* if, and only if, the nature of *X* requires *Y*.

Regarding freewill, someone might claim a free action requires the ability to do otherwise since it is essential to the nature of a free action. On this view, the ability to do

otherwise could be described as metaphysically necessary for freewill. In philosophy of mind, some say mental causation is impossible for a nonphysical mind due to the nature of causation, which requires spatial extension (for example see Kim, 2005, ch. 3). Spatial extension is allegedly metaphysically necessary for causation and thus mental causation is *metaphysically impossible* for nonphysical minds that are not spatial. Of course, there are substantive replies (see Bailey et al., 2011; Owen, 2018b). Nevertheless, these examples are helpful for introducing the concepts of metaphysical necessity and impossibility, which can be used to describe metaphysical sufficiency in the present context:

*X is metaphysically sufficient for Y if, and only if, given X then Y is metaphysically necessary and not-Y is metaphysically impossible due to the natures of X and Y.*

Metaphysical sufficiency is relevant to contemporary philosophical discussions about powers. For example, Nancy Cartwright and Pemberton (2013) argue that Aristotelian powers that produce specific effects due to their natures are central to modern science. Accordingly, a chemist discerns that H<sub>2</sub>O has the power to dissolve NaCl whenever the conditions are adequate (e.g., salt is placed in a cup of water). H<sub>2</sub>O causes NaCl to dissolve with regularity since the manifestation of this power is metaphysically sufficient for the effect. Given the nature of the power, when it is manifested in the appropriate conditions an effect of a particular nature follows.

Let us now apply metaphysical sufficiency to NCC. Neural state N<sup>1</sup> is metaphysically sufficient for conscious state C<sup>1</sup> if, and only if, given N<sup>1</sup> then C<sup>1</sup> is metaphysically necessary and C<sup>1</sup> not manifesting is metaphysically impossible due to the natures of N<sup>1</sup> and C<sup>1</sup>. If Chalmers's definition of NCC is understood according to metaphysical sufficiency it would be better off than if it were understood according to logical sufficiency. For one, it would seem to comport more with what NCC researchers have in mind when they are looking for neural correlates. After all, researchers are concerned with what natural characteristics of particular neurons correspond to consciousness, which might pertain to something specific about their nature such as their morphological structure, molecular compositions, biophysical properties, or function (i.e., spiking rates, oscillation rhythms, spike synchronization, etc.). The definition would also be internally consistent.

Despite its strengths, however, the definition would still suffer from shortcomings that could be avoided. The zombies discussed in the previous section also raise trouble for NCC defined in terms of metaphysical sufficiency. To begin with, if zombies are not merely logically possible but also metaphysically possible, the definition of NCC would be inconsistent with such a possibility and therefore lack external consistency. For if a neural state N<sup>1</sup> were metaphysically sufficient for conscious state C<sup>1</sup> then it would be metaphysically impossible for N<sup>1</sup> to be present without C<sup>1</sup>. But such would be metaphysically possible if zombies were metaphysically possible. To the definitions benefit, the metaphysical possibility of zombies is more debatable than the logical possibility of zombies and more difficult to demonstrate.

Nevertheless, the definition of NCC would remain vulnerably dependent on the outcome of such debates if it is inconsistent with one side in the debate.

The definition would also lack theoretical neutrality in two ways. First, it would lack metaphysical neutrality because it would be inconsistent with versions of dualism that permit the metaphysical possibility of zombies. While this concern also pertains to the metaphysical possibility of zombies, it is not the same concern as that just mentioned above regarding external inconsistency. For that issue depends on the metaphysical possibility being a real and demonstrated possibility. This concern regarding metaphysical neutrality does not depend on the possibility being real or demonstrated, but rather on the fact that the possibility would be ruled out in virtue of how NCC are defined; and such an issue should not be decided by how NCC are defined. Yet more importantly for present purposes, this would entail a significant burden—the need to demonstrate that zombies are not metaphysically possible—in order to justify the metaphysical possibility of NCC and thus legitimate the search for NCC.

The second way the definition would lack theoretical neutrality pertains to empirical neutrality. In the previous section, we argued that the definition of NCC understood according to logical sufficiency would rule out the possibility that a neural correlate could possibly correlate with a different conscious state, which is an empirical matter that should be decided by empirical investigation rather than by definition. We gave the example of C-fibers firing correlating with pain and another conscious state rather than pain. The charge was that the definition would lack empirical neutrality since it would rule out this possibility prior to empirical investigation. The same would be true if the definition were understood according to metaphysical sufficiency. For if a neural state N<sup>1</sup> were metaphysically sufficient for conscious state C<sup>1</sup> then C<sup>1</sup> would necessarily follow given the presence of N<sup>1</sup> and it would be impossible for another conscious state C<sup>2</sup> to follow instead of C<sup>1</sup>. This would be the case in every metaphysically possible world, including the actual world. Yet, again, whether N<sup>1</sup> could be the NCC of either C<sup>1</sup> or C<sup>2</sup> should be decided by empirical investigation, not by definition.

All things considered, if “sufficient” in Chalmers's definition meant “metaphysically sufficient” then the definition would be better off than if it meant “logically sufficient” (see Table 3). This is because the definition would comport well with standard usage and enjoy internal consistency. However, the definition would be inconsistent with the metaphysical possibility of zombies and it would lack empirical neutrality, two shortcomings that are best to avoid. Yet, there is a third sense of sufficiency worth considering.

## Physical Sufficiency

The third relevant sense of sufficiency is physical sufficiency. The natural and physical sciences focus on physical necessities and sufficiencies. In some respects, physical sufficiency is easier to grasp than the previous types of sufficiency since we are very familiar with it. Nevertheless, some precision is needed to clarify exactly what is meant by *physical sufficiency*.



**TABLE 2** | NCC defined according to logical sufficiency.**Logically sufficient NCC**

Neural state  $N^1$  is logically sufficient for conscious state  $C^1$  if, and only if, given  $N^1$  then  $C^1$  is logically necessary and  $\neg C^1$  is logically impossible.

Reflects standard use	Reasonable and coherent	Theoretically neutral
Lacks consistency with the concept of NCC that informs the empirical methodology NCC researchers employ.	Lacks external coherence as it is inconsistent with the logical possibility of zombies.	Lacks empirical neutrality since it logically rules out prior to empirical investigation the possibility of a neural state correlating with two different conscious states.

**TABLE 3** | NCC defined according to metaphysical sufficiency.**Metaphysically sufficient NCC**

Neural state  $N^1$  is metaphysically sufficient for conscious state  $C^1$  if, and only if, given  $N^1$  then  $C^1$  is metaphysically necessary and  $\neg C^1$  is metaphysically impossible due to the natures of  $N^1$  and  $C^1$ .

Reflects standard use	Reasonable and coherent	Theoretically neutral
Consistent with standard usage since researchers try to identify what characteristics of neurons are sufficient for consciousness, which might pertain to their <i>whole nature</i> or <i>something specific about their nature</i> such as their morphological structure, molecular compositions, biophysical properties, or function.	Lacks external coherence as it is inconsistent with the metaphysical possibility of zombies.	Lacks metaphysical neutrality since it rules out metaphysical views of consciousness that entail the metaphysical possibility of zombies. Lacks empirical neutrality, since it rules out prior to empirical investigation the possibility of a neural state correlating with two different conscious states.

In a reductivist milieu, “physical” is often taken as a referent to something regarding fundamental physics. But we do not here assume a reductivist view of physical objects. Subatomic particles are physical objects as are atoms, we assume, and likewise for trees, mammals, bodily organs, and cells. And physical laws pertaining to micro objects like the Higgs boson but also to macro objects like cells, organs, mammals, and trees. Laws of physics as well as biological laws are included<sup>15</sup>. Some might say we have neglected a relevant type of sufficiency that must be considered, namely nomological sufficiency, which focuses on natural laws. However, nomological sufficiency is not specifically addressed here since our view of physical sufficiency encompasses nomological sufficiency.

In describing physical sufficiency, once again, it will be helpful to begin with physical necessity, which physical sufficiency mirrors.  $X$  is *physically necessary* for  $Y$  if, and only if,  $X$  is required for  $Y$ , which is *physically impossible* without  $X$ , due to physical laws. For example, oxygen is necessary for human life, which is physically impossible without oxygen, due to physical laws. As for physical sufficiency, we can describe it in this context as follows:

*$X$  is physically sufficient for  $Y$  if, and only if,  $X$  satisfies the requirements of physical laws for  $Y$ .*

If one physical event is physically sufficient for a second event, the second event follows with consistent regularity due to physical laws. A temperature of zero degrees Celsius is sufficient for  $H_2O$  to change from a liquid state to a solid state. The structure of a Boeing 747 together with forces exerted by its fully functioning CFM-56 jet engines

are sufficient for the airliner to fly through the atmosphere. Understanding specific physical conditions that are physically sufficient to bring about a particular kind of effect is vital to technological development, in which we intend to bring about specific kinds of effects with consistent regularity. It is also vitally important to the development of medical treatments where it is desired to eliminate particular diseases by eliminating the physical conditions physically sufficient for the disease.

Now let's apply physical sufficiency to NCC. Neural state  $N^1$  is a neural correlate physically sufficient for conscious state  $C^1$  if, and only if,  $N^1$  satisfies the requirements of the physical laws in the actual world for  $C^1$ . One significant benefit of understanding the sufficiency of neural correlates in terms of physical sufficiency pertains to the first desiderata. It clearly comports with how NCC researchers use the term. Correspondingly, it also fits with how the research in the search for NCC is carried out in an empirical manner as opposed to simply relying on logical deductions. For the search for NCC is aimed at identifying the physical states and processes at the level of neurons and coalitions of neurons that support consciousness. And these must be identified *via* rigorous empirical investigation in which certain neural candidates are postulated as NCC and empirically tested by mapping the conscious states of subjects with neural states identified *via* brain imaging. Sometimes these empirical postulates are verified and sometimes they are falsified.

What matters in the search for NCC are physical facts about the neural states sufficient for consciousness that are known *via* empirical investigation; they are not logically deducible *via a priori* reasoning alone. Given this, it is not surprising that Koch (2004, p. 205), who co-instigated the contemporary search for NCC with Crick<sup>16</sup>, would write:

<sup>15</sup>By ‘law’ we merely mean consistent regularities throughout the physical world. There are various ways of ontologically describing and accounting for such regularities (see Hume, 2007, section VII; Heil, 2012, ch. 6).

<sup>16</sup>See Crick and Koch (1990).

**TABLE 4** | NCC defined according to physical sufficiency.**Physically sufficient NCC**

Neural state  $N^1$  is a neural correlate physically sufficient for conscious state  $C^1$  if, and only if,  $N^1$  satisfies the requirements of the physical laws in the actual world for  $C^1$ .

Reflects standard use	Reasonable and coherent	Theoretically neutral
Consistent with standard usage since researchers try to identify what characteristics of neurons in the actual world given its physical laws are sufficient for consciousness, which might pertain to their whole nature or something specific about their nature such as their morphological structure, molecular composition, biophysical properties, or function (i.e., spiking rates, oscillation rhythms, spike synchronization, etc.).	In addition to being internally consistent, the definition is externally consistent with the logical and metaphysical possibility of zombies. And “zombie agents” can conceptually aid the search for NCC interested in the physical differences that matter to consciousness.	Theoretically neutral as it is compatible with a range of views regarding the metaphysics of consciousness. Empirically neutral as it leaves open for empirical investigation the possibility of a neural state correlating with multiple, different conscious states.

*It does not feel like anything to be a zombie...Some argue that the logical possibility of their existence implies that consciousness does not follow from the natural laws of the universe, that it is an epiphenomenon. From this point of view, whether or not people feel makes no difference to themselves, to their offspring, and to the world at large. To Francis and me, this point of view seems sterile. We are interested in the real world, not in a logically possible never-never land where zombies roam. And, in the real world, evolution gave rise to organisms with subjective feelings.*

If the search for NCC is a search for the neural states physically sufficient for consciousness, then it is not threatened by the logical or metaphysical possibility of zombies.

A world with zombies might be logically and metaphysically possible. But what NCC research is focused on is what neuronal characteristics are physically sufficient for consciousness given the physical laws present in the actual world, not all logically or metaphysically possible worlds. This search is compatible with both the logical and metaphysical possibility of zombies since it is focused on what is physically sufficient given the physical laws of nature. It is aimed at uncovering the physical neuronal states corresponding to consciousness with consistent regularity, the regularity of biological laws. Some neural states correspond with the absence of consciousness, which Koch and Crick (2001) call “zombie agents,” while others regularly correspond with the presence of consciousness due to the physical laws of nature. The aim of NCC research is to distinguish the latter from the former (see Crick and Koch, 2003, p. 120; Koch, 2012, p. 90). In sum, when the definition of NCC is understood according to physical sufficiency the logical and metaphysical possibility of zombies does nothing to threaten it, and the idea of “zombie agents” can be a useful conceptual tool.

The definition also enjoys theoretical neutrality regarding an empirical issue that it would lack if understood according to the previous two senses of sufficiency. It would not decide prior to empirical investigation whether a particular neural state  $N^1$  could systematically correlate with one conscious state  $C^1$ , or rather, a different conscious state  $C^2$ . This issue is left open for empirical investigation to decide, as it discerns what the physical laws of nature permit. In addition, the definition enjoys theoretical neutrality metaphysically speaking in that the existence of physically sufficient NCC is compatible with a range of views in the metaphysics of mind—from physicalist views like the identity theory, to hylomorphism, to Cartesian dualism (see

Block and Stalnaker, 1999; Swinburne, 2013; Owen, 2018a)<sup>17</sup>. Given this, the legitimacy of NCC research would be less affected by changing tides in philosophy of mind.

However, a word of caution is in order. It is easy to confuse physical sufficiency with a causal notion of neural correlation. Although a correlation is not necessarily indicative of causation. And if “sufficient” in the definition of NCC meant that every conscious state has a neural correlate that causes it to exist, then the definition would lack theoretical neutrality. At worst, it would entail epiphenomenalism, assuming the effect cannot be identical to its cause and that the neural state is always the cause of the conscious state. At best, but concerning nevertheless, the definition would prematurely rule out views that explain the correlations in terms of constitution, supervenience, grounding, or some other relation that might not be causal. So physical sufficiency should not be confused with physical causation, which is only one possible explanation of a physical sufficiency.

In the final analysis, if “sufficient” in Chalmers’s definition of an NCC is understood in terms of physical sufficiency, it would appear to meet each desiderata (see Table 4). It would fit common usage, be reasonable and coherent internally as well as externally, and be theoretically neutral both metaphysically and empirically.

## THEORETICAL IMPLICATIONS

The definition of NCC understood according to physical sufficiency meets all the desiderata for a good definition of NCC outlined in section Defining NCC. That said, NCC so defined might fail to meet all the expectations some consciousness researchers have for NCC.

A clear implication of NCC being physically sufficient is that, despite the pragmatic mental health benefits locating NCC can produce, the search for NCC will leave profound questions about consciousness unanswered. By themselves, NCC will not solve the hard problem of consciousness or tell us what its nature is (see Chalmers, 1995). Moreover, identifying neural correlates in humans we already know are conscious will not necessarily tell us what other candidates for being conscious are actually conscious. When NCC are properly understood it becomes apparent that

<sup>17</sup>Hence Frith and Rees (2017 p. 3) can claim Descartes “. . . was the first to think seriously about the neural correlates of consciousness” (see Koch, 2015).

the search for NCC is limited in its scope of what it can tell us about consciousness (see Chalmers, 1998; Koch, 2019, ch. 7; Owen, 2019b).

Many consciousness researchers are principally interested in the pragmatic benefits of identifying NCC. They want to identify neural states sufficient for undesired mental states such as anxious feelings, so we can treat mental illnesses by influencing their neural underpinnings. Such researchers can find the theoretical neutrality of NCC beneficial since it precludes NCC research from being tethered to issues like the hard problem and thus influenced by relevant debates. This can allow the search for NCC to march forward without being unduly hindered by slower progression in other areas of consciousness research.

That said, many consciousness researchers have additional aspirations, such as solving the hard problem and identifying empirical indicators of the presence of consciousness in unresponsive medical patients, animals, and perhaps machines. It is important to acknowledge the limitations of what NCC can tell us concerning such issues, to elucidate their epistemic role. Identifying the neural correlate of a human person's feeling of pain will not tell us what exactly pain is, why such feelings exist, whether unresponsive medical patients feel pain, nor whether an iPhone feels pain. However, that does not entail NCC are irrelevant to these issues. NCC data is not sufficient by itself to answer certain questions, but when coupled with a warranted theoretical framework, NCC data provides important information for further inferences or extrapolations.

In this section, we illustrate the applicability of a theoretically motivated approach to consciousness research by considering two particular issues. To begin with, we consider the utility of applying a theoretical framework to empirically discerning the presence of consciousness in human patients suffering from unresponsive wakefulness syndrome (UWS)<sup>18</sup>. Subsequently, we provide an example regarding Alzheimer's disease and the persistence of the self, which illustrates the implications of theoretical frameworks in other areas of research that are related to consciousness studies.

To be clear, we are not here advocating for any particular theory or model that we discuss in these examples. Our only aim is to illustrate the epistemic role theoretical frameworks can play in consciousness research, and how they can be applied to the limited NCC data in order to make further important inferences regarding consciousness. Once this is accepted, the necessary work that will remain is rationally adjudicating between theories and discerning which theory(s) is most likely true and therefore may be employed in the ways suggested. This will inevitably be a long-term task, which will become more viable as research technology and our understanding of the brain progress.

## Unresponsive Patients

There are cases in which patients with UWS come out of an unresponsive state and recall being conscious while previously completely unresponsive. Consider a situation in which a patient gives no behavioral indication of responding to external stimuli. Furthermore, suppose the patient gives no indication *via* brain

activity (e.g., neural activity in the motor cortex when the patient is asked to imagine playing tennis). It is possible that the patient, while unresponsive, is nevertheless conscious.

Since medical practitioners do not have access to the patients' first person perspective and the patient does not provide any way of verifying her consciousness from a third person perspective *via* a response, an empirical method of verification is needed. The neuronal footprint of consciousness in the form of the full NCC (i.e., the neural correlate of being conscious vs. not being conscious) becomes vitally applicable at this point. However, verifying an NCC in a healthy human subject depends on verbal reports or a corresponding physiological indicator that has been previously verified by a report from the conscious subject (see Tsuchiya et al., 2015); and if it is true that NCC are physically sufficient and zombies are metaphysically or logically possible, how can we know that a full NCC is indicative of consciousness when a patient has undergone brain damage and there is no way for the subject to give us any verification?

A theoretical framework might provide an answer, assuming that the framework is itself warranted. Suppose that there is neural activity in the brain of our unresponsive patient in the posterior cerebral cortex, amongst a coalition of neurons in a temporo-parietal-occipital area<sup>19</sup>. Suppose further that the neural activity is localized, there are no long-range action potentials sent from this coalition of neurons to other brain regions. Assume, however, that the activity is reciprocal and intrinsic within the coalition of neurons. To demonstrate the epistemic utility of a theoretical framework in such cases, let us contrast two theories—the GNW theory and the Integrated Information Theory—and what they would imply in this case based on the above stipulated neurobiological data.

The GNW theory of consciousness is a computational theory (see Dehaene et al., 2017, p. 492)<sup>20</sup>. The theory's originator, Bernard Baars, hypothesized that within the human brain there is a global workspace that houses information and makes it available to the specialized processing systems throughout the brain (see Baars, 1988). Since the workspace's capacity is limited, various information signals compete for the privileged position of being the globally available representation in the workspace. That which comes to occupy the workspace is conscious; the signals that do not make it into the workspace are not conscious. Contemporary proponents of GNW have applied the theory to the neurophysiology of the neocortex (see Dehaene and Changeux, 2011; Dehaene, 2014). Per GNW, an indicator of consciousness is a global broadcast of information involving the activity of a prefrontal-parietal network of long-range cortical neurons corresponding with activity in high-level sensory cortices that receive the broadcast. This makes the information globally available for various functional processes (e.g., speech, memory, action) and thus conscious content, according to GNW.

Given the neurobiological data stipulated above regarding the unresponsive patient, a prognosis guided by the GNW theory would declare the patient unconscious. After all, the localized

<sup>18</sup>Also called a vegetative state (see Laureys et al., 2010).

<sup>19</sup>Koch et al. (2016a,b) discuss a posterior cortical "hot zone" in this area.

<sup>20</sup>For a lucid summary of the theory which has informed our own synopsis here, as well as an objection to it, see Koch (2019, Ch. 13).

neural activity makes no long-range projections. Nothing about the description of the scenario suggests there is information being processed that is globally available to areas throughout the brain for corresponding function, which is required for consciousness, according to GNW. On the other, a prognosis based on the same data that is instead guided by the Integrated Information Theory of consciousness (for brevity IIT) might be more optimistic in this case.

IIT was developed by the Italian psychiatrist and neuroscientist, Giulio Tononi. It starts with five self-evident axioms about the nature of consciousness and infers corresponding postulates about the nature of the physical substrate of consciousness. In short, according to this theory, consciousness involves information that is integrated, and the physical substrate of consciousness is also integrated in that it exemplifies a structure in the central nervous system that exhibits a maximal intrinsic cause-effect power called *Phi* and symbolized by  $\Phi$ . This power manifested by the physical substrate consisting of a causal structure in the central nervous system is consciousness. Thus, given a causal structure that manifests an intrinsic causal power in the central nervous system, consciousness is present because it is the causal power being manifested, according to IIT. Some leading proponents of IIT aim to develop a consciousness meter capable of measuring a patient's level of consciousness by measuring the intrinsic causation manifested in the cortex (see Massimini and Tononi, 2018; Koch, 2019). The greater the  $\Phi$  measurement, the higher the level of consciousness. Likewise, a lower  $\Phi$  measurement indicates a lower level consciousness, and a negative measurement indicates unconsciousness. Yet as long as there is a positive  $\Phi$  measurement, which indicates intrinsic causation manifested in the cortex, consciousness is present.

Given the neurobiological data stipulated above regarding the unresponsive patient, a prognosis guided by IIT would be based on a measurement of the intrinsic causation manifested by the localized reciprocal neural activity. The reciprocal nature of the neural activity would be worth investigating to see if there is a manifestation of intrinsic causation present. If it turned out that there was some degree, however minor, of positive  $\Phi$  then consciousness is present, according to IIT; whereas according to GNW this would not be sufficient for consciousness.

This example illustrates the practical difference two different theoretical frameworks can make to a medical prognosis regarding consciousness in unresponsive human patients<sup>21</sup>. By extension, one could see how the theories might be extrapolated to empirically test for consciousness in other biological organisms with a nervous system similar to ours. The GNW theorist would look for a global workspace with information available to other brain regions for corresponding function,

whereas the IIT theorist would look for intrinsic causation. There is also interest in applying these theories to evaluating alleged machine consciousness (see Dehaene et al., 2017; Koch and Tononi, 2017). Certainly, applying such theories beyond biological nervous systems is more speculative at this point and caution is in order (see Owen, 2019a). Moreover, there is more immediate applicability pertaining to human consciousness beyond just discerning the presence of consciousness. There are also implications for neighboring fields of study. For an example relevant to psychology, let us consider theoretically informed ways of thinking about Alzheimer's disease and the persistence of the self.

## Alzheimer's and the Self

When medial temporal lobe structures vital to the formation and retention of long-term memories become dysfunctional because of plaque, tangles, and other molecular detritus of advanced dementia, autobiographical episodic memories of Alzheimer's patients are often irretrievably lost<sup>22</sup>. Does this loss of memory about one's past experiences also entail a loss of the self, according to which the person who once was discontinues? Different theoretical frameworks motivate different answers.

One answer is that the self discontinues when the neural correlates supporting conscious recall of autobiographical episodic memories no longer function properly and access to such memories is lost. This conclusion is motivated by a psychological continuity view of personal identity often attributed to John Locke and his widely influential work *An Essay Concerning Human Understanding* [see especially (Locke, 1975) Bk II, Ch. 27, sect. 19–20]. A person's ontological identity is thought to be grounded in the individual's psychological continuity of past experiences that requires conscious recall of autobiographical episodic memories. Therefore the loss of conscious access to such memories that is the result of dysfunctional neural correlates underpinning that access entails the loss of the self, or the person, that once was. The empirical data garnered from brain imaging about the neural correlates does not by itself justify such a conclusion, far from it. Rather, such a conclusion is inescapably theory loaded, based on the idea that psychological continuity is what grounds the persistence of the person.

However, the conclusion that the self is lost does not follow given a different theoretical framework such as the Mind-Body Powers model of NCC (Owen, 2018a). According to the human ontology informing the model, a human person is a substance with mental powers, such as the capacity to consciously recall autobiographical episodic memories, that requires the co-manifestation of particular bodily powers. The bodily powers that are the partner-powers of mental powers require adequate biological structures at the neuronal level that manifest the requisite bodily powers. The loss of or damage to neural tissue that manifest such bodily powers that must be co-manifested with particular mental powers for their full exercise can imply that the person loses their ability to naturally exercise the

<sup>21</sup>Here there are multiple related ethical questions that require careful consideration, which we are indebted to a reviewer for raising. For one, how much justification must a theory have before we allow it to inform medical decisions? When theories do play a role, how much preference, or weight, should be given to them in the decision-making process? When there are different potential diagnoses informed by different theories of consciousness, who should have the final say about whether a patient is or is not conscious? These are very important questions that we want to note here but will leave to be addressed by forthcoming works.

<sup>22</sup>Owen is indebted to Koch here, and we also benefited from insightful comments from a referee for this journal.



mental power. But this change in the substance—i.e., the human person—actually requires that the substance persist through the change, otherwise the substance would not actually be undergoing change but just cease to exist. This framework allows for the continuity of the person. The individual continues to exist, but takes on the attribute of being mentally impaired in that the individual cannot manifest a mental power she was once capable of manifesting (i.e., the conscious recall of autobiographical memories). A person who becomes paraplegic due to damage to their spinal cord persists despite losing a bodily power that required the neural architecture lost. Likewise, given the framework of the Mind-Body Powers model of NCC, an Alzheimer's patient can continue to exist despite a loss of the neural substrate needed to exercise a mental power.

It may turn out to be a case that the theoretical frameworks discussed in this section are false or require significant development and alterations. We are not here advocating for any particular theory. The utility of a theoretical approach in consciousness research and related fields is our concern. Given that NCC are physically sufficient for corresponding consciousness, there are limits to just how much the search for NCC will tell us about the nature of consciousness, what is conscious, and what consciousness indicates about human ontology. Therefore, a reliable theoretical framework can provide additional premises upon which we can reach further conclusions given the data verified *via* NCC research. Clearly, a crucial step at this point in consciousness research is discerning which theories are viable and therefore can be reliably applied to interpreting NCC data. This will require extensive research testing various theoretical frameworks on multiple grounds.

## PHILOSOPHICALLY POLLUTING CONSCIOUSNESS SCIENCE?

Before concluding, let us address a concern that those who think consciousness research must always begin with the empirical data might have about the suggested theoretical approach. The concern is that, whether the advocates of theories acknowledge it or not, theoretical frameworks include philosophical tenets that would be part of a theoretical starting point, and this could compromise the empirical rigor of the science of consciousness. For example, GNW appeals to a computational or functionalist view of consciousness that is a view in philosophy of mind (see Dehaene et al., 2017, p. 492). Moreover, IIT's starting axioms are allegedly directly known from the first-person subjective perspective and they are not known from a third-person perspective observing the brain (see Oizumi et al., 2014; Tononi et al., 2016). And the psychological continuity view of the self pertains to human ontology, while the Mind-Body Powers model appeals to an Aristotelian powers ontology (see Owen, 2018a, section 2).

We acknowledge that philosophical, pre-empirical ideas play an important role in this approach. However, the idea that consciousness research must begin with the raw empirical data unadulterated by philosophical principles is itself a pre-empirical, philosophical idea and it is difficult to see how it could be empirically validated. Thus, consciousness

research founded on this ideological foundation is foundationally informed by an epistemological principle about consciousness research methodology. It is guilty of the very charge its advocate levels against the theoretical approach to consciousness research. Yet, what matters is not whether philosophical tenets inform our consciousness research, but whether the tenets we allow to inform our research are rationally valid (see Guta, 2015). And if we do not acknowledge the role such tenets play in a research approach then we are more likely to be guided by unacknowledged, unanalyzed, and untested philosophical principles that can influence our interpretations of empirical data. By contrast, the theoretical approach makes explicit the ideological starting points that are informing our research methods, data acquisition, and interpretations. This not only makes the influence of such starting points more apparent but also makes it easier to scrutinize them and to evaluate the legitimacy of their influence. So those worried about philosophical tenets watering down empirical rigor in consciousness research would actually benefit from adopting the theoretical approach we are advocating.

On our view, proposed theoretical frameworks should be evaluated on multiple grounds. For one, the logical coherence of the framework must be rigorously analyzed. It is fiscally prudent to do this at the beginning of the evaluation process, since logical analysis of a theory's internal coherence is often less costly than empirical research requiring expensive technology. A thorough logical analysis of a theory might expose a fatal logical inconsistency, allowing researchers and funding bodies to avoid spending valuable resources on empirically testing an incoherent theory that's not possibly true (see Tahko, 2012, pp. 39–42). However, multiple competing theories can be internally coherent. Therefore, theoretical virtues used to evaluate the legitimacy of scientific theories in general should be used to evaluate competing theoretical frameworks in consciousness research. To use the example theories discussed above in sections Must the correlation hold across all cases or only across specific types of cases (i.e., cases with ordinary brain function in an ordinary environment, cases with a normal brain but unusual inputs, cases with varying stimulation, or cases with abnormal brain function due to lesions)?, researchers should consider whether GNW or IIT has more theoretical virtues. Does one theory have more explanatory scope or explanatory power? Is IIT or GNW simpler, in that one is more capable of explaining relevant data without postulating unnecessary entities? Which theory coheres best with widely accepted theories in related research fields? Such questions can be used to weed out inadequate theories with less theoretical virtues and to identify more promising, theoretically virtuous, theories.

After analyzing competing theories based on logical coherence and theoretical virtues, multiple theories might remain viable contenders, but empirical research can be used to confirm or falsify remaining candidates. The empirically testable predictions that are implied by a theory must be tested, with the outcomes confirming or disconfirming the theory. Although this immensely valuable work will be empirical, here too, logical and philosophical analysis is useful. It can clarify which predictions are strictly entailed by the theory, probably in light of the theory,

merely consistent with the theory, or simply an opinion of an advocate of the theory that's not part of the theory itself. Furthermore, philosophical analysis can clarify what aspects of consciousness are relevant to particular predictions and what the implications are. For example, philosophers often distinguish between access and phenomenal consciousness, which is relevant to the hypothetical case above involving an unresponsive patient and the applicability of IIT and GNW<sup>23</sup>. For GNW, one might think, concerns access consciousness whereas IIT concerns phenomenal consciousness and therefore this must be remembered when employing or comparing the theories and their predictions. One might even object to the use of IIT and GNW in our example on the grounds that the two theories are not about "consciousness" in the same sense. This is a fair point, which actually highlights the importance of philosophical analysis to a theoretical approach.

While this approach can potentially advance consciousness research, we make no claim that it will make it easier. In addition to arduous empirical testing, responsible theoretically motivated consciousness research requires rigorous philosophical analysis to ensure that logically coherent, theoretically virtuous, and empirically adequate theories are recognized as such, and the opposite are discarded.

## RESULTS

After analyzing three relevant senses of "sufficient" it appears that the definition of NCC is best understood according to

<sup>23</sup>On access versus phenomenal consciousness, see footnote 4 above.

## REFERENCES

- Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746. doi: 10.1016/j.neubiorev.2011.12.003
- Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. New York, NY: Cambridge University Press.
- Bailey, A. M., Rasmussen, J., and Horn, L. V. (2011). No pairing problem. *Philos. Stud.* 154, 349–360. doi: 10.1007/s11098-010-9555-7
- Bayne, T., and Hohwy, J. (2013). "Consciousness: theoretical approaches," in *Neuroimaging of Consciousness*, eds A. E. Cavanna, A. Nani, H. Blumenfeld and S. Laureys (Heidelberg: Springer), 23–35.
- Beebe, H. (2006). *Hume on Causation*. New York, NY: Routledge.
- Block, N. (1995). On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–247. doi: 10.1017/S0140525X00038188
- Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci.* 9, 46–52. doi: 10.1016/j.tics.2004.12.006
- Block, N., and Stalnaker, R. (1999). Analysis, dualism, and the explanatory gap. *Philos. Rev.* 108, 1–46. doi: 10.2307/2998259
- Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., and Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *J. Neurosci.* 37, 9603–9613. doi: 10.1523/JNEUROSCI.3218-16.2017
- Cartwright, N., and Pemberton, J. (2013). "Aristotelian powers: without them, what would modern science do?" in *Powers and Capacities in Philosophy: The New Aristotelianism*, eds R. Groff and J. Greco (New York, NY: Routledge), 93–112.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219.

physical sufficiency. Yet if NCC are physically sufficient for consciousness, there are limits to what NCC data alone can tell us in certain areas of consciousness research. A way of overcoming these limitations is to take a theoretically-motivated approach to certain issues. Once proven viable, a theory can provide additional information that can provide grounds for making further inferences in light of data about NCC. In the long-term, properly understanding NCC will strengthen the vitality of the search for NCC, enabling it to empirically progress unhindered by philosophical debates about the nature of consciousness. While there are limits to what this search can do, one thing it has done and will continue to do is foster scientific interest in consciousness studies.

## DATA AVAILABILITY

All datasets analyzed for this study are included in the manuscript and the supplementary files.

## AUTHOR CONTRIBUTIONS

MO formulated the argument and wrote the manuscript incorporating feedback provided by MG in substantial comments on a prior version.

## FUNDING

MO's research was funded by the Tiny Blue Dot Foundation.

- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York, NY: Oxford University Press.
- Chalmers, D. J. (1998). "On the search for the neural correlate of consciousness" in *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, eds S. R. Hameroff, A. W. Kaszniak and A. C. Scott (Cambridge, MA: MIT Press), 219–229.
- Chalmers, D. J. (2000). "What is a neural correlate of consciousness?," in *Neural Correlates of Consciousness*, ed. T. Metzinger (Cambridge, MA: MIT Press), 17–40.
- Chalmers, D. J. (2010). *The Character of Consciousness*. New York, NY: Oxford University Press.
- Crick, F., and Koch, C. (1990). Toward a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275.
- Crick, F., and Koch, C. (2003). A framework for consciousness. *Nat. Neurosci.* 6, 119–126. doi: 10.1038/nn0203-119
- Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. New York, NY: Penguin Books.
- Dehaene, S., and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018
- Dehaene, S., Lau, H., and Kouider, S. (2017). What is consciousness, and could machines have it? *Science* 358, 486–492. doi: 10.1126/science.aan8871
- Farahany, N. A., Greely, H. T., Hyman, S., Koch, C., Grady, C., Pasca, S. P., et al. (2018). The ethics of experimenting with human brain tissue. *Nature* 556, 429–432. doi: 10.1038/d41586-018-04813-x
- Fink, S. B. (2016). A deeper look at the "neural correlate of consciousness". *Front. Psychol.* 7:1044. doi: 10.3389/fpsyg.2016.01044
- Frith, C. D., and Rees, G. (2017). "A brief history of the scientific approach to the study of consciousness," in *The Blackwell Companion to Consciousness*, eds S. Schneider and M. Velmans (Oxford, UK: Wiley Blackwell), 3–16.

- Guta, M. P. (2015). Consciousness, first-person perspective and neuroimaging. *J. Conscious. Stud.* 22, 218–245.
- Guta, M. P. (2019). "Introduction," in *Consciousness and the Ontology of Properties*, ed. M. P. Guta (New York, NY: Routledge), 1–12.
- Haldane, J. (1998). A return to form in the philosophy of mind. *Ratio* 11, 253–277. doi: 10.1111/1467-9329.00070
- Hardcastle, V. G., and Stewart, C. M. (2009). "fMRI: a modern cerebrascop? The case of pain," in *The Oxford Handbook of Philosophy and Neuroscience*, ed. J. Bickle (New York, NY: Oxford University Press), 179–199.
- Heil, J. (2012). *The Universe As We Find It*. New York, NY: Oxford University Press.
- Hohwy, J. (2007). The search for the neural correlates of consciousness. *Philos. Compass* 2, 461–474. doi: 10.1111/j.1747-9991.2007.00086.x
- Hohwy, J. (2009). The neural correlates of consciousness: new experimental approaches needed? *Conscious. Cogn.* 18, 428–438. doi: 10.1016/j.concog.2009.02.006
- Hohwy, J., and Bayne, T. (2015). "The neural correlates of consciousness: causes, confounds and constituents," in *The Constitution of Phenomenal Consciousness: Toward a Science and Theory*, ed. S. M. Miller (Philadelphia, PA: John Benjamins Publishing Company), 155–176.
- Hohwy, J., and Frith, C. D. (2004). The neural correlates of consciousness: room for improvement, but on the right track: comment. *J. Conscious. Stud.* 11, 45–51.
- Hume, D. (2007). *An Enquiry Concerning Human Understanding*. ed. P. Millican (New York, NY: Oxford University Press).
- Jensen, M., and Overgaard, M. (2011). Neural plasticity and consciousness. *Front. Psychol.* 2:191. doi: 10.3389/fpsyg.2011.00191
- Kandel, E. R., and Hudspeth, A. J. (2013). "The brain and behavior," in *Principles of Neural Science*, 5th Edn. eds E. R. Kandel, J. R. Schwartz, T. M. Jessell, S. A. Siegelbaum and A. J. Hudspeth (New York, NY: McGraw-Hill), 5–20.
- Kim, J. (2005). *Physicalism, Or Something Near Enough*. Princeton, New Jersey, NJ: Princeton University Press.
- Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Englewood, Colorado: Roberts and Company Publishers.
- Koch, C. (2012). *Consciousness: Confessions of a Romantic Reductionist*. Cambridge, MA: The MIT Press.
- Koch, C. (2015). Two natural philosophers, centuries apart, converse about the mind. *Sci. Am. Mind* 26, 28–31. doi: 10.1038/scientificamericanmind0315-28
- Koch, C. (2019). *The Feeling of Life Itself*. Cambridge, MA: MIT Press.
- Koch, C., and Crick, F. (2001). The zombie within. *Nature* 411:893. doi: 10.1038/35082161
- Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016a). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22
- Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016b). Posterior and anterior cortex—where is the difference that makes the difference? *Nat. Rev. Neurosci.* 17:666. doi: 10.1038/nrn.2016.105
- Koch, C., and Tononi, G. (2017). Can we quantify machine consciousness? *IEEE SPECTRUM* Available online at: <http://spectrum.ieee.org/computing/hardware/can-we-quantify-machine-consciousness>. Accessed November 29, 2018.
- Koons, R. C., and Bealer, G. (Eds.) (2010). *The Waning of Materialism*. New York, NY: Oxford University Press.
- Laureys, S., Celesia, G. G., Cohadon, F., Lavrijsen, J., León-Carrión, J., Sannita, W. G., et al. (2010). Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. *Eur. J. Neurol.* 8:68. doi: 10.1186/1741-7015-8-68
- Locke, J. (1975). *An Essay Concerning Human Understanding*. Edited by P. Nidditch. Oxford: Oxford University Press.
- Massimini, M., and Tononi, G. (2018). *Sizing Up Consciousness: Towards an Objective Measure of the Capacity for Experience*. New York, NY: Oxford University Press.
- Metzinger, T. (2000a). "Introduction," in *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, ed. T. Metzinger (Cambridge, MA: The MIT Press), 1–12.
- Metzinger, T. (Ed.). (2000b). *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: The MIT Press.
- Mogensen, J. (2011). Reorganization of the injured brain: implications for studies of the neural substrate of cognition. *Front. Psychol.* 2:7. doi: 10.3389/fpsyg.2011.00007
- Mormann, F., and Koch, C. (2007). Neural correlates of consciousness. *Scholarpedia* 2:1740. doi: 10.4249/scholarpedia.1740
- Muñoz-Céspedes, J. M., Rios-Lago, M., Paul, N., and Maestu, F. (2005). Functional neuroimaging studies of cognitive recovery after acquired brain damage in adults. *Neuropsychol. Rev.* 15, 169–183. doi: 10.1007/s11065-005-9178-5
- Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput. Biol.* 10:e1003588. doi: 10.1371/journal.pcbi.1003588
- Overgaard, M. (2017). The status and future of consciousness research. *Front. Psychol.* 8:1719. doi: 10.3389/fpsyg.2017.01719
- Overgaard, M., and Mogensen, J. (2011). A framework for the study of multiple realizations: the importance of levels of analysis. *Front. Physiol.* 2:79. doi: 10.3389/fpsyg.2011.00079
- Owen, M. (2018a). Aristotelian causation and neural correlates of consciousness. *Topoi* 1–12. doi: 10.1007/s11245-018-9606-9
- Owen, D. (2018b). "Neo-Thomistic hylomorphism applied to mental causation and neural correlates of consciousness," in *Doctoral dissertation* (Birmingham, UK: University of Birmingham).
- Owen, M. (2019a). Exploring common ground between integrated information theory and aristotelian metaphysics. *J. Conscious. Stud.* 26, 163–187.
- Owen, M. (2019b). "Neural correlates of consciousness and the nature of the mind," in *Consciousness and the Ontology of Properties*, ed. M. P. Guta (New York, NY: Routledge), 241–260.
- Penfield, W. (1938). The cerebral cortex in man: the cerebral cortex and consciousness. *Arch. Neurol. Psychiatry* 40, 417–422. doi: 10.1001/archneurpsyc.1938.02270090011001
- Penfield, W. (1975). *The Mystery of the Mind: A Critical Study of Consciousness and the Human Brain*. Princeton, NJ: Princeton University Press.
- Shulman, R. G. (2013). *Brain Imaging: What It Can (and Cannot) Tell Us About Consciousness*. New York, NY: Oxford University Press.
- Storm, J. F., Boly, M., Casali, A. G., Massimini, M., Olcese, U., Pennartz, C. M. A., et al. (2017). Consciousness regained: disentangling mechanisms, brain systems, and behavioral responses. *J. Neurosci.* 37, 10882–10893. doi: 10.1523/jneurosci.1838-17.2017
- Swinburne, R. (2013). *Mind, Brain and Free Will*. Oxford: Oxford University Press.
- Tahko, T. E. (2012). "In defence of aristotelian metaphysics," in *Contemporary Aristotelian Metaphysics*, ed. T. E. Tahko (New York, NY: Oxford University Press), 26–43.
- Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461. doi: 10.1038/nrn.2016.44
- Tononi, G., and Koch, C. (2008). The neural correlates of consciousness: an update. *Ann. N Y Acad. Sci.* 1124, 239–261. doi: 10.1196/annals.1440.004
- Tsuchiya, N., Wilke, M., Frässle, S., and Lamme, V. A. F. (2015). No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn. Sci.* 19, 757–770. doi: 10.1016/j.tics.2015.10.002
- van Gulick, R. (2017). "Consciousness," in *Stanford Encyclopedia of Philosophy (Summer 2017 Edition)*, ed. E. N. Zalta. Available online at: <https://plato.stanford.edu/archives/spr2018/entries/consciousness/>
- Willyard, C. (2015). Rise of the organoids: biologists are building banks of mini-organs and learning a lot about human development on the way. *Nature* 520–522.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Owen and Guta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Characterizing the Dynamical Complexity Underlying Meditation

Anira Escrichs<sup>1,2\*</sup>, Ana Sanjuán<sup>1</sup>, Selen Atasoy<sup>3</sup>, Ane López-González<sup>1</sup>, César Garrido<sup>4</sup>, Estela Càmarà<sup>2,5</sup> and Gustavo Deco<sup>1,6\*</sup>

<sup>1</sup> Computational Neuroscience Group, Department of Information and Communication Technologies, Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain, <sup>2</sup> Cognition and Brain Plasticity Unit, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain, <sup>3</sup> Department of Psychiatry, University of Oxford, Oxford, United Kingdom, <sup>4</sup> Radiology Unit, Hospital Clínic Barcelona, Barcelona, Spain, <sup>5</sup> Department of Cognition, Development and Educational Psychology, University of Barcelona, Barcelona, Spain, <sup>6</sup> Institució Catalana de la Recerca i Estudis Avançats, Barcelona, Spain

## OPEN ACCESS

### Edited by:

Olivia Gosseries,  
University of Liège, Belgium

### Reviewed by:

Andrea Piarulli,  
University of Pisa, Italy  
Silvia Scarpetta,  
University of Salerno, Italy

### \*Correspondence:

Anira Escrichs  
anira.escrichs@upf.edu  
Gustavo Deco  
gustavo.deco@upf.edu

**Received:** 26 January 2019

**Accepted:** 27 June 2019

**Published:** 10 July 2019

### Citation:

Escrichs A, Sanjuán A, Atasoy S, López-González A, Garrido C, Càmarà E and Deco G (2019) Characterizing the Dynamical Complexity Underlying Meditation. *Front. Syst. Neurosci.* 13:27. doi: 10.3389/fnsys.2019.00027

Over the past 2,500 years, contemplative traditions have explored the nature of the mind using meditation. More recently, neuroimaging research on meditation has revealed differences in brain function and structure in meditators. Nevertheless, the underlying neural mechanisms are still unclear. In order to understand how meditation shapes global activity through the brain, we investigated the spatiotemporal dynamics across the whole-brain functional network using the Intrinsic Ignition Framework. Recent neuroimaging studies have demonstrated that different states of consciousness differ in their underlying dynamical complexity, i.e., how the broadness of communication is elicited and distributed through the brain over time and space. In this work, controls and experienced meditators were scanned using functional magnetic resonance imaging (fMRI) during resting-state and meditation (focused attention on breathing). Our results evidenced that the dynamical complexity underlying meditation shows less complexity than during resting-state in the meditator group but not in the control group. Furthermore, we report that during resting-state, the brain activity of experienced meditators showed higher metastability (i.e., a wider dynamical regime over time) than the one observed in the control group. Overall, these results indicate that the meditation state operates in a different dynamical regime compared to the resting-state.

**Keywords:** ignition, whole-brain, meditation, resting-state, fMRI, integration, dynamical complexity

## 1. INTRODUCTION

During the last 2,500 years, contemplative traditions have explored the nature of the mind through self-discipline and self-observation. Meditation per se is not a philosophy or a religious practice, but a method of mental training which enables to cultivate a variety of human abilities, ranging from developing a clearer mind and enhancing attention to cultivating altruistic love and compassion toward other beings (Ricard et al., 2014).

In the last decade, MRI studies exploring the neural correlates of meditation have revealed important insights into how this mental training changes brain function and structure (Brewer et al., 2011; Kilpatrick et al., 2011; Froeliger et al., 2012; Hasenkamp et al., 2012; Taylor et al., 2013; Garrison et al., 2014; Marchand, 2014; Tang et al., 2015; Panda et al., 2016; Kyeong et al., 2017; Mooneyham et al., 2017; Marusak et al., 2018). Yet, little is known about how meditation influences the capability to transmit information across the whole-brain functional network.



Recently, it has been proposed that a brain state can be defined by measuring how the broadness of communication is elicited and distributed through the brain over time, i.e., by characterizing its underlying dynamical complexity (Deco et al., 2017). Investigating the propagation of the neural activity by measuring their dynamical implications (Hutchison et al., 2013) across the whole-brain network may help to explain the fundamental principles of the underlying mechanisms of different brain states (Deco et al., 2011, 2015; Sporns, 2013; Allen et al., 2014). Theoretical methods have been successfully applied to characterize different states of consciousness such as wakefulness, sleep, anesthesia or psychedelic states (Tagliazucchi and Laufs, 2014; Tagliazucchi et al., 2014; Atasoy et al., 2017, 2018; Deco et al., 2017; Jobst et al., 2017).

Here, we investigate the brain's macro-scale mechanisms underlying meditation as well as meditation-induced long-term changes in resting-state using the Intrinsic Ignition Framework (Deco and Kringelbach, 2017; Deco et al., 2017). This data-driven method allows to study the spatiotemporal dynamics across the whole-brain functional network by measuring the effect of naturally occurring local activation events on whole-brain integration.

## 2. METHODS

### 2.1. Participants

A total of forty participants were recruited for this experiment. Half of the participants were experienced meditators (mean (SD) age = 39.8 (10.29); education years = 13.6; mean (SD) hours meditation experience = 9526.9 (8619.8); 7 females) and were recruited from Vipassana communities of Barcelona. All of them had a minimum of 1,000 h of meditation experience and confirmed that they maintained daily practice (>1 hour/day). The other half were well-matched control participants with no prior meditation experience (mean (SD) age = 39.75 (10.13); education years = 13.8; 7 females). Participants reported no history of neurological disorder, provided written informed consent, and were compensated for their participation. The study was approved by the Ethics Committee of the Bellvitge Hospital in accordance with the Helsinki Declaration on ethical research.

### 2.2. Resting-State and Meditation fMRI

A total of 450 brain volumes in each condition were analyzed ( $\approx 15$  min). During rest, participants were asked to look at a fixation cross on the screen, remain as motionless as possible, not to think about anything in particular as well as not to fall asleep. After resting acquisition, all participants were engaged in meditation. Meditators were asked to practice anapanasati meditation (focused attention on breathing). In this type of meditation, subjects try to concentrate all their attention on natural breathing, and when they realize that the mind wanders, they need to recognize it and come back to natural breathing without judgment. Controls were instructed in meditation before being scanned following the instructions as taught by S.N. Goenka (Hart, 1987), who was a Vipassana meditation teacher. Controls confirmed that they understood the procedure after the simulation.

### 2.3. MRI Data Acquisition

MRI images were acquired on a 3T TIM TRIO scanner (Siemens, Erlangen, Germany) using 32-channel receiver coil. The high-resolution T1-weighted images were acquired with 208 slices in the sagittal plane, repetition time (TR) = 1,970 ms, echo time (TE) = 2.34 ms, TI = 1,050 ms, flip angle =  $9^\circ$ , field of view (FOV) = 256 mm, voxel size  $1 \times 1 \times 1$  mm. Resting-state and meditation fMRI were performed by a single shot gradient-echo EPI sequence (TR = 2,000 ms; TE = 29 ms; FOV = 240 mm; in-plane resolution 3 mm; 32 transversal slices with thickness = 4 mm; flip angle =  $80^\circ$ ).

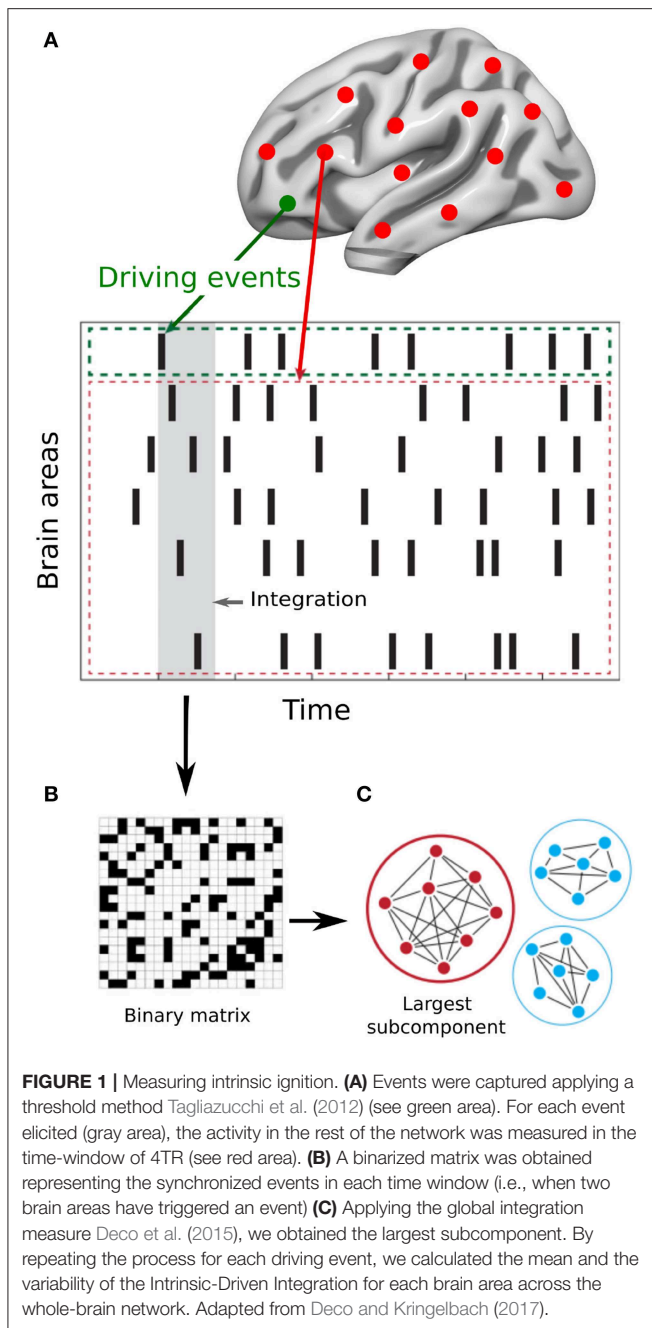
### 2.4. Preprocessing

Preprocessing was computed using the Data Processing Assistant for Resting-State fMRI (DPARSF) (Chao-Gan and Yu-Feng, 2010). Preprocessing included: manually reorienting T1 and EPI images; discarding the first 10 volumes due to magnetic field inhomogeneities; slice-timing correction; realignment for head motion correction; T1 co-registration to functional image; European regularization segmentation; removal of spurious variance through linear regression: six parameters from the head motion correction, the global mean signal, the white matter signal, and the cerebrospinal fluid signal, CompCor; removal of the linear trend in the time-series; spatial normalization to the Montreal Neurological Institute (MNI); spatial smoothing with 6 mm FWHM Gaussian Kernel; and band-pass temporal filtering (0.01–0.25 Hz) (Biswal et al., 1995; Lowe et al., 1998). Finally, we extracted the time-series according to a resting-state atlas of 268 nodes, which ensures the functional homogeneity within each node (Shen et al., 2013).

One meditator was removed due to incidental findings in the MRI session. In addition, 3 controls during meditation and 1 control during rest were excluded due to a head rotation  $>2$  mm or  $2^\circ$ . Moreover, the frame-wise displacement (FD) (Jenkinson et al., 2002) was calculated due to its consideration of voxel-wise differences in motion in its derivation (Yan et al., 2013). Subjects with head motion  $>2$  standard deviations above the group average and movement in more than 25% of time points were excluded from the analysis. FD correction led to the exclusion of 1 control during meditation. Therefore, the final sample of the study included: 19 controls during rest and 16 controls during meditation, 19 meditators during rest and 19 meditators during meditation. After exclusion, no significant differences in terms of age, educational level and gender were observed between groups.

### 2.5. Intrinsic Ignition Framework

The Intrinsic Ignition Framework (Deco and Kringelbach, 2017) measures the degree of elicited whole-brain integration of spontaneously occurring events across time. **Figure 1** describes the algorithm to obtain the intrinsic integration across events of each brain area. First, the time-series are filtered within the narrowband 0.04–0.07 Hz to avoid artifacts (Glerean et al., 2012). Then, for each brain area, driving events are captured for each timepoint and fixed as a binary signal by transforming the filtered time-series into z-scores,  $z_i(t)$ . A threshold  $\theta$  is imposed given by the sum of the mean and the standard deviation of the signal in each brain area, such that the binary sequence  $\sigma(t) = 1$  if  $z_i(t) > \theta$  and is crossing the threshold from below and



$\sigma(t) = 0$  otherwise (Tagliazucchi et al., 2012). If a brain area has triggered an event (Figure 1A green line) then the integration in the rest of the network is measured within the set time window of 4TR (Figure 1A gray time window). A binary matrix is constructed (Figure 1B) representing the synchronized events in each timepoint (i.e., when two brain areas have triggered an event). Afterwards, the global integration measure (Deco et al., 2015) is defined as the largest component in the binarized connectivity matrix, given by the length of the connected component considered as an adjacency matrix (Figure 1C). Finally, the Intrinsic-Driven Mean Integration (IDMI) is defined

as the averaged integration across events, and the variability as the standard deviation of the Intrinsic-Driven Integration. We would like to remark the similitude of our quantitative measure of ignition and the avalanche framework (see, for example, Beggs and Plenz, 2003).

## 2.6. Surrogate Analysis

To ensure that the observed results were not obtained by chance, we applied a surrogate data testing method. Specifically, we randomly permuted the original timeseries across time and measured the ignition in each spontaneous event on the shuffled data. After repeating the process 50 times, we tested whether the empirical ignition values were significantly higher than the surrogates' ignition values.

## 2.7. Statistical Analyses

Here, we compared the IDMI and the variability values for each group (controls and meditators) between conditions (resting and meditation), and we examined if there were differences between groups in the same condition (resting and meditation). Furthermore, we validated our results by comparing the real conditions vs. the randomized ignition data. To do so, we used a Monte-Carlo permutation method. We randomly shuffled the labels between conditions to obtain two new simulated conditions (10,000 permutations). Then, we evaluated how many times the difference between the simulated conditions was higher than the difference between the real conditions. This is, we computed the  $p$ -value of the null hypothesis that the two random distributions show higher difference than the real conditions. Additionally, we applied the Bonferroni correction for multiple comparisons.

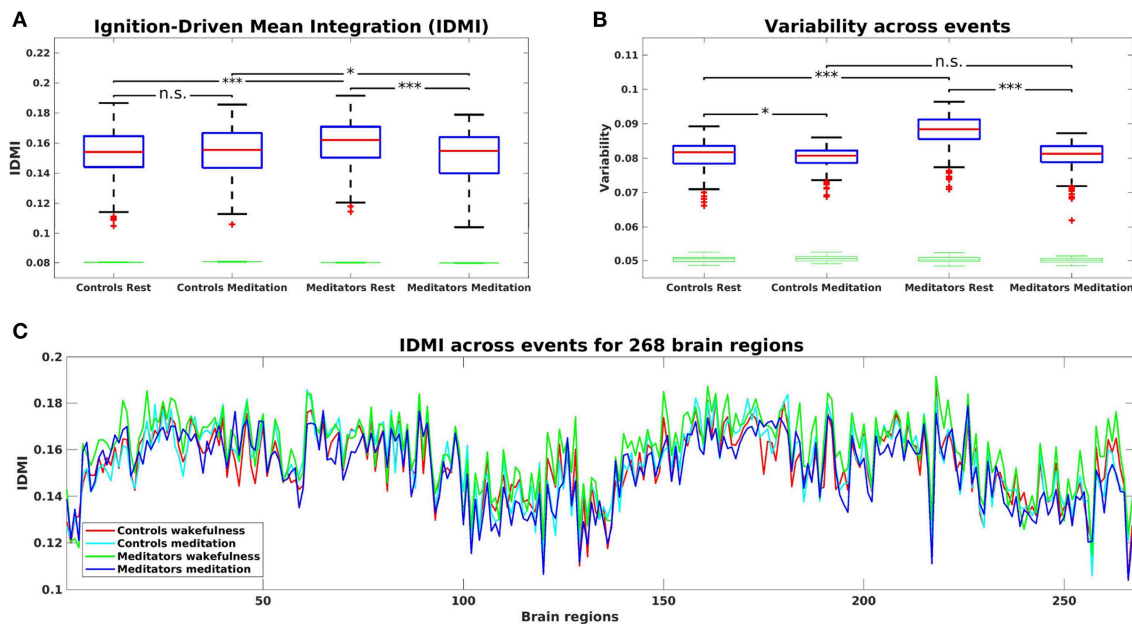
## 3. RESULTS

### 3.1. Intrinsic-Driven Mean Integration (IDMI)

Figure 2A shows the IDMI for each group and brain state, while Figure 2C shows the IDMI for each group and each brain area. The IDMI captures the spatial diversity as differences in average intrinsic ignition profiles across the different nodes. The brain activity of meditators during resting-state showed the highest values of the IDMI compared to the control group ( $p < 0.001$ , Monte-Carlo simulations after Bonferroni correction). Furthermore, this value decreased significantly when meditators were engaged in meditation ( $p < 0.001$ , Monte-Carlo simulations after Bonferroni correction). In contrast, controls did not show any differences between resting-state and meditation conditions.

### 3.2. Variability of Intrinsic-Driven Integration

Next, we calculated the variability of the Intrinsic-Driven Integration in both states (resting-state and meditation) for each group (controls and meditators). Figure 2B shows the variability for each group and brain state. The variability describes the heterogeneity of each brain area, which is closely connected to its local metastability (Deco and Kringelbach, 2017). Thus, it describes how the local activity



**FIGURE 2 | (A)** Mean of the Intrinsic-Driven Integration (IDMI) for each group during resting-state and meditation state. The IDMI was higher in meditators than in controls during resting-state and lower in meditators during meditation. No significant differences were observed in controls between conditions. Furthermore, we show the box plot from the surrogate IDMI data (on the bottom in green). The randomized data were significantly smaller than the original time-series, showing the robust statistical comparisons. **(B)** Both controls and meditators showed higher local metastability across the whole-brain during resting-state compared to meditation. However, the effect was significantly larger for meditators. Furthermore, the metastability in resting-state was significantly higher for experienced meditators than for controls. *P*-values are based on Monte-Carlo simulation after Bonferroni correction, \* $p \leq 0.025$ , \*\*\* $p \leq 0.0005$  and n.s. represents not significant. **(C)** IDMI across events for each group during resting-state and meditation for 268 brain regions.

in each brain area changes across time. High levels of metastability in a node represent a more dynamic function over time, while lower levels represent greater stability. The brain activity of controls and experienced meditators showed higher functional variability (i.e., metastability) in resting-state than in meditation. Nevertheless, the effect was significantly larger for meditators ( $p < 0.001$ , Monte-Carlo simulations after Bonferroni correction) than for controls ( $p = 0.022$ , Monte-Carlo simulations after the Bonferroni correction). Furthermore, the metastability in resting-state was significantly larger for experienced meditators than for controls ( $p < 0.001$ , Monte-Carlo simulations after the Bonferroni correction).

## 4. DISCUSSION

A growing scientific interest lies in the characterization of the meditation state. Hasenkamp and colleagues (Hasenkamp et al., 2012) captured the interactions between four cognitive phases during meditation, but disregarded the dynamical properties that contain relevant spatiotemporal information. Mooneyham and colleagues applied a dynamical functional connectivity approach dissociating mental states during a meditation scan. The authors reported that after a 6 weeks intervention mindfulness program, subjects spent more time in the state of focused attention and less time in the state of mind-wandering (Mooneyham et al.,

2017). In addition, a study that applied graph theoretical analysis (Jao et al., 2016) characterized the degree of the hierarchical organization during meditation. This study revealed that the nodes that had the highest integration degree during rest had the lowest integration degree during meditation, and vice versa. Our work extends these findings by exploring the brain activity during meditation by characterizing the dynamical complexity in terms of how local information is broadcasted across the whole-brain.

Here, we have characterized the dynamical complexity underlying resting-state and meditation in healthy controls and experienced meditators as evidenced by the level of intrinsic ignition. Specifically, in meditators but not in controls, we observed a significant increase of intrinsic ignition during resting-state compared to meditation (Figure 2A). In addition, during resting-state, meditators showed the maximal variability of intrinsic ignition (i.e., metastability) across the whole network, revealing a state of maximum network switching (Figure 2B).

Our results showing an increase of intrinsic ignition during rest compared to meditation are consistent with recent studies on information propagation across the brain. Irmischer and colleagues found a shift from more complex brain dynamics during rest to a state of reduced information propagation during meditation, importantly, only in meditators (Irmischer et al., 2018). Furthermore, Gard and colleagues demonstrated using

graph theory that yoga and meditation practitioners showed greater network integration than controls during rest (Gard et al., 2014). In addition, the increase of metastability in meditators during resting-state is congruent with the increase of the temporal complexity of oscillations during rest in meditators as observed in the previously mentioned study (Irrmischer et al., 2018). Moreover, studies applying a dynamical functional connectivity approach found that individuals with high trait mindfulness transitioned more frequently between brain states at rest (Lim et al., 2018; Marusak et al., 2018).

To sum up, these results demonstrate that experienced meditators can voluntarily alter their whole-brain dynamics when engaged in a meditative state. Furthermore, expertise in meditation leads to increased ignition and metastability at rest. This means that expert meditators are able to regulate the level of exploration of the dynamical repertoire, restricting it during meditation, and enhancing it during rest.

## ETHICS STATEMENT

This study was approved by the Clinical Research Ethics Committee of the Bellvitge University Hospital in accordance with the Declaration of Helsinki. All subjects gave written informed consent to participate in the study.

## REFERENCES

- Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676. doi: 10.1093/cercor/bhs352
- Atasoy, S., Deco, G., Kringelbach, M. L., and Pearson, J. (2018). Harmonic brain modes: a unifying framework for linking space and time in brain dynamics. *Neuroscientist* 24, 277–293. doi: 10.1177/1073858417728032
- Atasoy, S., Roseman, L., Kaelen, M., Kringelbach, M. L., Deco, G., and Carhart-Harris, R. L. (2017). Connectome-harmonic decomposition of human brain activity reveals dynamical repertoire re-organization under LSD. *Sci. Rep.* 7:17661. doi: 10.1038/s41598-017-17546-0
- Beggs, J. M., and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23, 11167–1177. doi: 10.1523/JNEUROSCI.23-35-11167.2003
- Biswal, B., Yetkin, F. Z., Haughton, V. M., Hyde, J. S., Zerrin Yetkin, F., Haughton, V. M., et al. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Res. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409
- Brewer, J. A., Worhunsky, P. D., Gray, J. R., Tang, Y.-Y., Weber, J., and Kober, H. (2011). Meditation experience is associated with differences in default mode network activity and connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20254–20259. doi: 10.1073/pnas.1112029108
- Chao-Gan, Y., and Yu-Feng, Z. (2010). DPARSF: a MATLAB toolbox for "Pipeline" data analysis of resting-state fMRI. *Front. Syst. Neurosci.* 4:13. doi: 10.3389/fnsys.2010.00013
- Deco, G., Jirsa, V. K., and McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nat. Rev. Neurosci.* 12, 43–56. doi: 10.1038/nrn2961
- Deco, G., and Kringelbach, M. L. (2017). Hierarchy of information processing in the brain: a novel 'Intrinsic Ignition' framework. *Neuron* 94, 961–968. doi: 10.1016/j.neuron.2017.03.028
- Deco, G., Tagliazucchi, E., Laufs, H., Sanjuán, A., and Kringelbach, M. L. (2017). Novel intrinsic ignition method measuring local- global integration characterizes wakefulness and deep sleep. *eNeuro* 4, 106–17. doi: 10.1523/ENEURO.0106-17.2017

## AUTHOR CONTRIBUTIONS

AE and GD designed the study. EC and CG designed the MRI protocol. AE collected the data and wrote the first version of the manuscript. AE and AS pre-processed the fMRI data. AE, AS, and AL-G performed the analyses. AE, AS, GD, and SA interpreted the results. All authors contributed to manuscript revision, read and approved the submitted version.

## FUNDING

AE is supported by a Francisco J. Varela Award from the Mind and Life Europe. AS is supported by the Spanish Ministry of Economy and Competitiveness Grant FPD12013-17045. GD is supported by the Spanish Ministry Research Project PSI2016-75688-P (AEI/FEDER), by the European Union's Horizon 2020 FET Flagship Human Brain Project 785907 HBP SGA2, by the Catalan Research Group Support 2017 SGR 1545 and by the Foundation Marato TV3 2016.

## ACKNOWLEDGMENTS

We are particularly grateful to the Mind and Life Europe for supporting this project. We thank the participants for their participation in the study.

- Deco, G., Tononi, G., Boly, M., and Kringelbach, M. L. (2015). Rethinking segregation and integration: contributions of whole-brain modelling. *Nat. Rev. Neurosci.* 16, 430–439. doi: 10.1038/nrn3963
- Froeliger, B., Garland, E. L., Kozink, R. V., Modlin, L. A., Chen, N. K., McClernon, F. J., et al. (2012). Meditation-state functional connectivity (msFC): Strengthening of the dorsal attention network and beyond. *Evid. Based Complement. Alternat. Med.* 2012:680407. doi: 10.1155/2012/680407
- Gard, T., Taquet, M., Dixit, R., Hölzel, B. K., de Montjoye, Y.-A., Brach, N., et al. (2014). Fluid intelligence and brain functional organization in aging yoga and meditation practitioners. *Front. Aging Neurosci.* 6:76. doi: 10.3389/fnagi.2014.00076
- Garrison, K. A., Scheinost, D., Constable, R. T., and Brewer, J. A. (2014). BOLD signal and functional connectivity associated with loving kindness meditation. *Brain Behav.* 4, 337–347. doi: 10.1002/brb3.219
- Glerean, E., Salmi, J., Lahnakoski, J. M., Jääskeläinen, I. P., and Sams, M. (2012). Functional magnetic resonance imaging phase synchronization as a measure of dynamic functional connectivity. *Brain Connect.* 2, 91–101. doi: 10.1089/brain.2011.0068
- Hart, W. (1987). *The Art of Living: Vipassana Meditation as Taught by S. N. Goenka*. Onalaska, WA: Pariyatti.
- Hasenkamp, W., Wilson-Mendenhall, C. D., Duncan, E., and Barsalou, L. W. (2012). Mind wandering and attention during focused meditation: a fine-grained temporal analysis of fluctuating cognitive states. *NeuroImage* 59, 750–760. doi: 10.1016/j.neuroimage.2011.07.008
- Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., et al. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage* 80, 360–378. doi: 10.1016/j.neuroimage.2013.05.079
- Irrmischer, M., Houtman, S. J., Mansvelder, H. D., Tremmel, M., Ott, U., and Linkenkaer-Hansen, K. (2018). Controlling the Temporal Structure of Brain Oscillations by Focused Attention Meditation. *Hum. Brain Mapp.* 39, 1825–1838. doi: 10.1002/hbm.23971
- Jao, T., Li, C.-W., Vértes, P. E., Wu, C. W., Achard, S., Hsieh, C.-H., et al. (2016). Large-scale functional brain network reorganization during taoist meditation. *Brain Connect.* 6, 9–24. doi: 10.1089/brain.2014.0318



- Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17, 825–41. doi: 10.1006/nimg.2002.1132
- Jobst, B. M., Hindriks, R., Laufs, H., Tagliazucchi, E., Hahn, G., Ponce-Alvarez, A., et al. (2017). Increased stability and breakdown of brain effective connectivity during slow-wave sleep: mechanistic insights from whole-brain computational modelling. *Sci. Rep.* 7:4634. doi: 10.1038/s41598-017-04522-x
- Kilpatrick, L. A., Suyenobu, B. Y., Smith, S. R., Bueller, J. A., Goodman, T., Creswell, J. D., et al. (2011). Impact of mindfulness-based stress reduction training on intrinsic brain connectivity. *NeuroImage* 56, 290–298. doi: 10.1016/j.neuroimage.2011.02.034
- Kyeong, S., Kim, J., Kim, D. J., Kim, H. E., and Kim, J.-J. (2017). Effects of gratitude meditation on neural network functional connectivity and brain-heart coupling. *Sci. Rep.* 7:5058. doi: 10.1038/s41598-017-05520-9
- Lim, J., Teng, J., Patanaik, A., Tandi, J., and Massar, S. A. (2018). Dynamic functional connectivity markers of objective trait mindfulness. *NeuroImage* 176, 193–202. doi: 10.1016/j.neuroimage.2018.04.056
- Lowe, M., Mock, B., and Sorenson, J. (1998). Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations. *NeuroImage* 7, 119–132.
- Marchand, W. R. (2014). Neural mechanisms of mindfulness and meditation: Evidence from neuroimaging studies. *World J. Radiol.* 6, 471–479. doi: 10.4329/wjr.v6.i7.471
- Marusak, H. A., Elrahal, F., Peters, C. A., Kundu, P., Lombardo, M. V., Calhoun, V. D., et al. (2018). Mindfulness and dynamic functional neural connectivity in children and adolescents. *Behav. Brain Res.* 336, 211–218. doi: 10.1016/j.bbr.2017.09.010
- Mooneyham, B. W., Mrazek, M. D., Mrazek, A. J., Mrazek, K. L., Phillips, D. T., and Schooler, J. W. (2017). States of mind: characterizing the neural bases of focus and mind-wandering through dynamic functional connectivity. *J. Cogn. Neurosci.* 29, 495–506. doi: 10.1162/jocn-a-01066
- Panda, R., Bharath, R. D., Upadhyay, N., Mangalore, S., Chennu, S., and Rao, S. L. (2016). Temporal dynamics of the default mode network characterize meditation-induced alterations in consciousness. *Front. Hum. Neurosci.* 10:372. doi: 10.3389/fnhum.2016.00372
- Ricard, M., Lutz, A., and Davidson, R. J. (2014). Mind of the meditator. *Sci. Am.* 311, 38–45. doi: 10.1038/scientificamerican1114-38
- Shen, X., Tokoglu, F., Papademetris, X., and Constable, R. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage* 82, 403–415. doi: 10.1016/j.neuroimage.2013.05.081
- Sporns, O. (2013). Network attributes for segregation and integration in the human brain. *Curr. Opin. Neurobiol.* 23, 162–171. doi: 10.1016/j.conb.2012.11.015
- Tagliazucchi, E., Balenzuela, P., Fraiman, D., and Chialvo, D. R. (2012). Criticality in large-scale brain fMRI dynamics unveiled by a novel point process analysis. *Front. Physiol.* 3:15. doi: 10.3389/fphys.2012.00015
- Tagliazucchi, E., Carhart-Harris, R., Leech, R., Nutt, D., and Chialvo, D. R. (2014). Enhanced repertoire of brain dynamical states during the psychedelic experience. *Hum. Brain Mapp.* 35, 5442–5456. doi: 10.1002/hbm.22562
- Tagliazucchi, E., and Laufs, H. (2014). Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* 82, 695–708. doi: 10.1016/j.neuron.2014.03.020
- Tang, Y.-Y., Hölzel, B. K., and Posner, M. I. (2015). The neuroscience of mindfulness meditation. *Nat. Rev. Neurosci.* 16, 213–225. doi: 10.1038/nrn3916
- Taylor, V. A., Daneault, V., Grant, J., Scavone, G., Breton, E., Roffe-vidal, S., et al. (2013). Impact of meditation training on the default mode network during a restful state. *Soc. Cogn. Affect. Neurosci.* 8, 4–14. doi: 10.1093/scan/nsr087
- Yan, C.-G., Craddock, R. C., Zuo, X.-N., Zang, Y.-F., and Milham, M. P. (2013). Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. *NeuroImage* 80, 246–62. doi: 10.1016/j.neuroimage.2013.04.081

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Escrichs, Sanjuán, Atasoy, López-González, Garrido, Càmarà and Deco. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Indicators and Criteria of Consciousness in Animals and Intelligent Machines: An Inside-Out Approach

Cyriel M. A. Pennartz<sup>1,2\*</sup>, Michele Farisco<sup>3,4</sup> and Kathinka Evers<sup>3</sup>

<sup>1</sup>Department of Cognitive and Systems Neuroscience, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands, <sup>2</sup>Research Priority Area, Brain and Cognition, University of Amsterdam, Amsterdam, Netherlands, <sup>3</sup>Centre for Research Ethics and Bioethics, Uppsala University, Uppsala, Sweden, <sup>4</sup>Biogen, Biology and Molecular Genetics Institute, Ariano Irpino, Italy

## OPEN ACCESS

### Edited by:

Wolf Joachim Singer,  
Ernst Strüngmann Institut für  
Neurowissenschaften, Germany

### Reviewed by:

Karl Friston,  
University College London,  
United Kingdom  
José M. Delgado-García,  
Universidad Pablo de Olavide, Spain

### \*Correspondence:

Cyriel M. A. Pennartz  
c.m.a.pennartz@uva.nl

**Received:** 13 March 2019

**Accepted:** 24 June 2019

**Published:** 16 July 2019

### Citation:

Pennartz CMA, Farisco M and  
Evers K (2019) Indicators and Criteria  
of Consciousness in Animals  
and Intelligent Machines: An  
Inside-Out Approach.  
*Front. Syst. Neurosci.* 13:25.  
doi: 10.3389/fnsys.2019.00025

In today's society, it becomes increasingly important to assess which non-human and non-verbal beings possess consciousness. This review article aims to delineate criteria for consciousness especially in animals, while also taking into account intelligent artifacts. First, we circumscribe what we mean with "consciousness" and describe key features of subjective experience: qualitative richness, situatedness, intentionality and interpretation, integration and the combination of dynamic and stabilizing properties. We argue that consciousness has a biological function, which is to present the subject with a multimodal, situational survey of the surrounding world and body, subserving complex decision-making and goal-directed behavior. This survey reflects the brain's capacity for internal modeling of external events underlying changes in sensory state. Next, we follow an inside-out approach: how can the features of conscious experience, correlating to mechanisms inside the brain, be logically coupled to externally observable ("outside") properties? Instead of proposing criteria that would each define a "hard" threshold for consciousness, we outline six indicators: (i) goal-directed behavior and model-based learning; (ii) anatomic and physiological substrates for generating integrative multimodal representations; (iii) psychometrics and meta-cognition; (iv) episodic memory; (v) susceptibility to illusions and multistable perception; and (vi) specific visuospatial behaviors. Rather than emphasizing a particular indicator as being decisive, we propose that the consistency amongst these indicators can serve to assess consciousness in particular species. The integration of scores on the various indicators yields an overall, graded criterion for consciousness, somewhat comparable to the Glasgow Coma Scale for unresponsive patients. When considering theoretically derived measures of consciousness, it is argued that their validity should not be assessed on the basis of a single quantifiable measure, but requires cross-examination across multiple pieces of evidence, including the indicators proposed here. Current intelligent machines, including deep learning neural networks (DLNNs) and agile robots, are not indicated to be conscious yet. Instead of assessing machine consciousness by a brief Turing-type of test, evidence for it may gradually accumulate when we study machines ethologically

and across time, considering multiple behaviors that require flexibility, improvisation, spontaneous problem-solving and the situational conspectus typically associated with conscious experience.

**Keywords:** awareness, bird, episodic memory, goal-directed behavior, illusion, robot, rodent, visuospatial behavior

## INTRODUCTION

Over the past decades, it has become increasingly important to assess consciousness in non-human and/or non-verbal beings, such as animals and—perhaps in a not too distant future—machines accomplishing sophisticated cognitive tasks. Examples that underscore the relevance of such assessments include the question of animal sentience (including a capacity to suffer pain) in relation to animal welfare in bio-industrial farming, in procedures for ritual slaughter, animal experimentation for biomedical purposes, but also in domestic pet keeping. Second, in the spectrum of Artificial Intelligence (AI) techniques, we find computers and robots performing complex tasks, such as playing Jeopardy! and other knowledge games, Alpha-Go or navigating through indoor as well as outdoor environments while negotiating irregular terrains (Murphy et al., 2011; Ferrucci et al., 2013; Hassabis et al., 2017). AI is currently developing at such a rapid pace that supra-human performance can be claimed for several games, acquired through self-play (Silver et al., 2018). This development gives rise to the question whether machines may have a level of sentience or consciousness, and ipso facto may be entitled to certain rights and moral status. A third domain where assessment of consciousness based on non-verbal criteria is highly relevant is represented by people who cannot linguistically express themselves, ranging from infants to patients suffering from global aphasia, deeply demented patients, patients in a minimally conscious state or locked-in syndrome.

In this review article, we will suggest operational criteria that can facilitate researchers and policymakers in a range of connected fields of study (neuroscience, AI, psychology and behavioral sciences, medicine, philosophy, ethics and animal welfare) in attributing levels of consciousness to non-verbal beings which may assume varying states of information processing (such as sleep and wakefulness). Such criteria aid us in determining who or what may be conscious, but we must note that an entity might be conscious even if it fails to satisfy them. Like in the neuroimaging assessment of residual consciousness in patients with disorders of consciousness (Schnakers et al., 2009; Owen, 2014), we should keep in mind here that absence of evidence is not the same as evidence of absence of consciousness. At the same time, one must be cautious not to attribute consciousness to any arbitrary object in nature without further argumentation. As we will argue, many criteria can be best circumscribed as “indicators of consciousness,” a term which avoids the connotation of “criterion” with a hard threshold in judging or deciding whether someone or something is conscious or not.

With “operational” we mean that such indicators of consciousness should be sufficiently concrete to apply them

in the practice of observing the behavior or cognitive performance of the subject under scrutiny. We will use the term “consciousness” interchangeably with “awareness,” without denying that a distinction between them has been drawn in some previously proposed frameworks (Farisco et al., 2017; Kastrup, 2017). We will depart from a brief overview of features that may be reasonably attributed to conscious experience, contextualize these properties in terms of biological function, and then proceed to translate such properties into observable behaviors, physiological responses, anatomical traits and/or internally generated phenomena (such as computational operations reaching a certain degree of complexity) that may be considered indicators of consciousness. That this approach is necessarily indirect and does not yield definitive proof, but rather indications for consciousness, is an unavoidable consequence of the epistemological position that the conscious experience of other beings, including other humans, is essentially subjective and that relevant knowledge about another subject’s experience is always inferential (Evers and Sigman, 2013).

Defining such indicators of consciousness is not an easy task and we refer to previous work in this direction (Weiskrantz, 1995; Griffin and Speck, 2004; Seth et al., 2005; Butler and Cotterill, 2006; Edelman and Seth, 2009; Pennartz, 2015). First and foremost, each definition of a criterion or indicator of consciousness depends on assumptions being made on what a conscious experience is, if and how it normally arises from a brain-body complex interacting with its environment, what its biological function may be, and how it relates to basic computational operations performed by the brain or artificial machinery aiming to replace it, either fully or in part. In this respect, the current article differs from previous outlines of consciousness criteria that were based, in part, on Integration Information Theory (IIT; Boly et al., 2013; Tononi et al., 2016), Global Neuronal Workspace (Dehaene and Changeux, 2011; Baars et al., 2013), or combinations thereof (Seth et al., 2005). As explained in more detail below, we start from the view that consciousness has a biological function: it presents the subject with a multimodal, situational survey of its surrounding world (including its own body), which subserves the kind of complex decision-making that is associated with goal-directed, deliberate, planned behavior (Pennartz, 2015, 2018). This survey takes the form of a dynamic “world model” or representation that is coded or given a symbolic expression by groups of interconnected thalamo-cortical areas, which are essentially making a “best guess” of the external causes underlying changes in sensory inputs reaching the brain (see Gregory, 1980; Mumford, 1992; Friston, 2010; Olcese et al., 2018b). In line with this, a previously proposed framework, departing from philosophy (i.e., the Intrinsic Consciousness Theory; Farisco et al., 2017),

similarly advocates a view of consciousness as a modeling and goal-directed activity, reflecting the brain's ability to model the world and body, allowing the subject to infer the surrounding world's state on the basis of such models. In Intrinsic Consciousness Theory, consciousness is an abductive, probabilistic and dynamic feature of the brain that is shaped by interaction with the world and largely depends on previous experience. This abductive feature is realized through the brain's ability to infer the best explanation (or model) of the data received from its surroundings *via* sensory receptors. Although originally developed in the context of studies on the brain, Intrinsic Consciousness Theory does not exclude in principle that the functions we associate with consciousness may be manifested by artificial subjects.

While it has not been uncommon in the past to associate consciousness with volition and voluntary decision-making (Crick and Koch, 2003), these latter two concepts have been difficult to test or quantify in practice, whereas goal-directed behavior (see below) is a well-defined concept that has been operationalized in studies on animal behavior (Balleine and Dickinson, 1998; Dickinson, 2012). However, it has remained unclear whether goal-directed behavior may deliver equally clear indicators of consciousness for assessing conscious processing in intelligent machines, and to what extent the coupling of goal-directed behavior to consciousness is unconditionally valid or conditional on further requirements. As far as evidence for goal-directed behavior may be lacking in certain animal species, the question arises whether other indicators of consciousness may be derived from the concept of conscious experience as a multimodal, situational survey.

In short, we will consecutively address five questions: (i) What do we mean with “consciousness,” and which set of features can we uniquely associate with subjective experience? (ii) How many distinctions between goal-directed behavior and habitual actions help delineate indicators of consciousness in non-human and/or non-verbal subjects? (iii) Which other indicators of consciousness may be used in addition to those based on goal-directedness? (iv) To what extent can theoretically derived, quantitative measures of brain activity be used as indicators of consciousness? and (v) Can the indicators of consciousness held to be applicable to non-human animals be extrapolated to apply to intelligent machines as well—and if not, which operational approach may be proposed in addition or instead? In addressing these questions, we will follow an “inside-out” approach: starting from features that can be attributed to subjective experience, correlating to neural mechanisms inside the brain, we continue with outlining a biological function of consciousness in relation to behavior, and hence infer externally observable (“outside”) properties that can be reasonably associated with consciousness in terms of brain anatomy, physiology and behavior (**Figure 1**). Thus, the approach is to study how subjects “act out” their multimodal situational survey through externally observable phenomena. This approach is pluralistic in the sense that it is open to recognizing the validity of multiple sources of potential evidence for consciousness (e.g., physiological, behavioral, anatomical).

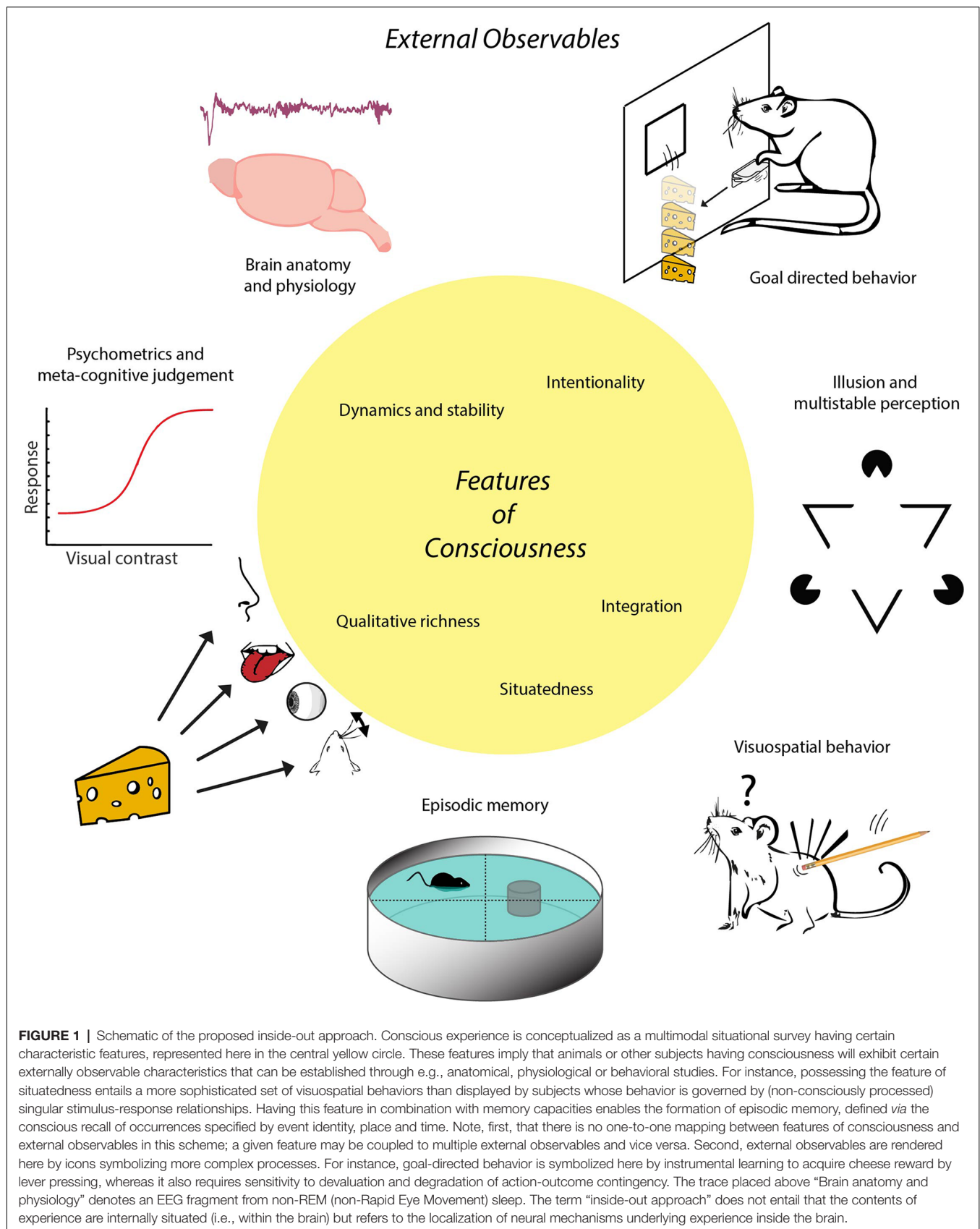
## WHAT DOES “CONSCIOUSNESS” MEAN AND WHAT ARE ITS KEY PROPERTIES?

In clinical practice, states of consciousness are mainly probed by prompting patients to report events with accuracy, usually by verbal expression (Laureys et al., 2004; Seth et al., 2005; Monti et al., 2010; Bruno et al., 2011). For instance, a physician may ask a patient whether he or she feels a touch to the skin, hears a particular tone or sees a stimulus presented on a screen. Such perceptual testing is routinely supplemented with verbal commands to generate a voluntary response; the physician may ask the patient: “can you raise your left arm” or “can you tell me your name” (Monti et al., 2009, 2010). Despite their practical usefulness, these traditional bedside criteria suffer from the limitation that they depend on the motor capacities of the individual, whereas patients suffering from disorders of consciousness and lacking capacities for behavioral expression may have residual consciousness that these traditional criteria perform fail to detect (Schnakers et al., 2009; Farisco et al., 2015; Farisco and Evers, 2016). Other patient groups e.g., suffering from paralysis, global aphasia or locked-in syndrome also lack capacities for behavioral communication, yet are recognized to remain conscious (Laureys et al., 2004; Casali et al., 2013; Owen, 2014).

A special case is constituted by patients who are fully locked-in and thus lack any communication channel with their families or medical personnel. Sometimes evidence on their conscious state emerges when they recover and recall from memory what they experienced during their locked-in state (Schiff, 2010; Sitt et al., 2014; Engemann et al., 2018), but in many other cases such *post hoc* evidence for conscious processing cannot be obtained. Evidence for active cognitive processing in behaviorally non-communicative patients may be obtained by functional magnetic resonance imaging (fMRI) imaging in combination with verbal commands to generate differentiated imagery (e.g., spatial vs. motor imagery), but such evidence is not fully conclusive as the findings are compatible with the possibility that ongoing cognitive processing might occur in the absence of awareness (Owen, 2013).

These and other considerations (Pennartz, 2015, 2018) emphasize the need to maintain a dissociation between consciousness *per se* and the repertoire of motor (including verbal) actions humans and non-human subjects may use to express what they experience. Thus, we define consciousness in the first instance by its experiential, sensory nature (Jackendoff, 1987). We take “experience” to denote the complex of conscious sensations occurring during states of wakeful processing of sensory inputs of external origin (perception), imagery (sensations due to internally driven processing under cognitive control) or dreaming (sensations due to internally driven processing, prevalent during REM sleep and less subject to cognitive control; Hobson, 2009). This points to a limitation in the scope of the current article; we do not theoretically deny the possibility of adopting a more inclusive view on subjective experience (see Farisco et al., 2017; Kastrup, 2017). Notably, many neuroscientists use the term “experience” more





**TABLE 1** | Key features of consciousness.

Qualitative richness	Conscious experience is specified by a wide, varied palette of sensory modalities and submodalities
Situatedness	Conscious experience is set in a specific spatiotemporal situation in which the subject is immersed
Intentionality	Conscious experience is about something other than is entailed by the underlying neuronal substrates; it depends on the subjective interpretation of external or internal inputs to the brain
Integration	The elements of a scene or situation we perceive are experienced as a unified whole
Dynamics and stability	Brain systems for conscious processing allow for both dynamic changes in experience as well as short-term stabilization of percepts

broadly, referring to life events that happen to organisms in general and lead to functional changes such as learning or memory storage (captured by terms such as “experience-dependent plasticity”; Pennartz, 2018). Perception, imagery and dreaming can be considered to constitute the main forms or modes by which consciousness is subjectively manifested, contrasting to unaware states such as dreamless sleep, anesthesia or coma.

Defining consciousness directly is notoriously difficult because one is forced to resort to related concepts such as wakefulness, experience, feeling and perceiving (or their unaware counterparts, such as deep sleep). When attempting to define these terms in turn, one is forced to resort to circularity. Therefore, we (Pennartz, 2015) and others have previously characterized (healthy, full-blown) conscious experience not by trying to define it directly, but as having a number of key features that characterize consciousness (though not necessarily uniquely; Hassin, 2013; **Table 1**):

- (i) **Qualitative richness:** conscious contents we experience are specified by a wide, varied palette of sensory modalities (vision, audition, somatosensation, olfaction, taste, vestibular sense) and submodalities (e.g., for vision: texture, motion, color, size, shape, depth). This notion does not imply that any experience is required to involve all or most modalities at the same time, but does imply that contents are constituted by modally specific elements that are experienced as distinct from one another (see Pennartz, 2009).
- (ii) **Situatedness:** whatever a subject in a healthy, normal condition experiences is set in a specific spatiotemporal situation, i.e., the subject is immersed in a multimodal situation characterized by a specific body position occupying a place in an environmental space and within a temporal framework. With “immersion” we mean that the subject is not looking at its own situation from a distance, but experiences its own body as being within the situation.
- (iii) **Intentionality:** experiences are fundamentally about something other than is entailed by the neuronal substrates (“vehicles”) underlying their generation. We can be conscious of exploding fireworks while the brain mechanisms at work to generate this experience are not exploding themselves (Searle, 2004; Olcese et al., 2018b). Our experiences fundamentally depend on an interpretation of the external or internally generated inputs to the brain which are processed by neural mechanisms that are of a different nature than the subjectively experienced contents. This process can alternatively be described as making subjective inferences on the causes of sensory inputs

reaching our brain, e.g., on objects emitting light, sound waves or other changes in energy impinging on sensory receptors. This is not to say that the brain would be insensitive to the causes and their ensuing sensory inputs. On the contrary, the latter affect the brain *via* the cranial nerves and spinal cord.

- (iv) **Integration:** the elements of a scene or situation we perceive are experienced as a unified whole, and this “in-one-piece” property is not compatible with subjects sustaining different aware experiences at the same time (see Tononi, 2004). We consider integration to be a broad, overarching term that comprises different forms and computational mechanisms, such as binocular fusion (i.e., the merging of different visual information originating from images projected on the left and right eye), temporal integration of visual information across saccades and eye blinks, integration of contours and other Gestalt features into whole objects, and multimodal integration (e.g., perceived audiovisual simultaneity in the presence of physical light-sound delays).
- (v) **Dynamics and stability:** brain systems for conscious processing allow for both dynamic changes in experience as well as for short-term stabilization of percepts, e.g., when viewing ambiguous pictures (e.g., Necker cube inversion; Borsellino et al., 1982), binocular rivalry (Tong et al., 1998), experiencing illusions (e.g., a Kanizsa triangle; Seghier and Vuilleumier, 2006; von der Heydt, 2013) or change detection (Huettel et al., 2001). Moreover, when moving one’s eyes and navigating through a stable environment, the subject experiences objects and scene elements as being stably positioned, indicating that the involved brain systems generate stable percepts despite a plethora of dynamic changes in sensory inputs.

Features (ii) and (iv) are often associated with having a first-person, egocentric perspective (Searle, 1992, 2015) or an “observing self” (Baars et al., 2003). Indeed a subject samples world and body states through one overall sensory apparatus, resulting in a visual perspective that is integrated with body position in space (Pennartz, 2015, 2018), but we add that this notion of egocentric perspective leans rather heavily on the dominance of vision in conscious experience, and is less obvious when subjects consciously manipulate objects or navigate allocentrically through their environments (i.e., experience can be partially object-centered and influenced by allocentric knowledge; Pennartz, 2015).

According to particular schools of thought, further features of consciousness may be added, such as aspects of language (Carruthers, 1989), self-consciousness and “higher-order

thought” (HOT; see Seth et al., 2005). HOT theories propose that consciousness critically depends on a form of meta-cognition by which subjects reflect on, evaluate and judge sensory contents (Cleeremans et al., 2007; Lau and Rosenthal, 2011). However, we and others (Koch et al., 2016) have argued elsewhere that human subjects retain the primary, sensory consciousness that we referred to above when such meta-cognitive or reflexive capacities are compromised or lost, for instance because of massive frontal lesions, drug-induced alterations in, or loss of, confidence judgments, or mental illness (Pennartz, 2015; Boly et al., 2017; Storm et al., 2017). Moreover, conscious sensations can exist without being granted a particular ontological status (e.g., as “veridical” or “hallucinatory”) by a separate meta-cognitive system.

Thus, we argue to keep the list of key properties of conscious experience compact, summarizing it as having a multimodal, unified and situational survey of the surrounding world and oneself. This survey is fundamentally inferential and thus subjective, given that more than one interpretation is possible when multiple options for making inferences are available. The multimodal, situational survey offers a “best guess” representation (see Gregory, 1980; Marcel, 1983) of what is happening around us, correcting for rapid sensorimotor changes that would disrupt a stable world view (e.g., eye blinks, saccades) and yet allowing for a dynamic updating. This characterization, however, is very much based on human subjective experience and its occurrence in animals or machines—in this form or in related manifestations—can of course not be directly verified. The core features do not reveal indicators of consciousness by which states of consciousness in non-verbal beings can be directly identified. Nonetheless, we will argue that they do point the way to finding reasonable, although indirect, indicators of consciousness that can be used in practice.

## BIOLOGICAL FUNCTION OF CONSCIOUSNESS

If one accepts the premise that conscious experience essentially corresponds to having a multimodal, inferential and situational world survey, what function could it subserve? Whereas it is not self-evident that consciousness *does* have a biological function, it is of note that evolution has led to the development of brain systems which are strongly associated with conscious sensing (i.e., thalamo-cortical systems for all modalities except, arguably, olfaction which may not require thalamic functioning; Pennartz, 2015) and pose high demands in terms of energy consumption. This evolutionary development underwrites that consciousness has survival and reproductive value, contributing to a subject’s innate inclination to persist and enhance in its integral functioning (Evers, 2009). To this one may raise the reductionist objection that consciousness could be an epiphenomenon, and that the “real work” for realizing meaningful sensorimotor behaviors is being performed by the neurons, without having to resort to an elusive phenomenon such as consciousness (Churchland, 1984).

However, this counterargument does not hold because of the *functional nature* of conscious experience as characterized above. When the neurons would, hypothetically, execute the “real work” without generating a multimodal, situational survey, consciousness would be lacking and organisms would have to grope around “in the dark” in their daily struggle for survival and reproduction (notably, darkness is only a metaphor in this case because a wholly nonconscious subject would not even be aware of darkness). At first glance, a proponent of the epiphenomenal argument may sit comfortably with this notion because, after all, subjects may seem to survive perfectly well as long as sensory stimuli are properly handled by sensorimotor neural circuits producing effective motor actions. However, the problem with this reasoning is that it only works well for behaviors generated in response to simple stimulus-response situations, such as reflexes and habits. Reflexes ensure simple and very fast motor reactions in response to low-dimensional sensory inputs, such as the heat of a candle flame resulting in immediate withdrawal of a nearby hand. Habits result from overtraining of stimulus-response associations (Balleine and Dickinson, 1998; Dickinson, 2012; de Wit et al., 2018) and thereby come to be executed more automatically and without (or at least with less) consciousness. We may be aware of the global sensory consequences of habitual actions but remain largely unaware of the sensory details or their underlying fine movements (Lisman and Sternberg, 2013; Pennartz, 2018).

Contrasting with reflexes and habits, complex decision-making cannot purely rely on low-dimensional or single-stimulus information to select actions quickly. With “complex” we mean that many variables need to be simultaneously taken into account to reach an optimal decision: not only estimates of single stimuli and their motivational value (in terms of reward or punishment), but also estimates of energetic effort, knowledge about the sensory-specific, social and reproductive consequences of stimuli, and the arrangement of, and coherence between, these variables in space and time. Even objects without explicit motivational value need to be taken into account, for instance, because they may become obstacles if the agent chooses a route to a more distant goal, or because they become instrumental if a different goal is selected in the course of action. Complex decision-making in novel or familiar situations not only requires organisms to have an internal evaluative system, specifying what they need to satisfy homeostatic variables both in terms of reward (acquisition of desirable items such as food and water) and punishment (avoidance or harmful and noxious stimuli), but also to have a model that is optimally informative about their current and future sensory world. Below we will go more deeply into two concepts strongly related to complex decision-making: goal-directed behavior and model-based learning.

It would be extremely difficult to negotiate such complex problem sets if subjects would have to work without some sort of situational survey offering a quick grasp on the situation at hand. We suggest that evolution has solved this problem by developing specialized brain systems actively generating rapidly updatable world surveys. We do not mean to equate consciousness with complex decision-making (or related meta-cognitive processes

**TABLE 2 |** Indicators of Consciousness.

Goal Directed Behavior and model-based learning	Goal-directed behavior is driven by a representation of the expected consequences of action and depends on knowledge of actions being causal for obtaining a desirable outcome Model-based learning is defined by the subject building an explicit, internal model of its state space, including specific stimulus–outcome relationships and enabling prospective cognition
Brain anatomy and physiology	In primates, conscious experience is associated with thalamocortical systems. In other vertebrates, similar brain structures indicate the presence of consciousness. In vertebrates, functional analogs of cortex and thalamus may support consciousness
Psychometrics and meta-cognitive judgment	Psychometric properties of stimulus detection and discrimination, coupled to meta-cognitive judgments on perceived stimuli, indicate perceptual similarities between humans and some animal species (e.g., monkeys, rodents, and birds)
Episodic memory	Autobiographical memory, i.e., memory of events ("what") a subject experienced at a particular place ("where") and time ("when"), indicates the presence of consciousness because it is linked to it by the definition of declarative memory
Illusion and multistable perception	Susceptibility to illusions and perceptual ambiguity has been demonstrated in non-human primates and several other species, and is coupled to intentionality, the key feature by which conscious systems can interpret sensory information in different ways
Visuospatial behavior	Having a stable, situational survey in the face of ongoing body and eye movements is coupled to specific visuospatial abilities, such as reaching into non-foveated parts of space and identifying objects as being stably positioned

such as confidence judgments) but claim that consciousness *suberves* (i.e., promotes and supports) it (see Griffin and Speck, 2004). In other words, consciousness is relevant for complex decision-making (which is not to say the former is a necessary condition for the latter). Conscious experience also continues in the absence of overt decisions or motor actions, such as during dreaming, daydreaming or passive observation without motor intentions or planning (Treserras et al., 2009; Pennartz, 2018).

## INDICATORS OF CONSCIOUSNESS IN ANIMALS

Departing from the premise that consciousness is functionally linked to complex decision-making, we can now begin to outline indicators of consciousness in animals in the context of their behavior, neuroanatomy and physiology (**Figure 1**), noting that for several classes of animals, multiple indicators of consciousness will be proposed as contributing evidence in favor or against the presence of consciousness (**Table 2**).

### Goal-Directed Behavior and Model-Based Learning

The distinction between habits and goal-directed behavior has been operationalized by Balleine and Dickinson (1998), Dickinson (2012) and is important for the current debate, even though goal-directed behavior should not be equated with complex decision-making in general. Agents display goal-directed behavior when this behavior is driven by a representation of the expected consequences of action and depends on knowledge of actions being causal for obtaining a desirable outcome (Dickinson, 2012). Operationally, this definition can be tested, first, by studying the consequences of devaluation of the outcome (e.g., by prior satiation of the animal to a particular food) in a behavioral setting where animals learn to react to stimuli by performing actions. Here in goal-directed behavior is marked by a rapid decline in action execution after outcome devaluation, whereas habitual responses to stimuli will persist given the long-term strength of the underlying stimulus-response association (Balleine and Dickinson, 1998). The neural system mediating goal-directed behavior in mammals is thought

to comprise a different set of brain structures (i.e., medial prefrontal and orbitofrontal cortices, hippocampus, ventral and dorsomedial striatum) than that implied in habit learning and performance (i.e., sensorimotor cortices and dorsolateral striatum; Balleine and Dickinson, 1998; Corbit and Balleine, 2000; Yin et al., 2004; Pezzulo et al., 2014; O'Hare et al., 2016; Pennartz, 2018); the current evidence on goal-directed behavior as defined above mainly pertains to rodents, marmosets and humans. Even though this definition of goal-directed behavior does not entail that it requires situations and decisions to be "complex" (in the sense meant above) for it to occur, it is reasonable to link the concepts of goal-directed behavior and complex decision-making on two grounds.

First, it allows behaviors to be distinguished from habitual (and by extension reflexive) behaviors, which can be placed largely in the domain of unaware information processing and non-complex decision-making. Second, the set of goal-directed behavior-related brain structures exhibits a strong overlap with brain areas implied in model-based learning, which intimately relates to complex decision-making and thereby contrasts with its habitual counterpart, model-free learning (Daw et al., 2005). Model-based learning is defined by the subject building an explicit, internal model of its state space, including specific stimulus–outcome relationships and enabling prospective cognition and on-the-fly decision-making (Daw and Dayan, 2014; Pennartz, 2018). Sometimes it is held to be identical to learning goal-directed behavior, but it is of note the two concepts are defined differently. Whereas goal-directed behavior emphasizes knowledge of outcome value and the causal dependence of outcome on operant action, model-based learning focuses on the agent building a sensory-specific model of stimulus-action-outcome relationships, on prospective activity that anticipates future events, and the ability to make decisions on the fly, spontaneously, based on generalizable internal models (Daw and Dayan, 2004). The importance of spontaneously arising behavior will be highlighted in "Consciousness in Intelligent Machines and Its Relations to Animal Cognition" section. Despite these different emphases, the two concepts conspire to suggest a common brain system supporting complex, non-habitual learning and planned decision-making, which not only depends on declarative memory recall but also on the on-line situational survey typically associated with



consciousness. Recent studies on neurophysiological correlates of goal representations and deliberation in mammals have underscored the common involvement of the set of brain structures already indicated (medial prefrontal and orbitofrontal cortices, hippocampus, ventral and dorsomedial striatum) in both goal-directed behavior and model-based learning (Johnson and Redish, 2005; Hok et al., 2012; Pezzulo et al., 2014; Genovesio et al., 2015; Wikenheiser and Redish, 2015). The complexity of processes underlying model-based learning exceeds that of basic forms of learning such as Pavlovian conditioning, which can proceed under unaware conditions such as anesthesia or sleep (Weinberger et al., 1984; Fifer et al., 2010). Thus, our position diverges from (Ressler, 2004) who argued that learning *per se* may be used as criterion for consciousness, as illustrated by conditioning in honeybees.

Having identified goal-directed behavior and model-based learning as indicators of consciousness that plausibly apply to primates and rodents, we will briefly compare the evidence with that obtained in another class of animals: birds. Space is lacking to review indicators of consciousness for amphibians, reptiles and invertebrate species, such as cephalopods and insects, noting that detailed knowledge on the indicators of consciousness reviewed here is often lacking for these species. goal-directed behavior and model-based learning have not yet been examined in great depth in non-mammalian tetrapods, but certain bird species display behaviors which are, arguably, tightly linked to goal-directed behavior, such as tool manufacturing and theory of mind (Butler and Cotterill, 2006). For instance, New Caledonian crows (*Corvus moneduloides*) perform complex multi-step procedures in fabricating tools (like wires bent into hooks) to gain access to food, which suggests the presence of goal representations and planning actions forming a causal chain leading up to reaching a pre-set goal (Weir et al., 2002; Butler and Cotterill, 2006; but see Suddendorf and Corballis, 2010). Likewise, theory-of-mind capacities (attributing a mental state to other subjects, allowing to predict the other's future behavior) have been suggested by specific food re-caching behaviors and deceptive strategies in scrub jays and ravens (Clayton et al., 2003; Bugnyar and Kotrschal, 2004). Because efficacious deceit or misinformation presupposes the representation of a pre-set goal (in this case, to secure cached food in view of competitors able to retrieve the same food), this behavior strongly suggests goal-directed behavior and, *a fortiori*, the internal modeling of another subject's mental state vis-à-vis one's own state.

At this point, we have functionally linked consciousness to planned or deliberate goal-directed behavior and model-based learning, but in order to define practically useful indicators of consciousness, it is necessary to scrutinize how tight this linkage may be. If an animal displays goal-directed behavior or model-based learning, does this imply that it *must* be conscious? There are two principal reasons arguing that this relationship is not that straightforward. First, neither the definition of goal-directed behavior nor model-based learning necessarily implies that the subject *must* have a multimodal, situational survey to be able to execute its behavior optimally. For goal-directed behavior, neither the causal action

contingency nor the anticipation of outcome value strictly requires consciousness. Similarly, in model-based learning, the acquisition of internal, sensory-specific stimulus-action-outcome models poses no formal requirement on the presence of conscious experience. Internal models of the current world are informative as regards upcoming decisions in two ways: they can be parsed into representations we are aware of (i.e., internally generated models of what is currently going on in our environment to cause the sensations we have) or remain unaware of (models of the causal structure of the environment in terms of what we cannot perceive directly, e.g., unobservable variables that cause you to have a dish with food in front of you).

Second, subjects—whether biological or artificial—may display goal-directed behavior or model-based learning that may arise in a different way than *via* a dependence on conscious experience, no matter how biologically useful this may be. This argument will be developed further in “Consciousness in Intelligent Machines and Its Relations to Animal Cognition” section. In summary, it appears more reasonable to consider goal-directed behavior and related behavioral expressions of model-based learning as indicators of consciousness rather than as indisputable evidence. With “indicator” we mean that a particular type of behavior or neural systems property yields externally observable evidence that the organism under scrutiny is likely to sustain some form of conscious experience—not necessarily having the same phenomenology as human conscious experience, and not necessarily implying proof of consciousness. Given these arguments, as well as other objections to regarding any specific type of non-verbal behavior as sole evidence for consciousness in animals (Weiskrantz, 1995; Seth et al., 2005), it is warranted to widen the search for indicators of consciousness beyond a single indicator of consciousness (such as goal-directed behavior) and next we will consider whether different indicators of consciousness produce a coherent picture for our two case studies (rodents and birds).

## Brain Anatomy and Physiology

The rationale for including these domains of study in our list of indicators of consciousness has been reviewed before and rests on extensive evidence, derived from studies in human patients and non-human primates, that conscious experience is generally dependent on the integrity of thalamocortical systems (Penfield, 1959; Zeki, 1993; Stoerig and Cowey, 1997; Weiskrantz, 1997; Koch, 2004; Key, 2015; Pennartz, 2015). Even though decorticate preparations may still allow interesting and adaptive behaviors to emerge if subcortical systems remain intact (Merker, 2007), extensive evidence on cortical and subcortical brain lesions has linked conscious experience in specific modalities or submodalities to the loss of specific neocortical areas, such as area MT/V5 to conscious motion vision, area V4/V4alpha to color vision, and the ventromedial structures of occipitotemporal cortex (comprising the fusiform and lingual gyri in humans) to form or shape vision (Milner and Goodale, 2008; Karnath et al., 2009; for a discussion of other sensory modalities, see Pennartz, 2015). In the absence of a functioning neocortex, subcortical systems such as the basal ganglia and superior

colliculus can continue to mediate sensorimotor behaviors such as orientation responses, feeding behavior and eye movements, but decorticate rats appear to be severely impaired in performing more complex visuospatial behaviors and become easily trapped on platforms and in alleys (Whishaw et al., 1981; for decorticate cats, see Bjursten et al., 1976). Here, we agree with Seth et al. (2005), proposing that animals be considered conscious when their brains possess at least functional analogs of cortex and thalamus.

Even though the rodent thalamocortical system is obviously much smaller in size than that in humans, it contains the same core components, such as the thalamic sensory relay nuclei, the nucleus reticularis, the intralaminar nuclei, and a complex of sensory cortical areas (Krubitzer, 2007) characterized by a hierarchical organization of lower and higher processing stations (Burkhalter, 2016; D'Souza et al., 2016). Similarly, the presence of neuromodulatory systems in the rodent brain stem and mes- and di-encephalon, acting as “enabling factors” for conscious processing, is well recognized, as well as the presence of cortico-basal ganglia thalamic loops subserving selection of action strategies, individual actions, skill learning and long-term goals (Hasselmo, 1995; Groenewegen and Uylings, 2000; Castro-Alamancos and Calcagnotto, 2001; Pennartz et al., 2011). Although major differences between rodent and human brains in terms of size, complexity and the presence of specialized areas should be acknowledged, we argue that the “size” argument will rather affect the complexity and/or intensity of the information the organism will be conscious of, and not so much the presence or absence of consciousness. Furthermore, rodent brains are likely lacking specialized areas such as a well-developed dorsolateral prefrontal cortex, but lesion studies in patients indicate that these areas are not required for consciousness (see e.g., Barbey et al., 2013 on dorsolateral prefrontal cortex). Moreover, prefrontal cortex is present in rodents in at least basic form (Groenewegen and Uylings, 2000; Uylings et al., 2003), and has been implicated in complex cognitive functions (e.g., prospection, evaluation, planning, instrumental and model-based learning, observational learning; Kametani and Kesner, 1989; Kesner et al., 1989; Mulder et al., 2003; Jurado-Parras et al., 2012; Daw and Dayan, 2014; Pezzulo et al., 2014; see also below).

A comparison of brain state physiology between rodents and humans confirms a great cross-species similarity in states associated with wakefulness (conscious state) vs. non-REM sleep (with deep sleep or slow-wave sleep representing an unaware state) and REM sleep (correlated with dream states, regarded as an altered form of consciousness; Hobson, 2009; Pennartz, 2015). Whereas wakeful and REM sleep states are globally characterized in both rodents and primates (including humans) by a desynchronized cortical EEG pattern, by low-power, high-frequency oscillations (e.g., gamma activity) and sparse, irregular firing of thalamic and cortical neurons, deep non-REM sleep is marked by strongly synchronized Delta waves (1–4 Hz), correlating with Down (quiet, hyperpolarized) and Up states (firing, depolarized) in both taxonomic orders. Again, interspecies differences between sleep-wake electrophysiology must be acknowledged, but when comparing rodents with

humans, the similarities in the correlations between, on the one hand, wakefulness, REM sleep and desynchronized thalamocortical states, and on the other hand, non-REM sleep and slow-waves, spindles and hippocampal ripples, is striking (e.g., Pennartz et al., 2002; Cantero et al., 2003; Steriade, 2006; Zhang et al., 2018). Thus, the presence of a thalamocortical system displaying similar physiological markers of wakeful and sleep states in rodents is proposed as an indicator of consciousness, although we add that the *absence* of such a system does not imply permanent lack of consciousness, because neural systems may, throughout evolution, have evolved to generate conscious experience in different ways. Moreover, the presence of a thalamocortical system in a desynchronized state may not be *sufficient* for sustaining conscious experience.

As concerns avian brains, it has been noted that a complex of pallial structures shows considerable similarities to the corticothalamic system in mammals, as this complex harbors multiple non-limbic, sensory and associational areas supplied by inputs from many sensory modalities (Butler and Cotterill, 2006). In addition to thalamic relay nuclei and basal ganglia-based loops, avian and mammalian brains share the presence of the thalamic reticular nucleus—as an important structure providing an inhibitory, pacing influence on the palliothalamic or corticothalamic system (Butler and Cotterill, 2006). Although birds lack a claustrum as well as the laminar cytoarchitecture typical of mammalian neocortex, these anatomical features may not be essential for consciousness. Indeed, a comparison of mammalian neocortical structures implied in the perception of different sensory modalities indicates that some “typical” neocortical features are probably not universally required across the modalities, such as a thalamic relay station and a receptive layer 4 in the case of olfaction (Pennartz, 2015). Moreover, the neural dynamics of wakefulness, REM and deep non-REM sleep show the same global patterns of transition in birds as in mammals (Campbell and Tobler, 1984; Kavanau, 2002). It is generally accepted that terrestrial mammals and birds share slow-wave and REM sleep phenomena (Libourel et al., 2018). In conclusion, similarities in brain anatomy and physiology between humans and other vertebrates generally appear to be useful indicators of consciousness and are consistent with some degree of consciousness in rodents and birds. For invertebrates, however, the lack of similarity does not necessarily entail absence of consciousness, necessitating consideration of other criteria.

## Psychometrics and Meta-cognitive Judgment

Following earlier work that established similarities between humans and monkeys in psychometric curves for stimulus detection and discrimination (e.g., Mountcastle et al., 1972; Britten et al., 1992; Spinozzi et al., 2004), more recent studies indicated that also rodents display sigmoid psychometric curves for stimulus detection and discrimination comparable to those of humans (Histed et al., 2012; Brunton et al., 2013; Meijer et al., 2018; see Fay, 1988; Andrew and Greenspan, 1999; Lemon, 2015). While rodents were traditionally viewed as “visually handicapped” due to the lower spatial resolution of

their vision (Huberman and Niell, 2011), this is a matter of gradation and should be considered in the context of several visual capabilities at which rodents excel, such as visual contrast sensitivity (Histed et al., 2012; Carandini and Churchland, 2013; Montijn et al., 2015). Moreover, rodents display superb detection and/or discrimination sensitivity in several non-visual modalities (e.g., olfaction, hearing and somatosensation; Fay, 1988; van Duuren et al., 2007; Guo et al., 2014). Their ability to integrate information from multiple senses is close to statistically optimal, as has been found in humans (Raposo et al., 2012). Similarly, detection and discrimination capacities have been well documented in several bird species, and their superior visual acuity and stereovision are particularly striking (Uhlrich et al., 1981; van der Willigen et al., 2010). Taken together with similarities in brain anatomy and physiology (see above), such comparable psychometric performance—found in spite of large differences in brain size and complexity—may be considered an indicator of consciousness, although there is an important caveat that should not be ignored.

Studies on blindsight in humans have made it clear that large lesions of area V1 can preserve detection and discrimination of stimuli that are projected in those parts of the visual field corresponding to the damaged part of the retinotopic map, whereas subjects report that they do not have a visual experience of the presented stimulus. Thus, psychometric performance as tested classically may to some extent be preserved, but when prompted for a “commentary” or meta-cognitive statement about the situation, subjects report an absence of stimulus awareness. In a landmark study, Cowey and Stoerig (1995) showed that monkeys with experimentally controlled V1 lesions exhibit blindsight in that they were able to make above-chance correct discriminations of stimulus location in a forced-choice paradigm, whereas they indicated they had not seen the stimulus when given the option to press a commentary key (which in this paradigm replaced the verbal-response option humans normally have). Although preliminary data from our lab suggest that also rodents can utilize a “Not Seen” response option, caution is warranted for two reasons.

First, even these metacognitive judgment responses can become automatized given a sufficient amount of training or repetition (Weiskrantz, 1995; compare this to a person stereotypically answering “yes, I am listening” to another person who constantly demands attention). This confound is also at play when considering the wider literature on metacognitive judgments in the context of perception, stimulus valuation and consciousness (for examples in the literature on humans, rodents or birds; see Persaud et al., 2007; Kepecs et al., 2008; Seth, 2008; van Wingerden et al., 2012; Watanabe et al., 2014). Thus, caution should be applied when considering potentially habitual expressions of judgment as additional evidence for consciousness. Second, choosing the “Not Seen” option on a physical response panel inevitably coheres with the amount and probability of pay-off attached to this option, compared to other response options. When, for instance, the reward probability for the “Not Seen” option is high compared to one or several of the “Seen” options, the animal may choose

“Not Seen” even when it did detect the presented stimulus. To our knowledge, these two problems have not yet been fully resolved in studies on animal behavior, but this can be done in principle by: (i) keeping reward parameters equal for detection vs. no-detection options; and (ii) testing whether responses in these paradigms conform to habitual or goal-directed behavior (see above).

Overall, we propose that the combination of psychometric performance and metacognitive judgment may be used as an indicator of consciousness, if appropriate controls are included to rule out habitual responding or other behavioral confounds.

## Episodic Memory

This type of memory is defined as autobiographical memory, that is memory of events (“what”) a subject experienced at a particular place (“where”) and time (“when”). Together with its decontextualized counterpart, semantic memory, it constitutes declarative memory: memory that humans can consciously recall and verbally report about (Milner et al., 1998; Kandel et al., 2000). Conscious recall is usually understood as being coupled to conscious experiencing of the event before being stored in declarative memory, warranting the proposal that declarative memory and conscious experiencing are tightly and bidirectionally linked to each other (Tulving, 2002; Pennartz, 2018). This opens the possibility to study consciousness through the “backdoor,” i.e., *via* retrievable traces that are left in episodic memory once a consciously experienced event is over. If we can show that certain species display memory capacities similar or identical to those found for human episodic memory, such evidence may provide a strong indicator of consciousness.

Evidence for episodic-like memory in rodents has accumulated over the past decades. Neurons in rodent hippocampus—the brain structure most unambiguously linked to episodic memory—not only display location-coding properties (“place cells”; O’Keefe and Dostrovsky, 1971; Wilson and McNaughton, 1993) but also, depending on behavioral task requirements, a form of temporal coding (“time cells”; Eichenbaum, 2014). The coding of events (“what”), however, is more multi-faceted and greatly depends on the nature of the event and the sensory modalities and submodalities involved. While events such as reward delivery, reward-predicting sensory cues, social interactions or salient but neutral sensory changes in the environment all affect hippocampal coding by firing-rate modulation (Leutgeb et al., 2005; Lansink et al., 2012; Aronov et al., 2017; Danjo et al., 2018), it should be kept in mind that many sensory, spatial and motor-related aspects of consciously experienced events are prominently coded in neocortical areas (see above; e.g., area MT/V5 for motion vision). The causal importance of the hippocampal memory system may not lie in an all-inclusive storage of what-where-when details of the experience, but rather in providing a spatiotemporal scaffold with which individual sensory elements, coded in connected neocortical structures, can be associated during the formation of long-term explicit memory. Upon partial cue presentation, this scaffold can concurrently subserve memory retrieval (pattern

completion; Teyler and DiScenna, 1986; Nakazawa et al., 2002; Leutgeb and Leutgeb, 2007; Eichenbaum, 2017).

In addition to neurophysiological research on episodic memory substrates in rodents, behavioral studies have addressed whether rats and mice possess an episodic-like memory that can be behaviorally utilized to retrieve what-where-when knowledge. Some studies suggest that rats are able to remember where, in a radial maze, a preferred reward was previously encountered, and how long ago this happened (Babb and Crystal, 2005; Dere et al., 2006); see also (Veyrac et al., 2015). Other studies, however, suggest that rats can remember some episodic aspects of food encounter events (e.g., location and quality of food items) but lack the capacity to alter caching strategies if food quality is subject to time-dependent decay (Roberts, 2002; Bird et al., 2003; McKenzie et al., 2005; Dere et al., 2006). Thus, whereas electrophysiological recording studies have underscored that the what-where-when aspects of episodic memory are differentially coded in corticohippocampal systems, the integral use of episodic memory in optimizing behavioral strategies remains to be investigated in more detail.

One important aspect of episodic memory should not go unmentioned in this discussion: the capacity for mental time travel (Tulving, 2002). The retrieval of episodic memories enables subjects to generate and re-experience events in the past and future, either in realistic or fictive situations. This has been argued to subserve prospective cognitive processes in both humans and rodents (Corballis, 2013), such as route planning (Pezzulo et al., 2014; Redish, 2016) and internal simulation of behavioral strategies, which can be captured by the term “internal hypothesis testing.” Behavioral and lesioning experiments have yielded evidence for prospective and retrospective cognition in rodents (but see Roberts and Feeney, 2009). In a 12-arm radial maze, Cook et al. (1985) obtained evidence for retrospective and prospective use of spatial memory from the strategies rats deployed to visit food locations in the maze, although, notably, the time span for prospecting was less than 1 h. Working in a similar behavioral paradigm addressing potential mental time travel, Kametani and Kesner (1989), Kesner et al. (1989) found that lesions of medial prefrontal cortex and parietal cortex impaired the use of prospective memory strategies.

Extensive work in rodent neurophysiology has raised consistent evidence for the replay of past behavioral experiences during sleep or during behavioral pauses in between action sequences, and this replay of neural sequences generally involves the hippocampus and its connected cortical and subcortical structures (Ji and Wilson, 2007; Pezzulo et al., 2014; Foster, 2017). As argued elsewhere (Pennartz, 2015), replay occurring during hippocampal ripples is probably not coupled to conscious experience. However, the hippocampus has also been shown to generate prospective neural sequences during active locomotion and deliberate (vicarious trial-and-error) behaviors (Johnson and Redish, 2005; Redish, 2016), which is more likely associated with active path planning, goal-directed behavior and wakeful, conscious processing (Kaplan et al., 2017; Pennartz, 2018).

Next to rodents and other mammalian species, evidence for episodic memory in bird species is relatively strong. Clayton and Dickinson (1998), Emery and Clayton (2001), Zentall et al. (2001), Clayton et al. (2003) and Salwiczek et al. (2010) showed that memory in scrub jays is marked not only by “what” happened at a particular location (“where,” i.e., where a food item was cached) but also how much time elapsed since the caching had taken place. In addition, they found evidence for forward planning of actions in the future (e.g., prospective caching) and anticipation of future needs (Correia et al., 2007) which aligns with a capacity for “mental time travel” as another hallmark of episodic memory (Tulving, 2002). However, the latter claim has been criticized by Suddendorf and Corballis (Suddendorf and Corballis, 2010), arguing that food-caching scrub jays did not anticipate on a future desire for specific food, as they did not prefer to store food that would become more desirable in the future. Despite such remaining uncertainties, the case for episodic memory as an indicator of consciousness can be argued to be at least as strong for these bird species as it is for rats or mice.

## Illusion and Multistable Perception

A further indicator of consciousness is related to its definition as multimodal, situational survey, marked by the key feature of intentionality. The subjective nature of conscious experience holds that perceived stimuli or situations may be interpreted in different ways, such as when we view ambiguous stimuli or are subject to illusions. From the perspective of psychophysics, illusion and multistability are important hallmarks of an inferential brain. Basically, it is impossible to experience an illusion unless the brain harbors an internal model that not only accounts for the causes of a particular pattern of sensory input, but also can cause perception to deviate from the actual process generating sensations, in case of ambiguities in their interpretation. Thus, the capacity to misperceive speaks to perception as an active process that discloses an important aspect of conscious processing. As an example of multistable perception, we refer to ambiguous figures such as the Necker cube, where the subject’s percept can switch between two quasi-stable states (e.g., front side of the cube being on the lower right or upper left). Susceptibility to illusions and perceptual ambiguity has been demonstrated in non-human primates and cats (e.g., Logothetis and Schall, 1989; Sheinberg and Logothetis, 1997; Nieder, 2002; Parker et al., 2002; von der Heydt, 2013) and has been utilized to examine whether neurons in lower and higher areas of the visual cortical hierarchy respond to sensory input *per se* or whether their activity correlates with subjective percepts (Leopold and Logothetis, 1999; von der Heydt, 2013; Panagiotaropoulos et al., 2014). Here, behavioral responses to illusion or multistability inducers can be used as an additional indicator of consciousness: specific behavioral patterns can serve to indicate how the subject “acts out” its subjective experience in case it is prone to illusion, whereas a different behavioral response will occur if the subject is not prone to it. Rhesus monkeys, for instance, have been shown to be susceptible to the Rotating Snake illusion (Agrillo et al., 2015; here the inducer of illusory motion is made up of static, interlocking circles that consist of adjacent blue and



yellow—or graytone—elements, giving rise to the illusion of seeing the circles rotate, especially in one's peripheral field of view; Murakami et al., 2006). This susceptibility was expressed by the monkeys making discriminatory choices between static vs. dynamic stimulus arrays. Susceptibility to illusions has been found even in fish, which challenges the hypothesis that cortical substrates are necessarily required to perceive illusory motion (Gori et al., 2014).

As regards rodents, recent evidence suggests that mice are susceptible to the illusion of perceiving large-field contours (Okuyama-Uchimura and Komai, 2016). In a two-choice visual discrimination task, mice were shown to discern illusory rectangular contours based on “Pacman”-type of figures, as often used in the Kanizsa triangle illusion. Further evidence for illusory percepts in mice comes from studies on the rubber tail illusion, which suggests that mice have a sense of body ownership (in this experiment, a mouse's tail and a fake rubber tail were synchronously stroked, and when subsequently the rubber tail was grasped the mice responded as if their own tail was being touched; Wada et al., 2016). In addition, mice are prone to the motion aftereffect, an illusion of motion that arises after prolonged exposure to motion of an object in one direction (Samonds et al., 2018). Finally, mice have a capacity for amodal completion (Kanizsa et al., 1993), which does not offer evidence for illusion susceptibility *per se* but does argue for an ability to integrate information across large portions of the visual field in order to make inferences about occluded parts of objects. To our knowledge, no evidence has been presented yet in favor of multistable perception in rodents. In conclusion, however, overall the evidence has been accumulating in favor of perception of visual and somatosensory illusions in rodents, and these findings add weight to the overall assessment of consciousness in this taxonomic order.

Similar evidence for susceptibility to illusions and perception of subjective contours has been found in birds. Using a visual setup with grating gaps and phase-shifted abutting gratings, Nieder and Wagner (1999) showed that barn owls can perceive subjective contours that are absent in the retinal image. In single-unit recordings from the visual Wulst area, they identified a significant fraction of neurons generated firing correlates of these contours. In a motion-aftereffect paradigm, Niu et al. (2006) recorded neurons in the pigeon's pretectum responding to real and subjective contours or producing after-responses to cessation of prolonged motion. Pepperberg and Nakayama (2016) provided evidence for subjective contour perception (Kanizsa figures) as well as amodal completion of occluded figures in the gray parrot. Evidence for multistable perception in pigeons was presented by Vetter et al. (2000), showing that subjects switched in their pecking responses between bistable apparent motion patterns made up by flashing LED displays.

## Visuospatial Behavior

Conscious subjects typically perceive non-mobile objects in their environment as stably positioned, even when they roam their environment and scan it by eye movements. This point is illustrated by monkeys from whom the striate cortex had been

surgically removed, resulting in blindsight (Humphrey, 1974; Stoerig and Cowey, 1997; note that other sensory modalities, such as somatosensation, remained intact). Humphrey described a rhesus monkey, Helen, that was nearly without striate cortex and was incapable of detailed shape discrimination or object recognition, but expressed a considerable capacity for visually guided behavior as shown by navigation through a room full of obstacles, by foraging for food items on the floor of a well-known arena, or catching passing flies. Here, the main point of interest in Humphrey's description lies in two aspects of Helen's behavior: first, unlike normal monkeys, Helen was unable to reach out to objects that were out of the line of direct sight and had not been foveated just beforehand. Whereas she was able to locate objects in the peripheral visual field by eye movements, she could not utilize visual information from the periphery to guide arm movements. This can be taken as an indication of a lack of visual situational survey.

Even more striking was the observation that the monkey, when confronted with a piece of black tape stuck to the floor in the midst of surrounding obstacles, would try to pick up the object again and again while roaming the arena, failing to notice that this was the same object at the same place, encountered several times before: “every time she moved away and then caught sight of the tape again she appeared to treat it as a new discovery.” Humphrey (1974) concluded that Helen's visual space was “self-centered,” unable as she was to place objects in a stable spatial framework—which corresponds remarkably well to the feature of consciousness of having a situational survey that has both a dynamic (updatable) and stable nature. Thus, the availability of a stable world representation in the presence of ego-motion can be “acted out” by the subject in its visuospatial behavior, and indeed we propose that this specific type of behavior can be used as yet another indicator of consciousness.

Whether rodents display similar visuospatial behavior, based on constructing a stable world survey, is a question that is hard to answer exactly at present, but it is well known that mice and rats, navigating through an arena, rapidly learn to ignore objects occupying a stable and familiar position in space, whereas they show exploratory behavior (i.e., approaches, sniffing) towards objects with novel properties including a changed location (Bekinschtein et al., 2013; Miranda et al., 2018). It should be noted, however, that the spontaneous location recognition task of Bekinschtein et al. (2013) does not only rely on stable object representation but also on memory for object location. Thus, more rigorous tests of stable object representation within perceived scenes should be designed and tested in rodents.

Notwithstanding the current incompleteness of evidence for visuospatial behavior as an indicator of consciousness in rodents, several other studies provide complementary indications. In a visuospatial attention task, Yang et al. (2017) showed that rats can attend to four different locations on a maze floor in order to conduct stimulus-guided locomotor responses, consistent with an ability to direct attention towards spots in a visuospatial survey. Going back to the decorticate rats studied in Whishaw et al. (1981), it is striking that these animals did not only get easily trapped in relatively simple spatial configurations,

but also failed to orient to a location in space where their bodies had been touched just before (**Figure 1**; an exception being a touch on their snout). When a stimulus was applied to body parts such as shoulders, limbs, paws or tail, the rats could display reactions such as turning or rearing, but their responses were not directed towards the stimulus or to the spatial location where the stimulus had been given. This suggests a failure to integrate spatial information about stimulus location, body positioning and directional responding, consistent with (although not conclusive for) the absence of a multimodal, situational survey in decorticate animals.

Visuospatial behavior has been less intensively investigated in birds than mammals, but it is noteworthy that several bird species show evidence of object constancy in perception. This phenomenon refers to the inference that an object will continue to exist after it has perceptually disappeared and occurs at different levels of complexity. Several Piagetian stages of object constancy have been indicated in ring doves, magpies and African gray parrots, which show behavioral signs of surprise and anger if a moving object that they are tracking disappears and is subsequently replaced by a different, unexpected object (Dumas and Wilkie, 1995; Pepperberg et al., 1997; Butler and Cotterill, 2006). Furthermore, evidence has been gathered to suggest that African gray parrots, facing changes that have been made in a visually presented array of objects, are able to indicate that something in the perceived situation has been changed, as well as what was changed (Pepperberg, 2002).

## CONCLUSIONS ON INDICATORS OF CONSCIOUSNESS

Which conclusions can be drawn on the validity of the indicators of consciousness for animal consciousness as proposed above? Would satisfaction of only one criterion be sufficient to conclude that a species under scrutiny possesses consciousness—or else, how many indicators of consciousness should minimally be met? First, none of the proposed indicators of consciousness is “hard” in the sense that its satisfaction would offer proof of consciousness. For every indicator of consciousness, it seems possible that one could devise a computational mechanism or multi-area model that could mimic the intended function, without being forced to invoke consciousness in one’s explanation. Yet, this very lack of a one-to-one correspondence between individual computational mechanisms and conscious experience should not be taken to mean that indicators of consciousness are worthless. The basic reason for this gap is that the key features of conscious experience (see “What Does ‘Consciousness’ Mean and What Are Its Key Properties?” section) reflect subjective experience, whereas the proposed indicators of consciousness were set to satisfy the requirement that they be testable by external observation of non-verbal subjects, in this case, animals. Externally observed behaviors, or explanatory schemes or models of neural mechanisms underlying it, should not be expected to reveal directly whether a subject’s experiences are qualitatively rich, have intentionality, etc., because these are not the sort of features

that computational models (i.e., models transforming input numbers into output numbers) could be expected to explain or reproduce (Pennartz, 2015). Vice versa, and contrary to Gutfreund (2017), we argue that an increased understanding of the computational mechanisms underlying cognitive phenomena (including conscious experience) in particular animal species does not make it less likely that this species is conscious, because computation and consciousness are not two phenomena standing in opposition to each other: they should be preferably conceived of as forming different levels of representation (computations being situated at a lower representational level than phenomenal consciousness; Pennartz, 2015).

Second, it is notable that for certain taxonomic orders (rodents) or classes (birds) of animals, the assessments across the various indicators of consciousness agree remarkably well with each other, as far as current evidence is permitting. Rodents and birds basically score positively on all indicators of consciousness (goal-directed behavior and model-based learning; anatomy and physiology; psychometrics and metacognitive judgment; episodic memory; illusion induction; visuospatial behavior), although for many criteria more empirical evidence is required to substantiate the particular claims. Rather than emphasizing any singular indicator of consciousness as a criterion for consciousness, we propose that the *consistency* amongst these indicators of consciousness serves to enhance or weaken the case for consciousness in any particular species. This proposal follows the logic that, if there is evidence for *X AND Y AND Z* (etc., where *X*, *Y* and *Z* are features of conscious experience), then the probability of some form of consciousness being present is increased. In other words, the integration of scores across the various indicators of consciousness can be used as a *graded* criterion for consciousness, in a similar vein as when unresponsive, brain-damaged patients are subjected to various neurological tests and their scores are summed up to determine the overall grade on the Glasgow Coma Scale (Teasdale et al., 2014). Another comparison coming to mind is nosological psychiatry, having resulted in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) where a number of symptoms of disorder *X* (e.g., bipolar depression) has to be present for a minimum period of time in order to diagnose the patient as suffering from *X* (American Psychiatric Association, 2013). At the same time, our approach differs from nosological psychiatry because it offers as yet a fairly general heuristic to deal with the problem of consciousness in animals and machines; a concrete and quantitative system (e.g., a linear scale with points) will take more time to develop. We also note that our methodology is species-dependent (e.g., the weight of the anatomy and physiology indicator will be different for mammals vs. invertebrates). Moreover, our indicators refer to larger aggregates of behavioral patterns (e.g., visuospatial behavior) rather than concrete elementary behaviors such as used on the Glasgow Coma Scale.

A logical consequence of a high consistency amongst indicators of consciousness, when applied to certain animal species, is to label the set of behaviors showing this consistency such that this is properly distinguished from behaviors showing either a low consistency or scoring negatively on all points. In

analogy to the term “episodic-like memory” that is used for animals that cannot make verbal declarations about what they exactly recall from memory, we propose the term “conscious-like behavior” as a summary term for animals displaying consistently positive scores on behavioral indicators of consciousness. Introducing this term is consistent with abductive reasoning as a heuristic—not, in this case, as a characteristic for how a conscious brain system works, but to assess various consciousness theories relative to one another: one makes an inference to the best explanation of the observable data, pertaining here to animal behavior and other indicators arguing in favor or against consciousness in non-human species.

## THEORETICALLY DERIVED MEASURES, CROSS-VALIDATION AND THE PROBLEM OF PANPSYCHISM

Consciousness can be decoupled from one’s verbal capacities but can also persist in the absence of observable motor behavior, such as during paralysis, dreaming and locked-in syndrome. It is therefore mandatory to ask whether criteria or indicators of consciousness can be derived from parameters pertaining purely to internal brain processes. Following up on brain physiology (indicator of consciousness #2), EEG recordings can be utilized in clinical practice to derive the bispectral index, which has been proposed as a measure of awareness vs. depth of anesthesia (for a critical review, see Stein and Glick, 2016). Although such indices are useful for evaluating consciousness in humans, they have two significant disadvantages: first, they may be specific to the human brain and may not apply to nervous systems that differ substantially in structure and/or function (e.g., of cephalopods). Second, the range of circumstances under which they are valid is incompletely known as they are not derived from a fundamental theory of consciousness. Indeed, a validated theory of consciousness may, in principle, yield a quantitative measure of consciousness that can be applied in addition to other (e.g., behavioral) criteria. Below we will highlight the importance of cross-validation when trying to apply theoretically derived measures in practice.

To illustrate this, it is instructive to consider Information Integration Theory (IIT; Tononi, 2004; Tononi et al., 2016), which assumes that conscious experience is essentially characterized by both differentiation and integration of information. IIT starts from a conceptual description of consciousness with specific axioms and infers the properties that a physical structure should have to be considered conscious (Oizumi et al., 2014; Tononi et al., 2016). The concept of consciousness IIT refers to is so-called primary consciousness, assumed as being equivalent to experience, which does not require any specific cognitive ability (Massimini and Tononi, 2018). To be conscious, according to IIT, a physical system must function not only through feed-forward mechanisms but also through re-entrant processes: a simple input-output mechanism cannot be conscious because its integrated information ( $\Phi$ ) is zero. Consciousness is equated with maximally integrated information, which is defined as the amount of information

generated by a complex of elements, over and above the information of its parts. In turn, information is defined as uncertainty reduction: if more experiential possibilities are ruled out, more information is available and thus  $\Phi$  increases. In contrast to functionalism, IIT argues that an exclusive focus on functions, ignoring physical structure, cannot explain consciousness (Tononi et al., 2016).

$\Phi$  is a gradable measure: different levels of consciousness are possible in different physical systems or even in the same system when this can assume different states of information exchange (e.g., during sleep vs. wakefulness). Even though IIT is not primarily a metaphysical but a scientific theory of consciousness, it defines consciousness as a property that emerges in physical systems organized in a specific way, not as a fundamental feature of reality (Tononi and Koch, 2015). IIT claims that consciousness is identical with the conceptual structure of physical systems characterized by certain postulates, so that consciousness is constitutive and fundamental to such systems, which are not restricted to human brains (Fallon, 2018). For IIT, the states of a neural system that are important for consciousness are those that have maximum cause-effect power on the system itself. Connected neuronal groups have a maximally irreducible cause-effect power on themselves, i.e., they specify a conceptual structure with the highest  $\Phi$  value. Although in practice  $\Phi$  cannot be computed for large systems of interconnected elements, it can be indirectly assessed through neuroimaging or electrophysiological measurements (Tononi et al., 2016). When comparing the indicators of consciousness and theoretical framework proposed here with IIT, both frameworks define consciousness by subjective and experiential features. We describe consciousness in representational and inferential terms, conceptualizing consciousness as resulting from inference about the world, from the generative activity of constructing a continuously updated multimodal, situational survey of the world around us and our bodies. In contrast, IIT focuses on information processing in systems of causally connected elements.

Previously some of us have argued (Pennartz, 2015) that IIT’s criterion for ascribing consciousness to a system is insufficient because integrated information processing constitutes only one component of the type of world-modeling activity we call consciousness. Although IIT is useful in thinking about ways to assess consciousness based on internal brain parameters, it is underconstrained and this raises several problems, of which only two can be highlighted here. First, the theory would be forced to attribute some degree of consciousness to too many systems throughout nature, including non-living entities such as weather systems. Weather systems do not qualify for consciousness under different, more conventional criteria. Systems consisting of many elements engaging in causal interactions and showing statistical dependence on one another will result in considerable amounts of  $\Phi$ , which would move IIT into the domain of panpsychism (Pennartz, 2015). In a blogpost<sup>1</sup>, S. Aaronson reinforced this objection by pointing to mathematical matrices marked by (very) high  $\Phi$  values, noting that devices such as

<sup>1</sup><https://www.scottaaronson.com/blog/?p=1799>

DVD players are capable of strong information integration based on the  $n$ -parity operations they routinely perform. One reply to this objection holds that panpsychism *per se* is not a problem for IIT—on the contrary, it would open up an unconventional view of consciousness that broadens our horizon of things in nature potentially possessing some level of consciousness (Koch, 2012).

A major objection against the argument of panpsychism—as a defense of IIT or in general—relates to a lack of testability, or more specifically, a lack of means for cross-validation. If a given system in nature, e.g., a cyclone, is characterized by a certain amount of  $\Phi$ , how could one validate that this system is indeed, to some extent, conscious? Do systems that should be considered unmistakably non-conscious (e.g., DVD players), yet display high  $\Phi$  values, serve as counterexamples disproving IIT? Vice versa, if a nervous system normally considered to be conscious exhibits low  $\Phi$ , would this argue against IIT, or does any non-zero  $\Phi$  value suffice to make the case for consciousness? Thus, a problem for IIT is that no criteria are offered to allow to test whether it is correct or incorrect. On the other hand, one could maintain that our “normal” criteria to label entities as conscious have been simply too narrow. This argument holds that IIT offers a fundamental, new law that does not require cross-validation against more conventional ways of conceptualizing consciousness.

We believe that this position is flawed because, once again, it would make the theory immune against any form of cross-examination and cross-validation. As with all phenomena throughout nature, a particular phenomenon  $X$  (e.g., magnetism or frozenness) can only be claimed to occur in a system if the system exhibits *manifest* properties indicating that  $X$  is occurring in that system. For instance, iron but not plastic can be said to have magnetic properties because it displays properties of attraction and repulsion in response to an applied magnetic field, even though both iron and plastic are composed of elementary particles obeying the Standard Model of particle physics, which includes electromagnetic interactions. Similarly, consciousness should be attributed only to systems displaying manifest, observable properties such as signs of wakefulness or specific conscious-like behaviors (see “Indicators of Consciousness in Animals” section) or physiological measures (when validated through their tight correlation with other indicators of consciousness).

Thus, the validity of theoretically derived measures of consciousness can only be assessed by testing them against other, usually externally observable phenomena that researchers can reasonably agree on as being indicative of consciousness. Again, this approach follows the logic of abductive reasoning, which compares hypotheses and infers the better model based on simplicity, effectiveness and the available evidence (Josephson and Josephson, 1996). Typically, we attribute awareness to systems that may also assume unaware states from time to time, such as deep sleep or anesthesia. This attribution follows the logic that the term “unawareness” only finds linguistic use if it arbitrates between contrasting states or conditions (aware vs. unaware). Even when consciousness would be more multifaceted than hitherto acknowledged, the point remains that,

if everything in our environment would always be conscious, the concept would lose its usefulness in daily or scientific discourse (Pennartz, 2015).

That IIT is underconstrained as a framework for consciousness raises a second important problem: the quantity it uses to assess consciousness ( $\Phi$ ) does not refer to anything beyond the system’s state it describes. It leaves the key feature of intentionality unresolved, as it does not explain how neural systems can generate contents about something that is different from themselves. In other words, IIT measures do not quantify to what extent a neural (or artificial) system generates beliefs about, or representations of, something. The lack of intentionality in IIT adds to the argumentation related to the problem of panpsychism signaled above.

Which other phenomena, indicative of consciousness, could then be used to test theoretically derived, quantitative measures? Not surprisingly, we refer back to the indicators of consciousness mentioned in “Indicators of Consciousness in Animals” section, noting that they are, on the one hand, linked to the neurorepresentational framework of multimodal, situational survey, but on the other hand are compatible with more than one theory and could thus serve as a more general testbed. Importantly, not only may individual indicators of consciousness provide estimates of conscious state, but especially the consistency amongst scores on different indicators of consciousness can function to provide an overall assessment.

One of the remaining problems is how the conscious status of behaviorally unresponsive patients and other behaviorally incapacitated subjects may be assessed. Currently, no universally accepted criteria are available, but several promising research directions may be indicated. First, decoding fMRI signals evoked by cognitive tasks (e.g., motor imagery) may be used to derive Yes/No answers from unresponsive patients (Laureys et al., 2004; Owen, 2015). A drawback of this approach is that cognition does not entail consciousness; cognitive activity (inferred *via* brain activity correlates) does not necessarily correspond to conscious experience (Jox, 2013; Peterson et al., 2013; Farisco et al., 2015). Second, the complexity of spatiotemporal, cortical EEG patterns evoked by TMS pulses can be used to stratify patients with disorders of consciousness (Gosseries et al., 2014; Casarotto et al., 2016). A third route, derived from the current neurorepresentational theory and currently under development, is to study the differences in representational capacity under aware vs. unaware conditions, i.e., to decode object and scene properties from large amounts of neural signals recorded in parallel and identify those components that are only present during aware but not unaware processing.

All of these neural-processing parameters can be combined with additional measures of local (within-area) and global (across-area) connectivity, such as of coherence, mutual information (and related measures of functional connectivity; e.g., Olcese et al., 2016; Mikulan et al., 2018) and directed information transfer (e.g., transfer entropy; Pal et al., 2016; Olcese et al., 2018a). The practical usefulness of these measures remains to be validated by comparison with clinical assessments either during the period of unresponsiveness or after recovery (Schiff, 2010).



## CONSCIOUSNESS IN INTELLIGENT MACHINES AND ITS RELATIONS TO ANIMAL COGNITION

The question of consciousness in intelligent machines, including AI algorithms and, in some cases, mobile robots, will be only briefly discussed here in light of the indicators of consciousness raised here, as it has been the subject of many reviews and books (e.g., Aleksander, 2001; Holland and Goodman, 2003; Reggia, 2013; Pennartz, 2015; Dehaene et al., 2017). The matter has become more urgent with the advent of deep learning neural networks (DLNNs), capable of recognizing complex input patterns and classifying complex scenes (LeCun et al., 2015; Schmidhuber, 2015), but also recent developments in robotics and AI based on deep reinforcement learning call for a reconsideration of the matter. Current reinforcement learning-based networks are capable of supra-human performance on Alpha-Go (Silver et al., 2017) and have been shown to learn 40 computer games simultaneously (Mnih et al., 2015). If intelligent machines would be capable of such advanced processing that they may be considered conscious, it becomes a pressing issue to discuss whether they are entitled to fundamental rights, such as the indemnification of pain and fear, and even the attribution of moral status. It is only 8 years ago that Koch and Tononi (2011) proposed an alternative Turing-type of test for consciousness in machines, namely whether they would be able to answer arbitrary questions about what is going on in a photograph. Today, DLNNs come remarkably close to this, having not only achieved complex scene classification, but also semantic labeling and verbal description of scenes (Karpathy and Fei-Fei, 2015; LeCun et al., 2015).

The fact that consciousness, as we know it, exists in living entities, does not entail that artificial entities could not theoretically be conscious. Several distinctions need to be drawn here. First, with respect to the nature of the “hardware.” In his critique of functionalism as a “scientific deviation as great as that of behaviorism it has attempted to supplant,” Edelman (1992) pointed out that the nature of the software (i.e., consciousness, in his view) perforce depends on the nature of the hardware (here: living brains and central nervous systems, vs. non-living machines), and that understanding the former presupposes knowledge of the latter. Here, the point is that consciousness cannot be understood well without any understanding of that which is conscious. Whereas in traditional computers a clear distinction can be made between hardware and software, it is unclear whether this dichotomy can be similarly applied to the brain and consciousness. Numerous questions arise: if consciousness would exist in an entity which, by its constitutive nature, is materially different from living brains, would it be similar to ours? By what reasoning may we justify an answer? If not, would this affect our abilities to (a) detect its consciousness, (b) understand or gain knowledge of it, and (c) communicate with it, provided some success in (a) and (b)?

Second, should we use the same term “consciousness” to refer to a different kind of entity (e.g., a living body vs. a

machine, and thus expand our definition accordingly), or create a new term to denote it? This question is epistemologically very important. On the one hand, linguistic innovations must be well-justified in order to avoid conceptual inflation. Yet, if we are too bound by common usage traditions, we hamper development of new thought. Our languages need to evolve to enable the expression of new ideas, knowledge and normative systems. This is not only a matter of communication but also of thought, for language shapes thought both epistemologically, in terms of what we can think and know, and normatively, in terms of the values we develop. These questions present us with conceptual choices that need to be made both philosophically (in terms of clarity, simplicity and logical coherence) and empirically (in terms of scientific justification, experimental validation and explanatory power).

With these questions in mind, there are multiple reasons to posit that the current generation of DLNNs is not conscious, and *a fortiori*, also has no understanding of presented scenes in a sense that resembles human understanding. Even though the viewpoint that state-of-the-art DLNNs have no consciousness is probably uncontroversial, it is important to expose the reasons why this would be the case. First, even DLNNs that are generally successful in semantically labeling scenes correctly can make rare but gross mistakes which reveal that they entirely miss the gist or emotional significance of a scene, or show that these networks lack any basic understanding of how the world physically “works” (Lake et al., 2017). Even after millions of training trials, current DLNNs are susceptible to erroneous classification that is often based on either irrelevant details of presented scenes, or has no clear correspondence to the way humans reach classification, based in part on our abilities for generalization, conceptual learning and selective attention (Schmidhuber, 2015; Lake et al., 2017). In a similar vein, current AI appears incapable of making counter-intuitive inferences on the basis of basic world knowledge and “common sense” and fails to show the associated ability to manifest artificial stupidity.

This first, empirically derived rationale is related to a second, theoretical argument against DLNN scene understanding and consciousness, which reintroduces the point that conscious experience exists by virtue of the brain’s internally producing a model of one’s current situation—our body and the world currently around us. As state-of-the-art architectures amongst DLNNs, deep convolutional neural networks (DCNNs) are characterized by feed-forward processing of inputs, which are transformed to hidden-layer responses before being converted into output (LeCun et al., 2015). There is no place in this feed-forward scheme to incorporate the genesis of an internal model, in contrast to the key properties laid out in “What Does ‘Consciousness’ Mean and What Are Its Key Properties?” section. For instance, neurons or groups of neurons in a DCNN cannot be claimed to have intentionality or qualitative richness. As currently designed, a DCNN is incapable of solving the basic problem of modality identification, i.e., the problem of representing the sensory modality (e.g., vision) providing inputs to a network as being phenomenally distinct from another modality (e.g., audition). Although several classes of DLNNs

are characterized by dense recurrent connectivity, the latter objection applies until proof to the contrary is provided. In this respect, models explicitly aiming to build internal models of the causes of sensory inputs and their environmental settings may offer a more promising avenue for further research (Rao and Ballard, 1999; Bastos et al., 2012; Dora et al., 2018). Recently, generative modeling has found its way into AI, for example in the form of generative adversarial networks (Goodfellow et al., 2014) and variational autoencoders (Kingma and Welling, 2014), offering a potential avenue towards better machine-based scene understanding.

Broadening the discussion to state-of-the-art robots which show “intelligent behavior” in the sense that they can negotiate difficult terrains, solve tasks like opening a door and many other sensorimotor problems (e.g., Murphy et al., 2011), we must next ask whether these artifacts could satisfy one or multiple behavioral indicators of consciousness laid out in “Indicators of Consciousness in Animals” section (**Figure 1**). For instance, a robot capable of opening a door may be argued to exhibit goal-directed behavior in that it may have been pre-programmed to do so, and has thus, in its software, a representation of a prespecified goal, such as escaping from the building where it was manufactured. Similarly, one may adjust the robot’s software so that it would display the appropriate psychometric responses given sensory stimuli, and program it to generate Seen/Unseen commentary responses in analogy to monkeys with blindsight. It would require a substantial amount of work to endow a robot with behaviorally expressed episodic memory, with appropriate reactions to illusion inducers, and with capacities allowing the robot, while roaming around, to approach and act on objects as being stationary in space (but see Eslami et al., 2018).

Even if these behavioral indicators of consciousness would be met, however, two principal objections remain standing against the claim that such robots are conscious. The first is that intelligent machines may rely on processors which are much faster than neurons, which are limited in speed because of their membrane time constant, axonal and synaptic delays, etc. Neuromorphic chips with acceleration factors in the order of  $10^4$ – $10^5$  as compared to biological neurons have been designed and tested (Friedmann et al., 2013, 2017). In the future, such artificial systems may develop alternative ways to satisfy behavioral indicators of consciousness by computing solutions for goal-directed behavioral problems, episodic memory, etc., in different ways than those evolved through biological evolution. Instead of attempting to generate a multimodal, situational survey of one’s environment and body, a robot may, for instance, have processors adopting a computational strategy of serial accumulation of evidence, based on a one-by-one scanning of elements in the environment, avoiding the need to construct the kind of overall, integrated survey we would normally consider “conscious.” It remains an open question whether this scenario might result in an alternative kind of consciousness. Whether and how alternative, non-conscious strategies would be effective is unknown, but in principle, it appears possible that intelligent machines could display such intelligent, non-automated behaviors without necessarily being conscious.

The second objection relates to the notion that intelligent robots may be pre-programmed to display goal-directed behavior and other conscious-like behaviors. There is nothing in the definition of goal-directed behavior that precludes pre-programmed behaviors from satisfying it. For instance, a humanoid robot such as Honda’s Asimo (Hirose and Ogawa, 2007), on display in Tokyo’s National Museum of Emerging Science and Innovation, is capable of kicking a soccer ball into a goal at a distance—which is a remarkable feat of sensorimotor engineering—but it does so in a pre-programmed fashion, and will stereotypically repeat its action pattern on the next show in the same museum. However, closely related to goal-directed behavior is the concept of model-based learning, which emphasizes the importance of systems that do not have to wait until a command or stimulus arrives in order to produce a desired behavior. Agents trained by model-based learning should be able to improvise in novel or unexpected situations, act on the fly, be flexible and act spontaneously when hitherto familiar circumstances are changing (Daw et al., 2005; Daw and Dayan, 2014; Pezzulo et al., 2014). Thus, a more appropriate test for robots, rather than displaying skilled performance *per se*, is to examine whether they can generalize their previously acquired knowledge to novel situations and produce spontaneous and adaptive improvisations when facing environmental changes that demand complex decisions.

These considerations accumulate to suggest a different approach to machine consciousness than a brief Turing-type of test which only provides a snapshot of a machine’s cognitive capacities obtained by interrogation. This approach should take into account that machines, having much faster processing units at their disposal, may use other strategies for solving complex problems than relying on conscious surveys of the environment (and are thereby less dependent on the functions of consciousness that apply in a biological context), and may display externally observable indicators such as goal-directed behavior by pre-programmed solutions. Thus, the biological function of consciousness proposed above may thus not be simply transferable upon machines, and additional criteria may well be needed. We have reviewed evidence for consciousness in several mammalian and avian species and—although this evidence is not completely unambiguous (and is certainly far from complete)—the force of the combined anatomic, physiological and behavioral-cognitive arguments makes a fairly strong case that rodents and birds do have consciousness. Monkeys make an even stronger case. Here, our point, however, is that also intelligent robots can be best evaluated through a process of prolonged, ethological observation as we have suggested for animals but which may be even more comprehensive in the case of robots in order to exclude pre-programmed solutions. In the case of immobile machines, observation of overt behaviors can be replaced by prolonged tests combining complex sensory stimulation patterns with flexible problem solving, interrogation and analysis of symbolically expressed responses.

The rationale is that the evidence for consciousness will optimally accumulate across the study of multiple behaviors (taxing goal-directed behavior, episodic memory, etc.), which should be maintained under varying environmental

circumstances (e.g., with novelty, detours) under which stereotypical, pre-programmed solutions can be excluded. Thus, our proposal to assess machine consciousness is as pluralistic as for animals, but more stringent in excluding AI solutions that need not rely on core features of consciousness. In Antarctic marine habitats, groups of collaborating Orcas display seal hunt behavior in which they collectively generate waves to drive the prey off an ice shelf (see Pitman and Durban, 2012). This is an adaptive, flexible form of behavior arising by virtue of taking into account where an agent's conspecifics are and what they do, how the ice shelf and sea lion are positioned and respond to oncoming waves, where the animal itself is relative to its fellow hunters and ice shelf, etc.—a complex form of behavior that will typically require a situational survey to produce a catch. If robots will be studied in a similar way, for a prolonged time and probing its reliance on its instantaneous capacity to generate survey updates, a robust test of machine consciousness may be within reach. To the best of our knowledge, no robots have been produced yet that would approach passing such a comprehensive test.

## AUTHOR CONTRIBUTIONS

CP, MF and KE all contributed to this article. “Biological Function of Consciousness” and “Indicators of Consciousness in Animals” sections were mainly written by CP. “Theoretically

Derived Measures, Cross-Validation and the Problem of Panpsychism” section was mainly written by MF and CP. “Consciousness in Intelligent Machines and Its Relations to Animal Cognition” section was mainly written by KE and CP. All authors contributed to “Introduction” section. MF made the tables.

## FUNDING

This research was supported by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2).

## ACKNOWLEDGMENTS

We would like to thank Sten Grillner for his comments on decorticate animals, Ned Block and David Chalmers for helpful discussions, Laura Bavelaar for supporting the sections involving consciousness in rats and birds, Angelica da Silva Lantyer for editing support, and Hanna Bodde for help with the artwork. We are grateful for numerous discussions about the topics of this review with colleagues participating in the First International Conference of the Human Brain Project, “Understanding Consciousness—a Scientific Quest for the 21st Century” (Barcelona, June 2018).

## REFERENCES

- Agrillo, C., Gori, S., and Beran, M. J. (2015). Do rhesus monkeys (*Macaca mulatta*) perceive illusory motion? *Anim. Cogn.* 18, 895–910. doi: 10.1007/s10071-015-0860-6
- Aleksander, I. (2001). *How to Build A Mind: Toward Machines with Imagination*. New York, NY: Columbia University Press.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. 5th Edn. Arlington, VA: American Psychiatric Association.
- Andrew, D., and Greenspan, J. D. (1999). Peripheral coding of tonic mechanical cutaneous pain: comparison of nociceptor activity in rat and human psychophysics. *J. Neurophysiol.* 82, 2641–2648. doi: 10.1152/jn.1999.82.5.2641
- Aronov, D., Nevers, R., and Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature* 543, 719–722. doi: 10.1038/nature21692
- Baars, B. J., Franklin, S., and Ramsoy, T. Z. (2013). Global workspace dynamics: cortical “binding and propagation” enables conscious contents. *Front. Psychol.* 4:200. doi: 10.3389/fpsyg.2013.00200
- Baars, B. J., Ramsoy, T. Z., and Laureys, S. (2003). Brain, conscious experience and the observing self. *Trends Neurosci.* 26, 671–675. doi: 10.1016/j.tins.2003.09.015
- Babb, S. J., and Crystal, J. D. (2005). Discrimination of what, when, and where: implications for episodic-like memory in rats. *Learn. Motiv.* 36, 177–189. doi: 10.1016/j.lmot.2005.02.009
- Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419. doi: 10.1016/s0028-3908(98)00033-1
- Barbey, A. K., Colom, R., and Grafman, J. (2013). Dorsolateral prefrontal contributions to human intelligence. *Neuropsychologia* 51, 1361–1369. doi: 10.1016/j.neuropsychologia.2012.05.017
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038
- Bekinschtein, P., Kent, B. A., Oomen, C. A., Clemenson, G. D., Gage, F. H., Saksida, L. M., et al. (2013). BDNF in the dentate gyrus is required for consolidation of “pattern-separated” memories. *Cell Rep.* 5, 759–768. doi: 10.1016/j.celrep.2013.09.027
- Bird, L. R., Roberts, W. A., Abrams, B., Kit, K. A., and Crupi, C. (2003). Spatial memory for food hidden by rats (*Rattus norvegicus*) on the radial maze: studies of memory for where, what, and when. *J. Comp. Psychol.* 117, 176–187. doi: 10.1037/0735-7036.117.2.176
- Bjurstén, L. M., Norrsell, K., and Norrsell, U. (1976). Behavioural repertory of cats without cerebral cortex from infancy. *Exp. Brain Res.* 25, 115–130. doi: 10.1007/bf00234897
- Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., and Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *J. Neurosci.* 37, 9603–9613. doi: 10.1523/jneurosci.3218-16.2017
- Boly, M., Seth, A. K., Wilke, M., Ingmundson, P., Baars, B., Laureys, S., et al. (2013). Consciousness in humans and non-human animals: recent advances and future directions. *Front. Psychol.* 4:625. doi: 10.3389/fpsyg.2013.00625
- Borsellino, A., Carlini, F., Riani, M., Tuccio, M. T., De Marco, A., Penengo, P., et al. (1982). Effects of visual angle on perspective reversal for ambiguous patterns. *Perception* 11, 263–273. doi: 10.1068/p110263
- Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* 12, 4745–4765. doi: 10.1523/jneurosci.12-12-04745.1992
- Bruno, M. A., Vanhaudenhuyse, A., Thibaut, A., Moonen, G., and Laureys, S. (2011). From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: recent advances in our understanding of disorders of consciousness. *J. Neurol.* 258, 1373–1384. doi: 10.1007/s00415-011-6114-x
- Brunton, B. W., Botvinick, M. M., and Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science* 340, 95–98. doi: 10.1126/science.1233912



- Bugnyar, T., and Kotrschal, K. (2004). Leading a conspecific away from food in ravens (*Corvus corax*)? *Anim. Cogn.* 7, 69–76. doi: 10.1007/s10071-003-0189-4
- Burkhalter, A. (2016). “The network for intracortical communication in mouse visual cortex,” in *Micro-, Meso- and Macro-Connectomics of the Brain*, eds H. Kennedy, D. C. Van Essen and Y. Christen (Cham: Springer), 31–43.
- Butler, A. B., and Cotterill, R. M. (2006). Mammalian and avian neuroanatomy and the question of consciousness in birds. *Biol. Bull.* 211, 106–127. doi: 10.2307/4134586
- Campbell, S. S., and Tobler, I. (1984). Animal sleep: a review of sleep duration across phylogeny. *Neurosci. Biobehav. Rev.* 8, 269–300. doi: 10.1016/0149-7634(84)90054-x
- Cantero, J. L., Atienza, M., Stickgold, R., Kahana, M. J., Madsen, J. R., and Kocsis, B. (2003). Sleep-dependent theta oscillations in the human hippocampus and neocortex. *J. Neurosci.* 23, 10897–10903. doi: 10.1523/JNEUROSCI.23-34-10897.2003
- Carandini, M., and Churchland, A. K. (2013). Probing perceptual decisions in rodents. *Nat. Neurosci.* 16, 824–831. doi: 10.1038/nn.3410
- Carruthers, P. (1989). Brute experience. *J. Philos.* 86, 258–269. doi: 10.2307/2027110
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* 5:198ra105. doi: 10.1126/scitranslmed.3006294
- Casaretto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fecchio, M., Napolitani, M., et al. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Ann. Neurol.* 80, 718–729. doi: 10.1002/ana.24779
- Castro-Alamancos, M. A., and Calcagnotto, M. E. (2001). High-pass filtering of corticthalamic activity by neuromodulators released in the thalamus during arousal: *in vitro* and *in vivo*. *J. Neurophysiol.* 85, 1489–1497. doi: 10.1152/jn.2001.85.4.1489
- Churchland, P. M. (1984). *Matter and Consciousness*. Cambridge, MA: MIT Press.
- Clayton, N. S., Bussey, T. J., and Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nat. Rev. Neurosci.* 4, 685–691. doi: 10.1038/nrn1180
- Clayton, N. S., and Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature* 395, 272–274. doi: 10.1038/26216
- Cleeremans, A., Timmermans, B., and Pasquali, A. (2007). Consciousness and metarepresentation: a computational sketch. *Neural Netw.* 20, 1032–1039. doi: 10.1016/j.neunet.2007.09.011
- Cook, R. G., Brown, M. F., and Riley, D. A. (1985). Flexible memory processing by rats: use of prospective and retrospective information in the radial maze. *J. Exp. Psychol. Anim. Behav. Process.* 11, 453–469. doi: 10.1037/0097-7403.11.3.453
- Corballis, M. C. (2013). Mental time travel: a case for evolutionary continuity. *Trends Cogn. Sci.* 17, 5–6. doi: 10.1016/j.tics.2012.10.009
- Corbit, L. H., and Balleine, B. W. (2000). The role of the hippocampus in instrumental conditioning. *J. Neurosci.* 20, 4233–4239. doi: 10.1523/jneurosci.20-11-04233.2000
- Correia, S. P., Dickinson, A., and Clayton, N. S. (2007). Western scrub-jays anticipate future needs independently of their current motivational state. *Curr. Biol.* 17, 856–861. doi: 10.1016/j.cub.2007.03.063
- Cowey, A., and Stoerig, P. (1995). Blindsight in monkeys. *Nature* 373, 247–249. doi: 10.1038/373247a0
- Crick, F., and Koch, C. (2003). A framework for consciousness. *Nat. Neurosci.* 6, 119–126. doi: 10.1038/nn0203-119
- Danjo, T., Toyozumi, T., and Fujisawa, S. (2018). Spatial representations of self and other in the hippocampus. *Science* 359, 213–218. doi: 10.1126/science.aao3898
- Daw, N. D., and Dayan, P. (2004). Neuroscience. Matchmaking. *Science* 304, 1753–1754. doi: 10.1126/science.1099898
- Daw, N. D., and Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130478. doi: 10.1098/rstb.2013.0478
- Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560
- de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., et al. (2018). Shifting the balance between goals and habits: five failures in experimental habit induction. *J. Exp. Psychol. Gen.* 147, 1043–1065. doi: 10.1037/xge0000402
- Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018
- Dehaene, S., Lau, H., and Kouider, S. (2017). What is consciousness and could machines have it? *Science* 358, 486–492. doi: 10.1126/science.aan8871
- Dere, E., Kart-Teke, E., Huston, J. P., and De Souza Silva, M. A. (2006). The case for episodic memory in animals. *Neurosci. Biobehav. Rev.* 30, 1206–1224. doi: 10.1016/j.neubiorev.2006.09.005
- Dickinson, A. (2012). Associative learning and animal cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 2733–2742. doi: 10.1098/rstb.2012.0220
- Dora, S., Pennartz, C., and Bohte, S. (2018). “A deep predictive coding network for inferring hierarchical causes underlying sensory inputs,” in *International Conference on Artificial Neural Networks* (Cham: Springer), 457–467.
- D’Souza, R. D., Meier, A. M., Bista, P., Wang, Q., and Burkhalter, A. (2016). Recruitment of inhibition and excitation across mouse visual cortex depends on the hierarchy of interconnecting areas. *Elife* 5:e19332. doi: 10.7554/eLife.19332
- Dumas, C., and Wilkie, D. M. (1995). Object permanence in ring doves (*Streptopelia risoria*). *J. Comp. Psychol.* 109, 142–150. doi: 10.1037/0735-7036.109.2.142
- Edelman, G. M. (1992). *Bright Air, Brilliant Fire: On the Matter of the Mind*. New York, NY: Basic Books.
- Edelman, D. B., and Seth, A. K. (2009). Animal consciousness: a synthetic approach. *Trends Neurosci.* 32, 476–484. doi: 10.1016/j.tins.2009.05.008
- Eichenbaum, H. (2014). Time cells in the hippocampus: a new dimension for mapping memories. *Nat. Rev. Neurosci.* 15, 732–744. doi: 10.1038/nrn3827
- Eichenbaum, H. (2017). On the integration of space, time, and memory. *Neuron* 95, 1007–1018. doi: 10.1016/j.neuron.2017.06.036
- Emery, N. J., and Clayton, N. S. (2001). Effects of experience and social context on prospective caching strategies by scrub jays. *Nature* 414, 443–446. doi: 10.1038/news011122-13
- Engemann, D. A., Raimondo, F., King, J. R., Rohaut, B., Louppe, G., Faugeras, F., et al. (2018). Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain* 141, 3179–3192. doi: 10.1093/brain/awy251
- Eslami, S. M. A., Jimenez Rezende, D., Besse, F., Viola, F., Morcos, A. S., Garnelo, M., et al. (2018). Neural scene representation and rendering. *Science* 360, 1204–1210. doi: 10.1126/science.aar6170
- Evers, K. (2009). *Neuroéthique: Quand la Matière S’éveille*. Paris: Odile Jacob.
- Evers, K., and Sigman, M. (2013). Possibilities and limits of mind-reading: a neurophilosophical perspective. *Conscious. Cogn.* 22, 887–897. doi: 10.1016/j.concog.2013.05.011
- Fallon, F. (2018). Integrated information theory, searle, and the arbitrariness question. *Rev. Philos. Psychol.* doi: 10.1007/s13164-018-0409-0 [Epub ahead of print].
- Farisco, M., and Evers, K. (2016). *Neurotechnology and Direct Brain Communication. New Insights and Responsibilities Concerning Speechless but Communicative Subjects*. New York, NY: Routledge.
- Farisco, M., Laureys, S., and Evers, K. (2015). Externalization of consciousness. Scientific possibilities and clinical implications. *Curr. Top. Behav. Neurosci.* 19, 205–222. doi: 10.1007/7854\_2014\_338
- Farisco, M., Laureys, S., and Evers, K. (2017). The intrinsic activity of the brain and its relation to levels and disorders of consciousness. *Mind Matter* 15, 197–219.
- Fay, R. R. (1988). Comparative psychoacoustics. *Hear. Res.* 34, 295–305. doi: 10.1016/0378-5955(88)90009-3
- Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., and Mueller, E. T. (2013). Watson: beyond Jeopardy! *Artif. Intell.* 199–200, 93–105. doi: 10.1016/j.artint.2012.06.009
- Fifer, W. P., Byrd, D. L., Kaku, M., Eigsti, I. M., Isler, J. R., Grose-Fifer, J., et al. (2010). Newborn infants learn during sleep. *Proc. Natl. Acad. Sci. U S A* 107, 10320–10323. doi: 10.1073/pnas.1005061107
- Foster, D. J. (2017). Replay comes of age. *Annu. Rev. Neurosci.* 40, 581–602. doi: 10.1146/annurev-neuro-072116-031538



- Friedmann, S., Frémaux, N., Schemmel, J., Gerstner, W., and Meier, K. (2013). Reward-based learning under hardware constraints—using a RISC processor embedded in a neuromorphic substrate. *Front. Neurosci.* 7:160. doi: 10.3389/fnins.2013.00160
- Friedmann, S., Schemmel, J., Grubl, A., Hartel, A., Hock, M., and Meier, K. (2017). Demonstrating hybrid learning in a flexible neuromorphic hardware system. *IEEE Trans. Biomed. Circuits Syst.* 11, 128–142. doi: 10.1109/tbcas.2016.2579164
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Genovesio, A., Cirillo, R., Tsujimoto, S., Mohammad Abdellatif, S., and Wise, S. P. (2015). Automatic comparison of stimulus durations in the primate prefrontal cortex: the neural basis of across-task interference. *J. Neurophysiol.* 114, 48–56. doi: 10.1152/jn.00057.2015
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). “Generative adversarial nets,” in *Proceedings of the 27th International Conference on Neural Information Processing Systems* (Cambridge, MA: MIT Press), 2672–2680.
- Gori, S., Agrillo, C., Dadda, M., and Bisazza, A. (2014). Do fish perceive illusory motion? *Sci. Rep.* 4:6443. doi: 10.1038/srep06443
- Gosseries, O., Thibaut, A., Boly, M., Rosanova, M., Massimini, M., and Laureys, S. (2014). Assessing consciousness in coma and related states using transcranial magnetic stimulation combined with electroencephalography. *Ann. Fr. Anesth. Reanim.* 33, 65–71. doi: 10.1016/j.annfar.2013.11.002
- Gregory, R. L. (1980). Perceptions as hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 181–197. doi: 10.1098/rstb.1980.0090
- Griffin, D. R., and Speck, G. B. (2004). New evidence of animal consciousness. *Anim. Cogn.* 7, 5–18. doi: 10.1007/s10071-003-0203-x
- Groenewegen, H. J., and Uylings, H. B. (2000). The prefrontal cortex and the integration of sensory, limbic and autonomic information. *Prog. Brain Res.* 126, 3–28. doi: 10.1016/s0079-6123(00)26003-2
- Guo, Z. V., Li, N., Huber, D., Ophir, E., Gutnisky, D., Ting, J. T., et al. (2014). Flow of cortical activity underlying a tactile decision in mice. *Neuron* 81, 179–194. doi: 10.1016/j.neuron.2013.10.020
- Gutfreund, Y. (2017). The neuroethological paradox of animal consciousness. *Trends Neurosci.* 40, 196–199. doi: 10.1016/j.tins.2017.02.001
- Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron* 95, 245–258. doi: 10.1016/j.neuron.2017.06.011
- Hasselmo, M. E. (1995). Neuromodulation and cortical function: modeling the physiological basis of behavior. *Behav. Brain Res.* 67, 1–27. doi: 10.1016/0166-4328(94)00113-t
- Hassin, R. R. (2013). Yes it can: on the functional abilities of the human unconscious. *Perspect. Psychol. Sci.* 8, 195–207. doi: 10.1177/1745691612460684
- Hirose, M., and Ogawa, K. (2007). Honda humanoid robots development. *Philos. Trans. A Math. Phys. Eng. Sci.* 365, 11–19. doi: 10.1098/rsta.2006.1917
- Histed, M. H., Carvalho, L. A., and Maunsell, J. H. (2012). Psychophysical measurement of contrast sensitivity in the behaving mouse. *J. Neurophysiol.* 107, 758–765. doi: 10.1152/jn.00609.2011
- Hobson, J. A. (2009). REM sleep and dreaming: towards a theory of protoconsciousness. *Nat. Rev. Neurosci.* 10, 803–813. doi: 10.1038/nrn2716
- Hok, V., Chah, E., Reilly, R. B., and O'Mara, S. M. (2012). Hippocampal dynamics predict interindividual cognitive differences in rats. *J. Neurosci.* 32, 3540–3551. doi: 10.1523/jneurosci.6449-11.2012
- Holland, O., and Goodman, R. (2003). Robots with internal models a route to machine consciousness? *J. Conscious. Stud.* 10, 77–109.
- Huberman, A. D., and Niell, C. M. (2011). What can mice tell us about how vision works? *Trends Neurosci.* 34, 464–473. doi: 10.1016/j.tins.2011.07.002
- Huetzel, S. A., Güzelde, G., and McCarthy, G. (2001). Dissociating the neural mechanisms of visual attention in change detection using functional MRI. *J. Cogn. Neurosci.* 13, 1006–1018. doi: 10.1162/089892901753165908
- Humphrey, N. K. (1974). Vision in a monkey without striate cortex: a case study. *Perception* 3, 241–255. doi: 10.1068/p030241
- Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.
- Ji, D., and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107. doi: 10.1038/nn1825
- Johnson, A., and Redish, A. D. (2005). Hippocampal replay contributes to within session learning in a temporal difference reinforcement learning model. *Neural Netw.* 18, 1163–1171. doi: 10.1016/j.neunet.2005.08.009
- Josephson, J. R., and Josephson, S. G. (1996). *Abductive Inference: Computation, Philosophy, Technology* Cambridge, MA: Cambridge University Press.
- Jox, R. J. (2013). Interface cannot replace interlocution: why the reductionist concept of neuroimaging-based capacity determination fails. *AJOB Neurosci.* 4, 15–17. doi: 10.1080/21507740.2013.827279
- Jurado-Parras, M. T., Gruart, A., and Delgado-García, J. M. (2012). Observational learning in mice can be prevented by medial prefrontal cortex stimulation and enhanced by nucleus accumbens stimulation. *Learn. Mem.* 19, 99–106. doi: 10.1101/lm.024760.111
- Kametani, H., and Kesner, R. P. (1989). Retrospective and prospective coding of information: dissociation of parietal cortex and hippocampal formation. *Behav. Neurosci.* 103, 84–89. doi: 10.1037/0735-7044.103.1.84
- Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). *Principles of Neural Science*. 4th Edn. New York, NY: McGraw-Hill.
- Kanizsa, G., Renzi, P., Conte, S., Compastella, C., and Guerani, L. (1993). Amodal completion in mouse vision. *Perception* 22, 713–721. doi: 10.1068/p220713
- Kaplan, R., King, J., Koster, R., Penny, W. D., Burgess, N., and Friston, K. J. (2017). The neural representation of prospective choice during spatial planning and decisions. *PLoS Biol.* 15:e1002588. doi: 10.1371/journal.pbio.1002588
- Karnath, H. O., Rüter, J., Mandler, A., and Himmelbach, M. (2009). The anatomy of object recognition—visual form agnosia caused by medial occipitotemporal stroke. *J. Neurosci.* 29, 5854–5862. doi: 10.1523/JNEUROSCI.5192-08.2009
- Karpathy, A., and Fei-Fei, L. (Eds) (2015). “Deep visual-semantic alignments for generating image descriptions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Piscataway: IEEE Press), 677–691.
- Kastrup, B. (2017). There is an ‘unconscious,’ but it may well be conscious. *Eur. J. Psychol.* 13, 559–572. doi: 10.5964/ejop.v13i3.1388
- Kavanau, J. L. (2002). REM and NREM sleep as natural accompaniments of the evolution of warm-bloodedness. *Neurosci. Biobehav. Rev.* 26, 889–906. doi: 10.1016/s0149-7634(02)00088-x
- Kepecs, A., Uchida, N., Zariwala, H. A., and Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455, 227–231. doi: 10.1038/nature07200
- Kesner, R. P., Farnsworth, G., and DiMattia, B. V. (1989). Double dissociation of egocentric and allocentric space following medial prefrontal and parietal cortex lesions in the rat. *Behav. Neurosci.* 103, 956–961. doi: 10.1037/0735-7044.103.5.956
- Key, B. (2015). Fish do not feel pain and its implications for understanding phenomenal consciousness. *Biol. Philos.* 30, 149–165. doi: 10.1007/s10539-014-9469-4
- Kingma, D. P., and Welling, M. (2014). *Auto-Encoding Variational Bayes*. LaJolla: International Conference on Learning Representations.
- Koch, C. (2004). *The Quest for Consciousness*. Englewood, Colorado: Roberts and Company.
- Koch, C. (2012). *Consciousness. Confessions of A Romantic Reductionist*. Cambridge, MA: MIT Press.
- Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22
- Koch, C., and Tononi, G. (2011). A test for consciousness. *Sci. Am.* 304, 44–47. doi: 10.1038/scientificamerican0611-44
- Krubitzer, L. (2007). The magnificent compromise: cortical field evolution in mammals. *Neuron* 56, 201–208. doi: 10.1016/j.neuron.2007.10.002
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2017). Building machines that learn and think like people. *Behav. Brain Sci.* 40:e253. doi: 10.1017/s0140525x16001837
- Lansink, C. S., Jackson, J. C., Lankelma, J. V., Ito, R., Robbins, T. W., Everitt, B. J., et al. (2012). Reward cues in space: commonalities and differences in neural coding by hippocampal and ventral striatal ensembles. *J. Neurosci.* 32, 12444–12459. doi: 10.1523/JNEUROSCI.0593-12.2012
- Lau, H., and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci.* 15, 365–373. doi: 10.1016/j.tics.2011.05.009

- Laureys, S., Owen, A. M., and Schiff, N. D. (2004). Brain function in coma, vegetative state, and related disorders. *Lancet Neurol.* 3, 537–546. doi: 10.1016/s1474-4422(04)00852-x
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Lemon, C. H. (2015). Perceptual and neural responses to sweet taste in humans and rodents. *Chemosens Percept.* 8, 46–52. doi: 10.1007/s12078-015-9177-8
- Leopold, D. A., and Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends Cogn. Sci.* 3, 254–264. doi: 10.1016/s1364-6613(99)01332-7
- Leutgeb, S., and Leutgeb, J. K. (2007). Pattern separation, pattern completion and new neuronal codes within a continuous CA3 map. *Learn. Mem.* 14, 745–757. doi: 10.1101/lm.703907
- Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L., and Moser, M. B. (2005). Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science* 309, 619–623. doi: 10.1126/science.1114037
- Libourel, P.-A., Barrillot, B., Arthaud, S., Massot, B., Morel, A.-L., Beuf, O., et al. (2018). Partial homologies between sleep states in lizards, mammals, and birds suggest a complex evolution of sleep states in amniotes. *PLoS Biol.* 16:e2005982. doi: 10.1371/journal.pbio.2005982
- Lisman, J., and Sternberg, E. J. (2013). Habit and nonhabit systems for unconscious and conscious behavior: implications for multitasking. *J. Cogn. Neurosci.* 25, 273–283. doi: 10.1162/jocn\_a\_00319
- Logothetis, N. K., and Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science* 245, 761–763. doi: 10.1126/science.2772635
- Marcel, A. J. (1983). Conscious and unconscious perception: an approach to the relations between phenomenal experience and perceptual processes. *Cogn. Psychol.* 15, 238–300. doi: 10.1016/0010-0285(83)90010-5
- Massimini, M., and Tononi, G. (2018). *Sizing up Consciousness: Towards An Objective Measure of the Capacity for Experience*. Oxford: Oxford University Press.
- McKenzie, T. L. B., Bird, L. R., and Roberts, W. A. (2005). The effects of cache modification on food caching and retrieval behavior by rats. *Learn. Motiv.* 36, 260–278. doi: 10.1016/j.lmot.2005.02.011
- Meijer, G. T., Pie, J. L., Dolman, T. L., Pennartz, C. M. A., and Lansink, C. S. (2018). Audiovisual integration enhances stimulus detection performance in mice. *Front. Behav. Neurosci.* 12:231. doi: 10.3389/fnbeh.2018.00231
- Merkel, B. (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav. Brain Sci.* 30, 63–81; discussion 134. doi: 10.1017/s0140525x07000891
- Mikulan, E., Hesse, E., Sedeno, L., Bekinschtein, T., Sigman, M., García, M. D. C., et al. (2018). Intracranial high-gamma connectivity distinguishes wakefulness from sleep. *Neuroimage* 169, 265–277. doi: 10.1016/j.neuroimage.2017.12.015
- Milner, A. D., and Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia* 46, 774–785. doi: 10.1016/j.neuropsychologia.2007.10.005
- Milner, B., Squire, L. R., and Kandel, E. R. (1998). Cognitive neuroscience and the study of memory. *Neuron* 20, 445–468. doi: 10.1016/S0896-6273(00)80987-3
- Miranda, M., Kent, B. A., Morici, J. F., Gallo, F., Saksida, L. M., Bussey, T. J., et al. (2018). NMDA receptors and BDNF are necessary for discrimination of overlapping spatial and non-spatial memories in perirhinal cortex and hippocampus. *Neurobiol. Learn. Mem.* 155, 337–343. doi: 10.1016/j.nlm.2018.08.019
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature* 518, 529–533. doi: 10.1038/nature14236
- Monti, M. M., Coleman, M. R., and Owen, A. M. (2009). Executive functions in the absence of behavior: functional imaging of the minimally conscious state. *Prog. Brain Res.* 177, 249–260. doi: 10.1016/s0079-6123(09)17717-8
- Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., et al. (2010). Willful modulation of brain activity in disorders of consciousness. *N. Engl. J. Med.* 362, 579–589. doi: 10.1056/NEJMoa0905370
- Montijn, J. S., Goltstein, P. M., and Pennartz, C. M. (2015). Mouse V1 population correlates of visual detection rely on heterogeneity within neuronal response patterns. *Elife* 4:e10163. doi: 10.7554/elifelife.10163
- Mountcastle, V. B., LaMotte, R. H., and Carli, G. (1972). Detection thresholds for stimuli in humans and monkeys: comparison with threshold events in mechanoreceptive afferent nerve fibers innervating the monkey hand. *J. Neurophysiol.* 35, 122–136. doi: 10.1152/jn.1972.35.1.122
- Mulder, A. B., Nordquist, R. E., Orgut, O., and Pennartz, C. M. (2003). Learning-related changes in response patterns of prefrontal neurons during instrumental conditioning. *Behav. Brain Res.* 146, 77–88. doi: 10.1016/j.bbr.2003.09.016
- Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol. Cybern.* 66, 241–251. doi: 10.1007/bf00198477
- Murakami, I., Kitaoka, A., and Ashida, H. (2006). A positive correlation between fixation instability and the strength of illusory motion in a static display. *Vis. Res.* 46, 2421–2431. doi: 10.1016/j.visres.2006.01.030
- Murphy, M. P., Saunders, A., Moreira, C., Rizzi, A. A., and Raibert, M. (2011). The LittleDog robot. *Int. J. Rob. Res.* 30, 145–149. doi: 10.1177/0278364910387457
- Nakazawa, K., Quirk, M. C., Chitwood, R. A., Watanabe, M., Yeckel, M. F., Sun, L. D., et al. (2002). Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science* 297, 211–218. doi: 10.1126/science.1071795
- Nieder, A. (2002). Seeing more than meets the eye: processing of illusory contours in animals. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 188, 249–260. doi: 10.1007/s00359-002-0306-x
- Nieder, A., and Wagner, H. (1999). Perception and neuronal coding of subjective contours in the owl. *Nat. Neurosci.* 2, 660–663. doi: 10.1038/10217
- Niu, Y. Q., Xiao, Q., Liu, R. F., Wu, L. Q., and Wang, S. R. (2006). Response characteristics of the pigeon's pretectal neurons to illusory contours and motion. *J. Physiol.* 577, 805–913. doi: 10.1113/jphysiol.2006.120071
- O'Hare, J. K., Ade, K. K., Sukharnikova, T., Van Hooser, S. D., Palmeri, M. L., Yin, H. H., et al. (2016). Pathway-specific striatal substrates for habitual behavior. *Neuron* 89, 472–479. doi: 10.1016/j.neuron.2015.12.032
- Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput. Biol.* 10:e1003588. doi: 10.1371/journal.pcbi.1003588
- O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* 34, 171–175. doi: 10.1016/0006-8993(71)90358-1
- Okuyama-Uchimura, F., and Komai, S. (2016). Mouse ability to perceive subjective contours. *Perception* 45, 315–327. doi: 10.1177/0301006615614440
- Olcese, U., Bos, J. J., Vinck, M., Lankelma, J. V., van Mourik-Donga, L. B., Schlumm, F., et al. (2016). Spike-based functional connectivity in cerebral cortex and hippocampus: loss of global connectivity is coupled to preservation of local connectivity during non-REM sleep. *J. Neurosci.* 36, 7676–7692. doi: 10.1523/JNEUROSCI.4201-15.2016
- Olcese, U., Bos, J. J., Vinck, M., and Pennartz, C. M. A. (2018a). Functional determinants of enhanced and depressed interareal information flow in nonrapid eye movement sleep between neuronal ensembles in rat cortex and hippocampus. *Sleep* 41:zsy167. doi: 10.1093/sleep/zsy167
- Olcese, U., Oude Lohuis, M. N., and Pennartz, C. M. A. (2018b). Sensory processing across conscious and nonconscious brain states: from single neurons to distributed networks for inferential representation. *Front. Syst. Neurosci.* 12:49. doi: 10.3389/fnsys.2018.00049
- Owen, A. M. (2013). Detecting consciousness: a unique role for neuroimaging. *Annu. Rev. Psychol.* 64, 109–133. doi: 10.1146/annurev-psych-113011-143729
- Owen, A. M. (2014). Disorders of consciousness: diagnostic accuracy of brain imaging in the vegetative state. *Nat. Rev. Neurol.* 10, 370–371. doi: 10.1038/nrneurol.2014.102
- Owen, A. M. (2015). Using functional magnetic resonance imaging and electroencephalography to detect consciousness after severe brain injury. *Handb. Clin. Neurol.* 127, 277–293. doi: 10.1016/b978-0-444-52892-6.00018-0
- Pal, D., Silverstein, B. H., Lee, H., and Mashour, G. A. (2016). Neural correlates of wakefulness, sleep, and general anesthesia: an experimental study in rat. *Anesthesiology* 125, 929–942. doi: 10.1097/aln.0000000000001342
- Panagiotaropoulos, T. I., Kapoor, V., and Logothetis, N. K. (2014). Subjective visual perception: from local processing to emergent phenomena of brain activity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130534. doi: 10.1098/rstb.2013.0534
- Parker, A. J., Krug, K., and Cumming, B. G. (2002). Neuronal activity and its links with the perception of multi-stable figures. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, 1053–1062. doi: 10.1098/rstb.2002.1112

- Penfield, W. (1959). The interpretive cortex; the stream of consciousness in the human brain can be electrically reactivated. *Science* 129, 1719–1725. doi: 10.1126/science.129.3365.1719
- Pennartz, C. M. A. (2009). Identification and integration of sensory modalities: neural basis and relation to consciousness. *Conscious. Cogn.* 18, 718–739. doi: 10.1016/j.concog.2009.03.003
- Pennartz, C. M. A. (2015). *The Brain's Representational Power-On Consciousness and the Integration of Modalities*. Cambridge, MA: MIT Press.
- Pennartz, C. M. A. (2018). Consciousness, representation, action: the importance of being goal-directed. *Trends Cogn. Sci.* 22, 137–153. doi: 10.1016/j.tics.2017.10.006
- Pennartz, C. M., Ito, R., Verschure, P. F., Battaglia, F. P., and Robbins, T. W. (2011). The hippocampal-striatal axis in learning, prediction and goal-directed behavior. *Trends Neurosci.* 34, 548–559. doi: 10.1016/j.tins.2011.08.001
- Pennartz, C. M., Uylings, H. B., Barnes, C. A., and McNaughton, B. L. (2002). Memory reactivation and consolidation during sleep: from cellular mechanisms to human performance. *Prog. Brain Res.* 138, 143–166. doi: 10.1016/s0079-6123(02)38076-2
- Pepperberg, I. M. (2002). In search of king Solomon's ring: cognitive and communicative studies of Grey parrots (*Psittacus erithacus*). *Brain Behav. Evol.* 59, 54–67. doi: 10.1159/000063733
- Pepperberg, I. M., and Nakayama, K. (2016). Robust representation of shape in a Grey parrot (*Psittacus erithacus*). *Cognition* 153, 146–160. doi: 10.1016/j.cognition.2016.04.014
- Pepperberg, I. M., Willner, M. R., and Gravitz, L. B. (1997). Development of Piagetian object permanence in a grey parrot (*Psittacus erithacus*). *J. Comp. Psychol.* 111, 63–75. doi: 10.1037/0735-7036.111.1.63
- Persaud, N., McLeod, P., and Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nat. Neurosci.* 10, 257–261. doi: 10.1038/nn1840
- Peterson, A., Naci, L., Weijer, C., Cruse, D., Fernández-Espejo, D., Graham, M., et al. (2013). Assessing decision-making capacity in the behaviorally nonresponsive patient with residual covert awareness. *AJOB Neurosci.* 4, 3–14. doi: 10.1080/21507740.2013.821189
- Pezzulo, G., van der Meer, M. A., Lansink, C. S., and Pennartz, C. M. (2014). Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn. Sci.* 18, 647–657. doi: 10.1016/j.tics.2014.06.011
- Pitman, R. L., and Durban, J. W. (2012). Cooperative hunting behavior, prey selectivity and prey handling by pack ice killer whales (*Orcinus orca*), type B, in Antarctic Peninsula waters. *Mar. Mamm. Sci.* 28, 16–36. doi: 10.1111/j.1748-7692.2010.00453.x
- Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580
- Raposo, D., Sheppard, J. P., Schrater, P. R., and Churchland, A. K. (2012). Multisensory decision-making in rats and humans. *J. Neurosci.* 32, 3726–3735. doi: 10.1523/JNEUROSCI.4998-11.2012
- Redish, A. D. (2016). Vicarious trial and error. *Nat. Rev. Neurosci.* 17, 147–159. doi: 10.1038/nrn.2015.30
- Reggia, J. A. (2013). The rise of machine consciousness: studying consciousness with computational models. *Neural Netw.* 44, 112–131. doi: 10.1016/j.neunet.2013.03.011
- Ressler, N. (2004). Rewards and punishments, goal-directed behavior and consciousness. *Neurosci. Biobehav. Rev.* 28, 27–39. doi: 10.1016/j.neubiorev.2003.10.003
- Roberts, W. A. (2002). Are animals stuck in time? *Psychol. Bull.* 128, 473–489. doi: 10.1037/0033-2909.128.3.473
- Roberts, W. A., and Feeney, M. C. (2009). The comparative study of mental time travel. *Trends Cogn. Sci.* 13, 271–277. doi: 10.1016/j.tics.2009.03.003
- Salwiczek, L. H., Watanabe, A., and Clayton, N. S. (2010). Ten years of research into avian models of episodic-like memory and its implications for developmental and comparative cognition. *Behav. Brain Res.* 215, 221–234. doi: 10.1016/j.bbr.2010.06.011
- Samonds, J. M., Lieberman, S., and Priebe, N. J. (2018). Motion discrimination and the motion aftereffect in mouse vision. *eNeuro* 5:ENEURO.0065-18.2018. doi: 10.1523/eneuro.0065-18.2018
- Schiff, N. D. (2010). Recovery of consciousness after brain injury: a mesocircuit hypothesis. *Trends Neurosci.* 33, 1–9. doi: 10.1016/j.tins.2009.11.002
- Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003
- Schnakers, C., Vanhaudenhuyse, A., Giacino, J., Ventura, M., Boly, M., Majerus, S., et al. (2009). Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. *BMC Neurol.* 9:35. doi: 10.1186/1471-2377-9-35
- Searle, J. R. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT press.
- Searle, J. R. (2004). *Mind*. Oxford: Oxford University Press.
- Searle, J. R. (2015). *Seeing Things As They Are*. Oxford: Oxford University Press.
- Seghier, M. L., and Vuilleumier, P. (2006). Functional neuroimaging findings on the human perception of illusory contours. *Neurosci. Biobehav. Rev.* 30, 595–612. doi: 10.1016/j.neubiorev.2005.11.002
- Seth, A. K. (2008). Post-decision wagering measures metacognitive content, not sensory consciousness. *Conscious. Cogn.* 17, 981–983. doi: 10.1016/j.concog.2007.05.008
- Seth, A. K., Baars, B. J., and Edelman, D. B. (2005). Criteria for consciousness in humans and other mammals. *Conscious. Cogn.* 14, 119–139. doi: 10.1016/j.concog.2004.08.006
- Sheinberg, D. L., and Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. U S A* 94, 3408–3413. doi: 10.1073/pnas.94.7.3408
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* 362, 1140–1144. doi: 10.1126/science.aar6404
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. (2017). Mastering the game of Go without human knowledge. *Nature* 550, 354–359. doi: 10.1038/nature24270
- Sitt, J. D., King, J. R., El Karoui, I., Rohaut, B., Faugeras, F., Gramfort, A., et al. (2014). Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *J. Clin. Neurosci.* 137, 2258–2270. doi: 10.1093/brain/awu141
- Spinozzi, G., De Lillo, C., and Castelli, S. (2004). Detection of grouped and ungrouped parts in visual patterns by tufted capuchin monkeys (*Cebus apella*) and humans (*Homo sapiens*). *J. Comp. Psychol.* 118, 297–308. doi: 10.1037/0735-7036.118.3.297
- Stein, E. J., and Glick, D. B. (2016). Advances in awareness monitoring technologies. *Curr. Opin. Anaesthesiol.* 29, 711–716. doi: 10.1097/aco.0000000000000387
- Steriade, M. (2006). Grouping of brain rhythms in corticothalamic systems. *Neuroscience* 137, 1087–1106. doi: 10.1016/j.neuroscience.2005.10.029
- Stoerig, P., and Cowey, A. (1997). Blindsight in man and monkey. *Brain* 120, 535–559. doi: 10.1093/brain/120.3.535
- Storm, J. F., Boly, M., Casali, A. G., Massimini, M., Olcese, U., Pennartz, C. M. A., et al. (2017). Consciousness regained: disentangling mechanisms, brain systems, and behavioral responses. *J. Neurosci.* 37, 10882–10893. doi: 10.1523/JNEUROSCI.1838-17.2017
- Suddendorf, T., and Corballis, M. C. (2010). Behavioural evidence for mental time travel in nonhuman animals. *Behav. Brain Res.* 215, 292–298. doi: 10.1016/j.bbr.2009.11.044
- Teasdale, G., Maas, A., Lecky, F., Manley, G., Stocchetti, N., and Murray, G. (2014). The Glasgow Coma Scale at 40 years: standing the test of time. *Lancet Neurol.* 13, 844–854. doi: 10.1016/s1474-4422(14)70120-6
- Teyler, T. J., and DiScenna, P. (1986). The hippocampal memory indexing theory. *Behav. Neurosci.* 100, 147–154. doi: 10.1037/0735-7044.100.2.147
- Tong, F., Nakayama, K., Vaughan, J. T., and Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21, 753–759. doi: 10.1016/s0896-6273(00)80592-9
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.* 5:42. doi: 10.1186/1471-2202-5-42
- Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461. doi: 10.1038/nrn.2016.44
- Tononi, G., and Koch, C. (2015). Consciousness: here, there and everywhere? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140167. doi: 10.1098/rstb.2014.0167
- Treserras, S., Boulouaou, K., Conchou, F., Simonetta-Moreau, M., Berry, I., Celsis, P., et al. (2009). Transition from rest to movement: brain correlates

- revealed by functional connectivity. *Neuroimage* 48, 207–216. doi: 10.1016/j.neuroimage.2009.06.016
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annu. Rev. Psychol.* 53, 1–25. doi: 10.1146/annurev.psych.53.100901.135114
- Uhlrich, D. J., Essock, E. A., and Lehmkuhle, S. (1981). Cross-species correspondence of spatial contrast sensitivity functions. *Behav. Brain Res.* 2, 291–299. doi: 10.1016/0166-4328(81)90013-9
- Uylings, H. B., Groenewegen, H. J., and Kolb, B. (2003). Do rats have a prefrontal cortex? *Behav. Brain Res.* 146, 3–17. doi: 10.1016/j.bbr.2003.09.028
- van der Willigen, R. F., Harmening, W. M., Vossen, S., and Wagner, H. (2010). Disparity sensitivity in man and owl: psychophysical evidence for equivalent perception of shape-from-stereo. *J. Vis.* 10, 10.1–10.11. doi: 10.1167/10.1.10
- van Duuren, E., Escamez, F. A., Joosten, R. N., Visser, R., Mulder, A. B., and Pennartz, C. M. (2007). Neural coding of reward magnitude in the orbitofrontal cortex of the rat during a five-odor olfactory discrimination task. *Learn. Mem.* 14, 446–456. doi: 10.1101/lm.546207
- van Wingerden, M., Vinck, M., Tijms, V., Ferreira, I. R., Jonker, A. J., and Pennartz, C. M. (2012). NMDA receptors control cue-outcome selectivity and plasticity of orbitofrontal firing patterns during associative stimulus-reward learning. *Neuron* 76, 813–825. doi: 10.1016/j.neuron.2012.09.039
- Vetter, G., Haynes, J. D., and Pfaff, S. (2000). Evidence for multistability in the visual perception of pigeons. *Vis. Res.* 40, 2177–2186. doi: 10.1016/S0042-6989(00)00070-5
- Veyrac, A., Allerborn, M., Gros, A., Michon, F., Raguet, L., Kenney, J., et al. (2015). Memory of occasional events in rats: individual episodic memory profiles, flexibility, and neural substrate. *J. Neurosci.* 35, 7575–7586. doi: 10.1523/JNEUROSCI.3941-14.2015
- von der Heydt, R. (2013). Neurophysiological constraints on models of illusory contours. *Cogn. Neurosci.* 4, 49–50. doi: 10.1080/17588928.2012.748024
- Wada, M., Takano, K., Ora, H., Ide, M., and Kansaku, K. (2016). The rubber tail illusion as evidence of body ownership in mice. *J. Neurosci.* 36, 11133–11137. doi: 10.1523/JNEUROSCI.3006-15.2016
- Watanabe, A., Grodzinski, U., and Clayton, N. S. (2014). Western scrub-jays allocate longer observation time to more valuable information. *Anim. Cogn.* 17, 859–867. doi: 10.1007/s10071-013-0719-7
- Weinberger, N. M., Gold, P. E., and Sternberg, D. B. (1984). Epinephrine enables Pavlovian fear conditioning under anesthesia. *Science* 223, 605–607. doi: 10.1126/science.6695173
- Weir, A. A., Chappell, J., and Kacelnik, A. (2002). Shaping of hooks in New Caledonian crows. *Science* 297:981. doi: 10.1126/science.1073433
- Weiskrantz, L. (1995). The problem of animal consciousness in relation to neuropsychology. *Behav. Brain Res.* 71, 171–175. doi: 10.1016/0166-4328(95)00042-9
- Weiskrantz, L. (1997). *Consciousness Lost and Found: A Neuropsychological Exploration*. Oxford: Oxford University Press.
- Whishaw, I. Q., Schallert, T., and Kolb, B. (1981). An analysis of feeding and sensorimotor abilities of rats after decortication. *J. Comp. Physiol. Psychol.* 95, 85–103. doi: 10.1037/h0077760
- Wikenheiser, A. M., and Redish, A. D. (2015). Hippocampal theta sequences reflect current goals. *Nat. Neurosci.* 18, 289–294. doi: 10.1038/nn.3909
- Wilson, M. A., and McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science* 261, 1055–1058. doi: 10.1126/science.8351520
- Yang, F. C., Jacobson, T. K., and Burwell, R. D. (2017). Single neuron activity and theta modulation in the posterior parietal cortex in a visuospatial attention task. *Hippocampus* 27, 263–273. doi: 10.1002/hipo.22691
- Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189. doi: 10.1111/j.1460-9568.2004.03095.x
- Zeki, S. (1993). *A Vision of the Brain*. London: Blackwell.
- Zentall, T. R., Clement, T. S., Bhatt, R. S., and Allen, J. (2001). Episodic-like memory in pigeons. *Psychon. Bull. Rev.* 8, 685–690. doi: 10.3758/bf03196204
- Zhang, H., Fell, J., and Axmacher, N. (2018). Electrophysiological mechanisms of human memory consolidation. *Nat. Commun.* 9:4103. doi: 10.1038/s41467-018-06553-y

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Pennartz, Farisco and Evers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Scaling of a Large-Scale Simulation of Synchronous Slow-Wave and Asynchronous Awake-Like Activity of a Cortical Model With Long-Range Interconnections

Elena Pastorelli<sup>1,2\*</sup>, Cristiano Capone<sup>1,3</sup>, Francesco Simula<sup>1</sup>, Maria V. Sanchez-Vives<sup>4,5</sup>, Paolo Del Giudice<sup>3</sup>, Maurizio Mattia<sup>3</sup> and Pier Stanislaw Paolucci<sup>1</sup>

<sup>1</sup> INFN, Sezione di Roma, Rome, Italy, <sup>2</sup> PhD Program in Behavioural Neuroscience, "Sapienza" University, Rome, Italy,

<sup>3</sup> National Center for Radiation Protection and Computational Physics, Istituto Superiore di Sanità, Rome, Italy, <sup>4</sup> Systems Neuroscience, IDIBAPS, Barcelona, Spain, <sup>5</sup> Department of Life and Medical Sciences, ICREA, Barcelona, Spain

## OPEN ACCESS

### Edited by:

Preston E. Garraghty,  
Indiana University Bloomington,  
United States

### Reviewed by:

Sacha Jennifer van Albada,  
Julich Research Centre, Germany  
Sergio E. Lew,  
University of Buenos Aires, Argentina

### \*Correspondence:

Elena Pastorelli  
elena.pastorelli@roma1.infn.it

**Received:** 20 March 2019

**Accepted:** 08 July 2019

**Published:** 23 July 2019

### Citation:

Pastorelli E, Capone C, Simula F, Sanchez-Vives MV, Del Giudice P, Mattia M and Paolucci PS (2019) Scaling of a Large-Scale Simulation of Synchronous Slow-Wave and Asynchronous Awake-Like Activity of a Cortical Model With Long-Range Interconnections. *Front. Syst. Neurosci.* 13:33. doi: 10.3389/fnsys.2019.00033

Cortical synapse organization supports a range of dynamic states on multiple spatial and temporal scales, from synchronous slow wave activity (SWA), characteristic of deep sleep or anesthesia, to fluctuating, asynchronous activity during wakefulness (AW). Such dynamic diversity poses a challenge for producing efficient large-scale simulations that embody realistic metaphors of short- and long-range synaptic connectivity. In fact, during SWA and AW different spatial extents of the cortical tissue are active in a given timespan and at different firing rates, which implies a wide variety of loads of local computation and communication. A balanced evaluation of simulation performance and robustness should therefore include tests of a variety of cortical dynamic states. Here, we demonstrate performance scaling of our proprietary Distributed and Plastic Spiking Neural Networks (DPSNN) simulation engine in both SWA and AW for bidimensional grids of neural populations, which reflects the modular organization of the cortex. We explored networks up to  $192 \times 192$  modules, each composed of 1,250 integrate-and-fire neurons with spike-frequency adaptation, and exponentially decaying inter-modular synaptic connectivity with varying spatial decay constant. For the largest networks the total number of synapses was over 70 billion. The execution platform included up to 64 dual-socket nodes, each socket mounting 8 Intel Xeon Haswell processor cores @ 2.40 GHz clock rate. Network initialization time, memory usage, and execution time showed good scaling performances from 1 to 1,024 processes, implemented using the standard Message Passing Interface (MPI) protocol. We achieved simulation speeds of between  $2.3 \times 10^9$  and  $4.1 \times 10^9$  synaptic events per second for both cortical states in the explored range of inter-modular interconnections.

**Keywords:** spiking neural network, slow wave activity, asynchronous activity, long range interconnections, distributed simulation, strong scaling, weak scaling, large-scale simulation

# 1. INTRODUCTION

At the large scale, the neural dynamics of the cerebral cortex result from an interplay between local excitability and the pattern of synaptic connectivity. This interplay results in the propagation of neural activity. A case in point is the spontaneous onset and slow propagation of low-frequency activity waves during the deep stages of natural sleep or deep anesthesia (Hobson and Pace-Schott, 2002; Destexhe and Contreras, 2011; Sanchez-Vives and Mattia, 2014; Reyes-Puerta et al., 2016).

The brain in deep sleep expresses slow oscillations of activity at the single-neuron and local network levels which, at a macroscopic scale, appear to be synchronized in space and time as traveling waves (slow-wave activity, SWA). The “dynamic simplicity” of SWA is increasingly being recognized as an ideal test bed for refining and calibrating network models composed of spiking neurons. Understanding the dynamical and architectural determinants of SWA serves as an experimentally grounded starting point to tackle models of behaviorally relevant, awake states (Han et al., 2008; Curto et al., 2009; Luczak et al., 2009). A critical juncture in such a logical sequence is the description of the dynamic transition between SWA and asynchronous, irregular activity (AW, asynchronous wake state) as observed during fade-out of anesthesia, for instance, the mechanism of which is still a partially open problem (Curto et al., 2009; Steyn-Ross et al., 2013; Solovey et al., 2015). To help determine the mechanism of this transition, it may be of interest to identify the factors enabling the same nervous tissue to express global activity regimes as diverse as SWA and AW. Understanding this repertoire of global dynamics requires high-resolution numerical simulations of large-scale networks of neurons which, while keeping a manageable level of simplification, should be realistic with respect to both non-linear excitable local dynamics and to the spatial dependence of the synaptic connectivity (as well as the layered structure of the cortex) (Bazhenov et al., 2002; Hill and Tononi, 2005; Potjans and Diesmann, 2014; Krishnan et al., 2016).

Notably, efficient brain simulation is not only a scientific tool, but also a source of requirements and architectural inspiration for future parallel/distributed computing architectures, as well as a coding challenge on existing platforms. Neural network simulation engine projects have focused on: flexibility and user friendliness, biological plausibility, speed and scalability [e.g., NEST (Gewaltig and Diesmann, 2007; Jordan et al., 2018), NEURON (Hines and Carnevale, 1997; Carnevale and Hines, 2006), GENESIS (Wilson et al., 1989), BRIAN (Goodman and Brette, 2009; Stimberg et al., 2014)]. Their target execution platforms can be either homogeneous or heterogeneous (e.g., GPGPU-accelerated) high-performance computing (HPC) systems, (Izhikevich and Edelman, 2008; Nageswaran et al., 2009; Modha et al., 2011), or neuromorphic platforms, for either research or application purposes [e.g., SpiNNaker (Furber et al., 2013), BrainScaleS (Schmitt et al., 2017), TrueNorth (Merolla et al., 2014)].

From a computational point of view, SWA and AW pose different challenges to simulation engines, and comparing the simulator performance in both situations is an important element

in assessing the general value of the choices made in the code design. During SWA, different and limited portions of the network are sequentially active, with a locally high rate of exchanged spikes, while the rest of the system is almost silent. On the other hand, during AW the whole network is homogeneously involved in lower rate asynchronous activity. In a distributed and parallel simulation framework, this raises the question of whether the computational load on each core and the inter-process communication traffic are limiting factors in either cases. We also need to consider that activity propagates for long distances across the modeled cortical patch, therefore the impact of spike delivery on the execution time depends on the chosen connectivity. Achieving a fast and flexible simulator, in the face of the above issues, is the purpose of our Distributed and Plastic Spiking Neural Networks (DPSNN) engine. Early versions of the simulator (Paolucci et al., 2013) originated from the need for a representative benchmark developed to support the hardware/software co-design of distributed and parallel neural simulators. DPSNN was then extended to incorporate the event-driven approach of Mattia and Del Giudice (2000), implementing a mixed time-driven and event-driven strategy similar to the one introduced in Morrison et al. (2005). Here, we report the performances of DPSNN in both slow-wave (SW) and AW states, for different sizes of the network and for different connectivity ranges. Specifically, we discuss network initialization time, memory usage and execution times, and their strong and weak scaling behavior.

# 2. MATERIALS AND METHODS

## 2.1. Description of the Distributed Simulator

DPSNN has been designed to be natively distributed and parallel, with full parallelism also exploited during the creation and initialization of the network. The full neural system is represented in DPSNN by a network of C++ processes equipped with an agnostic communication infrastructure, designed to be easily interfaced with both Message Passing Interface (MPI) and other (custom) software/hardware communication systems. Each C++ process simulates the activity of one or more clusters of neighboring neurons and their set of incoming synapses. Neural activity generates spikes with target synapses in other processes; the set of “axonal spikes” is the payload of the associated exchanged messages. Each axonal spike carries the identity of its producing neuron and its original time of emission [AER, Address Event Representation (Lazzaro et al., 1993)]. Axonal spikes are only sent to target processes where at least one target synapse exists.

The memory cost of point-like neuron simulation is dominated by the representation of recurrent synapses which, in the intended biologically plausible simulations, are numbered in thousands per neuron. When plasticity support is switched off, the local description of each synapse includes only the identity of the target and source neurons, the synaptic weight, the transmission delay from the pre- to the post-synaptic neuron,

and an additional optional identifier for possible different types of synapse (see **Table 1**).

When the synaptic plasticity support is switched on, each synapse takes note of its own previous activation time. For every post-synaptic neuron spike and synaptic event a Spike-timing-dependent plasticity (STDP) contribution is computed using double floating point precision. Individual STDP events contribute to synaptic Long Term Potentiation (LTP) and Depression (LTD) through a first order low pass filter operating at a longer timescale than that of neural dynamics. Intermediate computations are performed in double precision, while the status of the low-pass filter is stored in the synaptic representation using single floating point precision. In the present work, synaptic plasticity is kept off.

During the initialization phase only, each synapse is represented at both the processes storing the source and target neuron because, in the current implementation, each process sends to each member in the subset of processes to be connected a single message requesting the creation of all needed synaptic connections. This memory overhead could be reduced by splitting the generation of connection requests in group, at the price of a proportional increase in the number of communication messages needed by the initialization phase. Indeed this is the moment of peak memory usage. Afterwards, the initialization synapses are stored exclusively in the process-hosting target neurons. For each process, the list of all incoming synapses is maintained. The synaptic list is double-ordered: synapses with the same delay are grouped together and are further ordered by presynaptic neuron index as in Mattia and Del Giudice (2000). Incoming spikes are ordered according to the identity of the presynaptic neurons. The ordering of both the synaptic list and incoming spikes helps a faster execution of the demultiplexing stage that translates incoming spikes into synaptic events because it increases the probability of contiguous memory accesses while exploring the synaptic list.

Because of the high number of synapses per neuron the relative cost of storing the neuron data structure is relatively small. It mainly contains the parameters and status variables used to describe the dynamics of the single-compartment point-like neuron. Moreover, it also contains the queue of input spikes (both recurrent and external) that arrived during the previous simulation time step, with their associated post-synaptic current value and time of arrival. External spikes are generated

as a Poissonian train of synaptic inputs. In the neuron data structure, double-precision floating-point storage is adopted for the variables used in the calculation of the membrane potential dynamics, while all the other variables and constant parameters related to the neuron are stored using single precision. Also, we opted for double precision floating point computations, because of the dominance of the cost related to the transport and memory accesses while distributing neural spikes to synapses.

In DPSNN there is no memory structure associated with external synapses: for each neuron, for each incoming external spike the associated synaptic current is generated on the fly from a Gaussian distribution with assigned mean and variance (see section 2.2); therefore external spikes have a computational cost but a negligible memory cost. The described queuing system ensures that the full set of synaptic inputs, recurrent plus external, are processed using an event-driven approach.

### 2.1.1. Execution Flow: Overview of the Mixed Time- and Event-Driven Simulation Scheme

There are two phases in a DPSNN simulation: (1) the creation of the structure of the neural network and its initialization; (2) the simulation of the dynamics of neurons (and synapses, if plasticity is switched on). For the simulation phase we adopted a combined event-driven and time-driven approach partially inspired by Morrison et al. (2005). Synaptic events drive the neural dynamics, while the message passing of spikes among processes happens at regular time steps, which must be set shorter than the minimum synaptic delay to guarantee the correct causal relationship in the distributed simulation. The minimum axo-synaptic delay sets the communication time step of the simulation, in our case 1ms.

**Figure 1** describes the main blocks composing the execution flow and the event- or time-driven nature of each block. The simulation phase can be broken down into the following phases: (1) identification of the subset of neurons that spiked during the previous time step, and (when plasticity is switched on) computation of an event-driven STDP contribution; (2) spikes are sent to the cluster of neurons where target synapses exist (inter-process communication blocks in the figure); (3) the list of incoming spikes to each process is placed into the double-ordered synaptic list, waiting for a number of time steps that match the synaptic delays, at which point the corresponding target synapses are activated; (4) synapses inject their event into queues that are local to their post-synaptic neuron and compute the STDP plasticity contribution; (5) each neuron sorts the lists of input events produced by recurrent and external synapses; (6) for each event in the queue, the neuron integrates its own dynamic equations using an event-driven solver. Periodically, at a slower time step (1 s in the current implementation), all synapses modify their efficacies using the integrated plasticity signal described above. Later sections describe the individual stages and data representation in further detail.

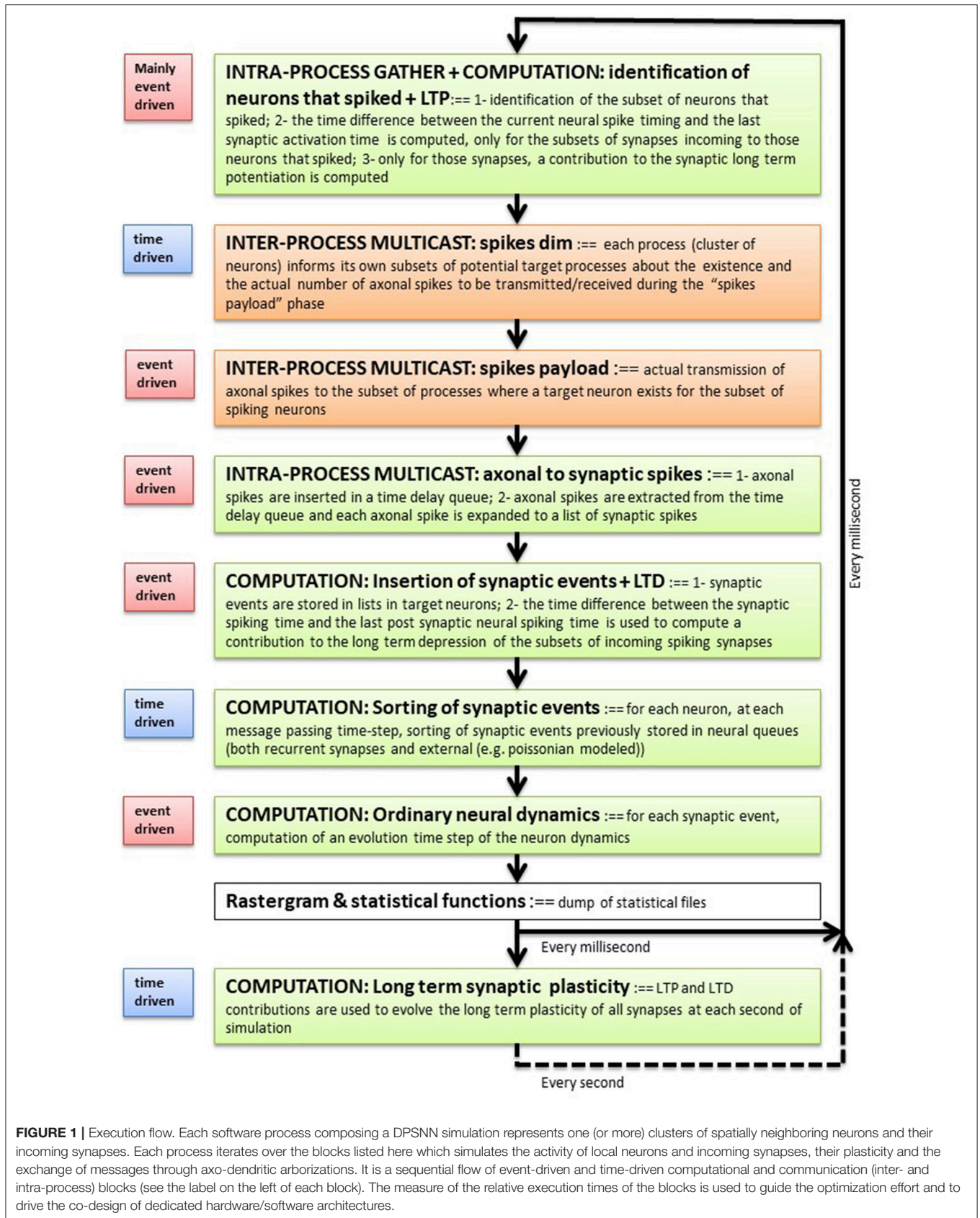
### 2.1.2. Spike Messages: Representation and Communication

Spike messages are defined according to an AER, where each spike is represented by the identity of the spiking neuron, and the spiking time. Spikes sharing a target process (i.e., targeting

**TABLE 1** | Representation of recurrent synapses. Static synapses cost 12 bytes per synapse. Plasticity support adds a cost of 8 bytes per synapse.

Synaptic representation							
Static (12 bytes/synapse)						Plasticity (8 bytes/synapse)	
Field	Source neuron ID	Target neuron ID	Weight	Delay	Kind	Last spiking time	Derivative
Size (byte)	4	4	2	1	1	4	4





**FIGURE 1 |** Execution flow. Each software process composing a DPSNN simulation represents one (or more) clusters of spatially neighboring neurons and their incoming synapses. Each process iterates over the blocks listed here which simulates the activity of local neurons and incoming synapses, their plasticity and the exchange of messages through axo-dendritic arborizations. It is a sequential flow of event-driven and time-driven computational and communication (inter- and intra-process) blocks (see the label on the left of each block). The measure of the relative execution times of the blocks is used to guide the optimization effort and to drive the co-design of dedicated hardware/software architectures.



the cluster of neurons managed by a process) are packed into “axonal spikes” messages that are delivered from the source process to its target processes during the communication phase. During execution, the “axon” arborization is managed by the target process in order to reduce the network communication load. Its construction during the initialization phase involves the following steps: in a “fan-out” step, each process notifies the processes which will host the target synapses of its own neurons. In the next “fan-in” step, each target process stores its own list of source neurons, and the associated double-ordered list of incoming synapses. For details about the initial construction of the connectivity infrastructure and the delivery of spiking messages see **Supplementary Material**.

### 2.1.3. Bidimensional Grids of Cortical Columns and Their Mapping Onto Processes

Neural networks are organized in this study as bidimensional grids of modules (mimicking cortical columns) as in Capone et al. (2019b). Each module is composed of 1,250 point-like neurons, and further organized into subpopulations. The set of cortical columns can be distributed over a set of processes (and processors): each process can either host a fraction of a column (e.g.,  $\frac{1}{2}$ ,  $\frac{1}{4}$ ), a whole single column, or several columns (up to 576 columns per process for the experiments reported in this paper). In this respect, our data distribution strategy aims to store in contiguous memories neighboring neurons and their incoming synapses. As usual in parallel processing, for a given problem size the performance tends to saturate and then to worsen, as it is mapped onto an excessively large number of processors, so that for each grid size an optimal partitioning onto processes is heuristically established, in terms of the performance measure defined in section 2.4.

Each process runs on a single processor; if hyperthreading is active, more than one process can be simultaneously executed on the same processor. For the measures in this paper, hyperthreading was deactivated, so that one hardware core always runs a single process. In DPSNN, distributing the simulation over a set of processes means that both the initialization and the run phases are distributed, allowing the scaling of both phases with the number of processes, as demonstrated by the measures reported in section 3.

## 2.2. Model Architecture and Network States

For each population in the modular network, specific parameters for the neural dynamics can be defined, as well as specific intra- and inter-columnar connectivity and synaptic efficacies. Connectivity among different populations can be modeled with specific laws based on distance-dependent probability, specific to each pair of source and target subpopulation. By suitably setting the available interconnections between different populations, cortical laminar structures can also be potentially modeled in the simulation engine.

The grids (see **Table 2**) are squares in a range of sizes ( $24 \times 24$ ,  $48 \times 48$ ,  $96 \times 96$ ,  $192 \times 192$ ). Each local module is always composed of  $K = 1,250$  neurons, further subdivided into subpopulations. We implemented a ratio of 4:1 between

**TABLE 2 |** Configurations used for the scaling measures of DPSNN.

Grid	Number of columns	Number of neurons	Number of synapses		MPI proc	
			Recurrent	Total	Min	Max
$24 \times 24$	576	0.72M	0.8G	1.1G	1	64
$48 \times 48$	2,304	2.88M	3.2G	4.4G	4	256
$96 \times 96$	9,216	11.52M	12.9G	17.6G	64	1,024
$192 \times 192$	36,864	46.08M	51.8G	70.3G	1024	1,024

excitatory and inhibitory neurons ( $K_I = 250$  neuron/module); in each module, excitatory neurons ( $E$ ) were divided into two subpopulations:  $K_F = 250$  (25%) strongly coupled “foreground” neurons ( $F$ ), having a leading role in the dynamics, and  $K_B = 750$  (75%) “background” neurons ( $B$ ) continuously firing at a relatively low rate. Populations on the grid are connected to each other through a spatial connectivity kernel. The probability of connection from excitatory neurons decreases exponentially with the inter-module distance  $d$ :

$$C_{ts\lambda}(d) = C_{ts\lambda}^0 \times \exp\left(\frac{-d}{\lambda}\right). \quad (1)$$

More specifically,  $d$  is the distance between the source ( $s = \{F, B\}$ ), and target ( $t = \{F, B, I\}$ ) module, and  $\lambda$  a characteristic spatial scale of connectivity decay.  $d$  and  $\lambda$  are expressed using inter-modular distance units (imd). For the simulations here described, the translation to physical units sets imd in the range of a few hundreds micrometers. Simulations are performed considering different  $\lambda$  values (0.4, 0.5, 0.6, 0.7) imd, but  $C_{ts\lambda}^0$  is set so as to generate the same mean number of projected synapses per neuron ( $M_t = 0.9 * K_t, t = \{F, B, I\}$ ) for all  $\lambda$  values. Connections originating from inhibitory neurons are local (within the same local module) and also in this case  $M_t = 0.9 * K_t, t = \{F, B, I\}$ . All neurons of the same type (excitatory/inhibitory) in a population share the same mean number of incoming synapses. The connectivity has open boundary conditions on the edges of the two-dimensional surface.

Synaptic efficacies are randomly chosen from a Gaussian distribution with mean  $J_{ts}$  and SD  $\Delta J_{ts}$ , chosen in different experiments so as to set the system in different working regimes and simulate different states. The procedure for the selection of the efficacies is based on a mean-field method described below. Each neuron also receives spikes coming from neurons belonging to virtual external populations, collectively modeled as a Poisson process with average spike frequency  $\nu_{ext}$  and synaptic efficacy  $J_{ext}$ . Excitatory neurons are point-like leaky integrate-and-fire (LIF) neurons with spike frequency adaptation (SFA) (Gigante et al., 2007; Capone et al., 2019b). SFA is modeled as an activity-dependent self-inhibition, described by the fatigue variable  $c(t)$ . The time evolution of the membrane potential  $V(t)$ , and  $c(t)$ , of

excitatory neurons between spikes is governed by:

$$\begin{cases} \dot{V} = -\frac{V-E}{\tau_m} - g_c \frac{c}{C_m} + \sum J_i \delta(t - t_i - \delta_i) + \sum J_{ext}^i \delta(t - t_i^{poiss}) \\ \dot{c} = -\frac{c}{\tau_c} + \alpha_c \sum_k \delta(t - t_{sp}^{(k)}) \end{cases} \quad (2)$$

$\tau_m$  is the membrane characteristic time,  $C_m$  the membrane capacitance, and  $E$  the resting potential. SFA is not considered for inhibitory neurons; that is, in (2), the second equation and the  $g_c \frac{c}{C_m}$  term in the first are absent. Incoming spikes, generated at times  $t_i$ , reached the target neuron with delay  $\delta_i$  and provoked instantaneous membrane potential changes of amplitude  $J_i$ . Alike, external stimuli produce a  $J_{ext}^i$  increment in the membrane potential, with  $t_i^{poiss}$  representing the spike times generated by a Poisson distribution of average  $v_{ext}$ . An output spike at time  $t_{sp}^{(k)}$  was triggered if the membrane potential exceeded a threshold  $V_\theta$ . On firing, the membrane potential was reset to  $V_r$  for a refractory period  $\tau_{arp}$ , whereas  $c$  was increased by the amount  $\alpha_c$ . Once the network connectivity and the neural dynamics have been defined, synaptic efficacies and external stimuli can be set to determine the dynamical states accessible to the system, by means of mean-field theory.

We elaborated a dynamical mean-field description for our simulations. We assume that the input received by the different neurons in a module are independent but are conditioned by the same (and possibly time-dependent) mean and variance of the input synaptic current (Brunel and Hakim, 1999; Mattia and Del Giudice, 2002).

The gain function  $\phi_i$  for this kind of neurons was firstly found by Ricciardi (1977) as the first-passage time (FPT) of the membrane potential to the firing threshold, for an integrate-and-fire neuron with stationary white noise input current, in the diffusion approximation.

Spike frequency adaptation is an important ingredient for the occurrence of slow oscillations in the mean field dynamics (Gigante et al., 2007; Capone et al., 2019b).

The mean-field dynamics for the average activity  $v_i, i = \{F, B, I\}$  is determined by the gain functions  $\phi_i$  as follows:

$$(F, B): \begin{cases} \dot{v}_i = \frac{\phi_i(\vec{v}, \vec{c}) - v_i}{\tau_E} \\ \dot{c}_i = -\frac{c_i}{\tau_c} + \alpha_c v_i \end{cases}; \quad (I): \begin{cases} \dot{v}_i = \frac{\phi_i(\vec{v}) - v_i}{\tau_I} \end{cases} \quad (3)$$

where  $\tau_E$  and  $\tau_I$  are phenomenological time constants. The interplay between the recurrent excitation embodied in the gain function, and the activity-dependent self-inhibition, is the main driver of the alternation between a high-activity (Up) state and a low-activity (Down) state (Mattia and Sanchez-Vives, 2012; Capone et al., 2019b).

In the mean-field description of the neural module, synaptic connectivity is chosen so as to match the total average synaptic input the neurons would receive inside the multi-modular network. The fixed points of the dynamics expressed by Equation (3) can be analyzed using standard techniques (Strogatz, 2018). The nullclines of the system (where  $\dot{v} = 0$  or  $\dot{c} = 0$ ) cross at fixed points that can be predicted to be either stable or unstable. In the

simulations described here, the strengths of recurrent synapses  $J_{t,s}$  connecting source population  $s$  and target population  $t$ , and of external synapses  $J_{t,ext}$ , is randomly chosen from a Gaussian distribution with mean  $J_{t,s}$  and variance  $\Delta J_{t,s} = 0.25 \times J_{t,s}$ .

We relied on mean-field analysis to identify neural and network parameters setting the network's modules in SW or AW dynamic regimes. **Figure 2A** shows an example of nullclines for the mean-field equations (3) of a system displaying SW. The black S-shaped line is the nullcline for the rate  $v$  while the red straight line is the one for the fatigue variable  $c$  (for details see Mattia and Sanchez-Vives, 2012; Capone et al., 2019b). The stable fixed point, at the intersection of the nullclines, has a low level of activity and is characterized by a small basin of attraction: the system can easily escape from it thanks to the noise, and it gets driven toward the upper branch of the  $v$  nullcline from which, due to fatigue, it is attracted back to the fixed point, thus generating an oscillation (see **Figure 2B**).

Network parameters can also be set in order to have an asynchronous state, mainly by setting a lower Foreground-to-Foreground (FF) synaptic efficacy, which generates a more linear  $v$  nullcline close to the fixed point (see **Figure 2C**). In this case the basin of attraction of the fixed point is larger, and oscillations do not occur, resulting in a stationary asynchronous state (see **Figure 2D**), in which neural activity fluctuates around the mean-field fixed point. **Table S1** reports the complete list of synaptic parameters for both SW and AW states.

## 2.3. Neural Activity Analysis

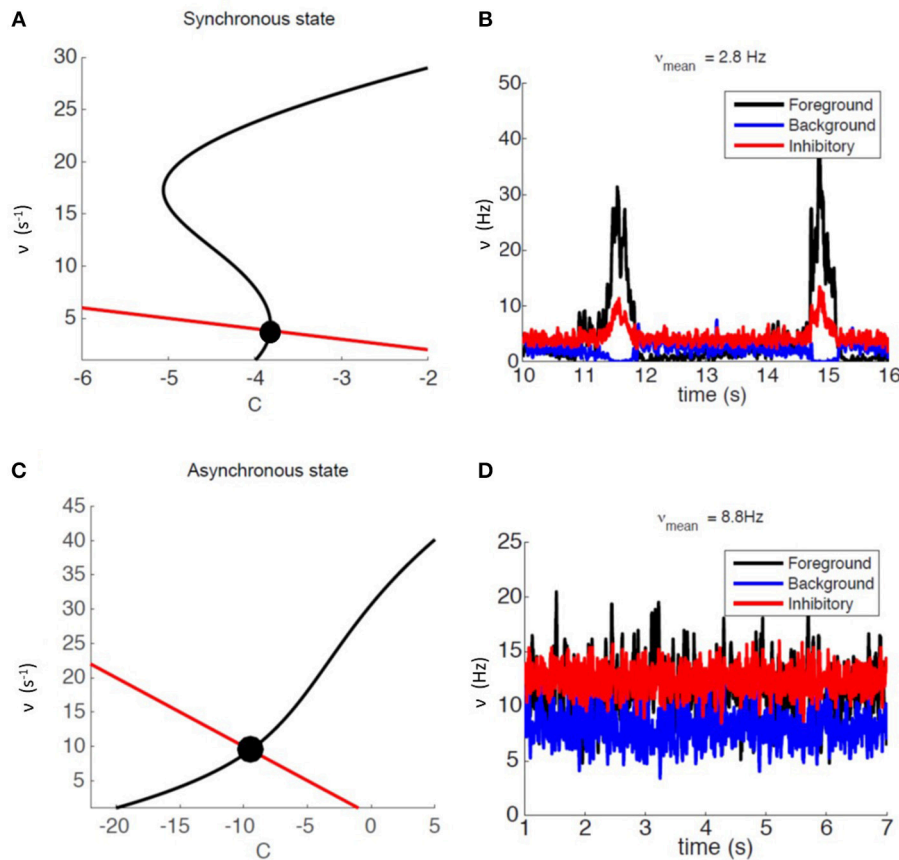
The simulation generates the spikes produced by each neuron in the network. From these, the time course of the average firing rate for each subpopulation in the network can be computed using an arbitrary sampling step, which here we set to be 5 ms. Power spectra are computed using the Welch method (see **Figures 6B,E** for examples of SW and AW power spectra including delta band).

A proxy to experimentally acquired multi-unit activity (MUA(t)) is computed as the average firing rate of a simulated module. In similarity with experimental data analysis (Capone et al., 2019b) we considered the logarithm of such signal ( $\log \frac{\text{MUA}(t)}{\text{MUA}_{down}}$ ) where the  $\text{MUA}_{down}$  is the average MUA(t) in the Down state. A zero mean white noise is added to emulate unspecific background fluctuations from neurons surrounding the module. The variance of such noise (0.5) was chosen to match the width of experimental  $\log \frac{\text{MUA}(t)}{\text{MUA}_{down}}$  distributions for bistable neural populations in their Down state.

For SW, the MUA(t) signal alternated between high and low activity states (Up and Down states).  $T(x, y)$ , the Down-to-Up transition time, is determined by a suitable threshold and is detected for each point in the spatial grid, defining the propagation wavefront.  $V(x, y)$ , the absolute value of wavefront speed, can be computed as

$$V(x, y) = \frac{1}{\sqrt{(\frac{\partial T(x, y)}{\partial x})^2 + (\frac{\partial T(x, y)}{\partial y})^2}} \quad (4)$$

The average speed during the collective Down-to-Up transitions is obtained by averaging  $V(x, y)$  over all the simulated positions  $(x, y)$ .



**FIGURE 2 |** Dynamical representation of the SW and AW states. **(A)** Phase space representation of mean-field analysis with a weakly stable fixed point. **(B)** Firing-rate time course for an example module, for foreground, background, and inhibitory sub-populations (respectively in black, blue, and red) in the SW state. **(C)** Phase space representation of mean-field analysis with a stable fixed point at a high level of activity. **(D)** Firing-rate time course for an example module, for foreground, background, and inhibitory sub-populations (respectively in black, blue, and red) in the AW state.

## 2.4. Performance Measures

Performance is quantified in terms of “strong scaling” and “weak scaling”: the former refers to keeping the problem size fixed and varying the number of hardware cores, while the latter refers to the increase in equal proportions of the problem size and the number of hardware cores.

As a performance measure we computed the equivalent synaptic events per second as the product of the total number of synapses (recurrent and external) and the number of spikes occurred across the whole simulation, divided by the elapsed execution time. This way, a comparison of the simulation cost among different problem sizes and hardware/software resources (core/processes) can be captured in a single graph. Similarly, we defined a convenient metric to evaluate the memory efficiency of a simulation: by dividing the total memory required by the simulation by the number of recurrent synapses to be represented. Indeed, as stated in section 2.1 we expect the memory usage to be dominated by the representation of synapses which are thousands per neuron.

Scalability measurements were taken on different neural network sizes, varying the size of the grid of columns and, for

each size, distributing it over a different span of MPI processes. We selected four grid sizes:  $24 \times 24$  columns, including 0.7M neurons and 1.1G synapses;  $48 \times 48$  columns including 2.8M neurons and 4.4G synapses;  $96 \times 96$  columns including 12M neurons and 17.6G synapses;  $192 \times 192$  columns including 46M neurons and 70G synapses. The number of processes over which each network size was distributed varied from a minimum, bounded by memory, and a maximum, bounded by communication or HPC platform constraints (see **Table 2**). For the  $192 \times 192$  configuration, only one measure was taken because of memory requirements, corresponding to a run distributed over 1,024 MPI processes/hardware cores.

Execution times for SW were measured across the time elapsed between the rising edges of two subsequent waves, and for AW across a time span of 3 s. In both cases, initial transients were omitted.

## 2.5. Validation of Results and Comparison of Performances

We used NEST version 2.12, the high-performance general purpose simulator developed by the NEST Initiative, as a

reference for the validation of results produced by the specialized DPSNN engine and for the comparison of its performances (speed, initialization time, memory footprint). The comparison was performed for both SW and AW dynamical states, for a network of 4.4G synapses ( $48 \times 48$  grid of columns). We assessed the correctness of results using power spectral density (PSD) analysis of the temporal evolution of the firing rates of each subpopulation. **Figure 3** reports an example of this comparison, showing the PSD match of NEST and DPSNN for each single subpopulation in a randomly chosen position inside the grid of columns. In AW state the average firing rate is about 8.8Hz with about 1125 synapses per neuron. Not surprisingly, NEST started to converge to the production of stable Power Spectral Densities only when the integration time step was reduced down to 0.1ms. We used this time step for all simulations used for comparison with DPSNN.

## 2.6. Execution Platforms

The benchmark measures assessing the DPSNN scalability in terms of run time, initialization time, and memory usage herein described, were performed on the Galileo server platform provided by the CINECA consortium. Galileo is a cluster of 516 IBM nodes, each of which includes 16-cores, distributed on two Intel Xeon Haswell E5-2630 v3 octa-core processors clocked at 2.40 GHz. The nodes are interconnected through an InfiniBand network. Due to the specific configuration of the server platform, the maximum-allowed partition of the server includes 64 nodes. Hyperthreading is disabled on all cores, therefore the number of MPI processes launched during each run exactly matches the hardware cores, with a maximum of 1,024 hardware cores (or equivalently MPI processes) available for any single run. In the present version of the DPSNN simulator, parameters cannot be changed at runtime, and dynamic equations are explicitly coded (i.e., no meta-description is available as an interface to the user).

Simulations used for validation and comparison between DPSNN and NEST were run on a smaller cluster of eight nodes interconnected through a Mellanox InfiniBand network; each node is a dual-socket server with a six-core Intel(R) Xeon(R) E5-2630 v2 CPU (clocked at 2.60 GHz) per socket with HyperThreading enabled, so that each core is able to run two processes.

## 3. RESULTS

### 3.1. Initialization Times

The initialization time, in seconds, is the time required to complete the setup phase of the simulator, which is necessary to build the whole neural network. Measures reported in **Figure 4A** show the scaling of the DPSNN initialization time for four network grid sizes, distributed over a growing number of hardware cores, which also correspond to MPI processes. Dashed colored lines represent strong scaling. Weak scaling is represented by the four points connected by the dotted black line (each point refers to a 4-fold increase in the network size and number of used cores with respect to the previous one). For the explored network sizes and hardware resources, the

initialization time speedup is almost ideal for fewer cores, while for the highest numbers of cores it is sub-ideal by a factor between 10 and 20%.

**Figure 4B** reports the scaling of the initialization time for two different values of the characteristic spatial scale of the decay of the connection probability ( $\lambda$ ), for a single grid size ( $96 \times 96$  columns). As already explained, for all simulations we kept the total number of synapses per neuron constant. Going from  $\lambda = 0.4$  to  $\lambda = 0.6$  imd, each excitatory neuron in a column has synaptic connections with neurons in a number of other columns growing from 44 to 78 (neglecting variations to proximity to columns' boundaries). Almost the same scaling is observed for both values of  $\lambda$ , with expected higher values for the initialization of a network with longer-range connectivity. Indeed, a dependence of the initialization time on synaptic connectivity range is expected, because it affects the proportion of target synapses residing in different columns, for which MPI messaging is needed for the connectivity setup, as explained in Materials and Methods.

### 3.2. Memory Occupation

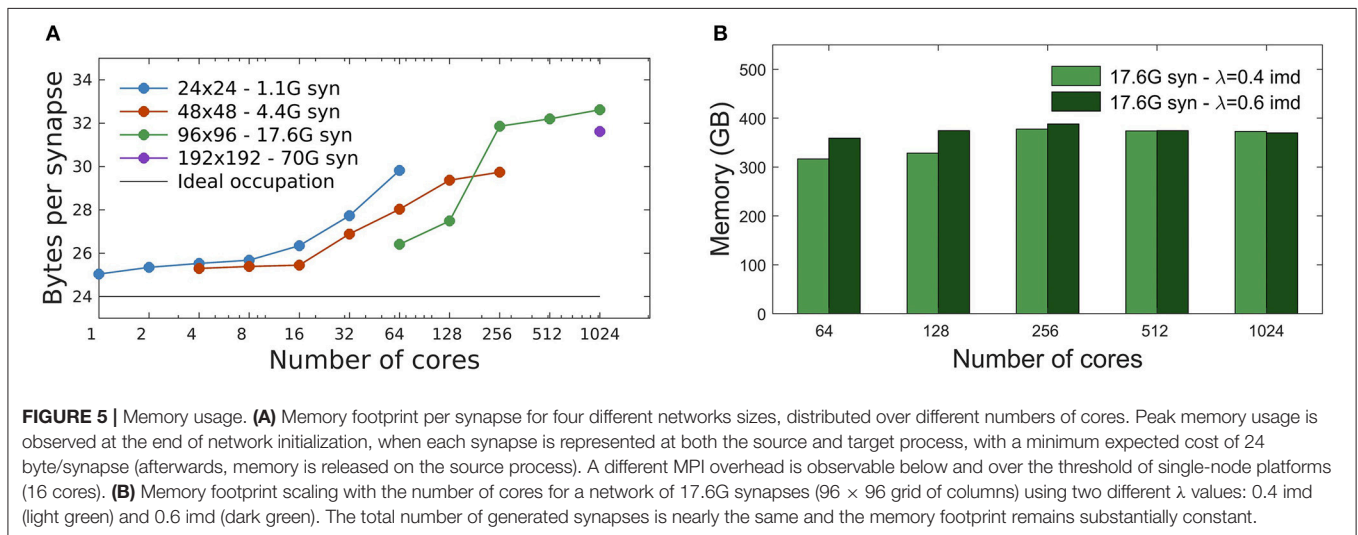
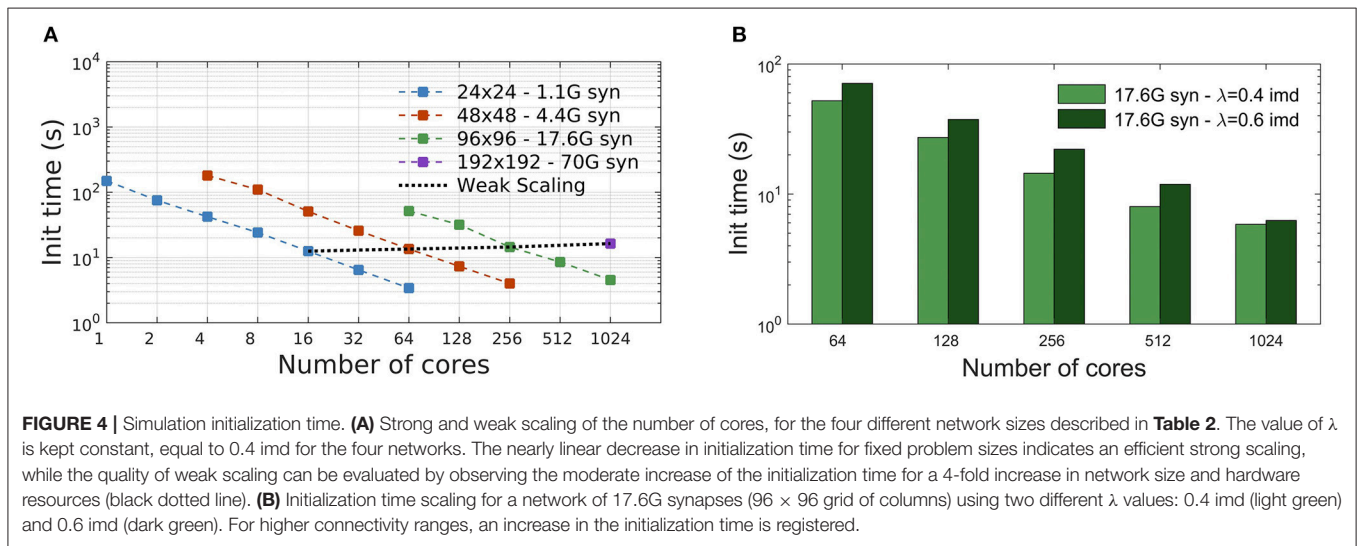
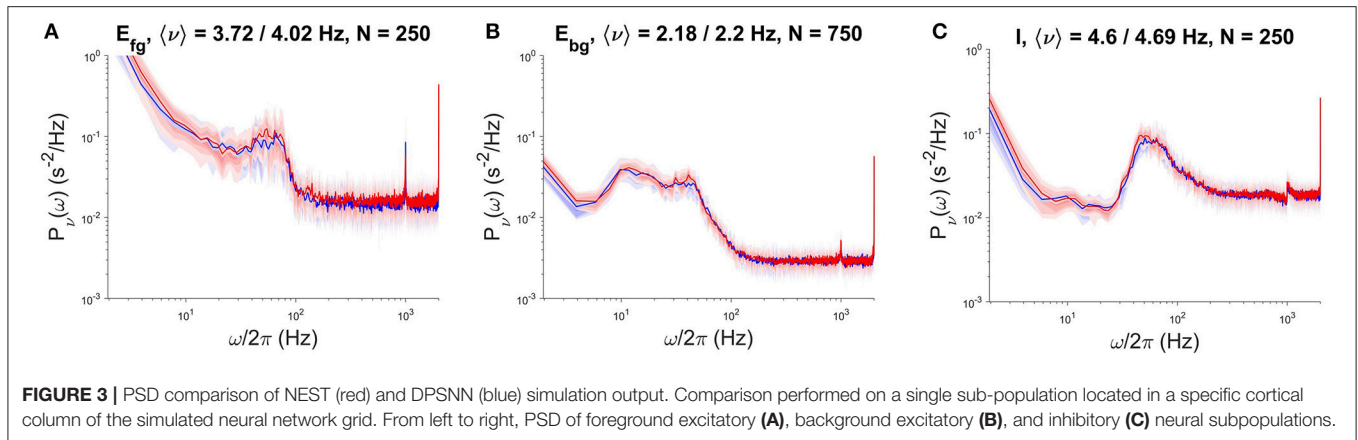
In DPSNN, we expect the memory usage to be dominated by the representation of synapses, which are thousands per neuron, with only a minor impact of the representation of neurons and external synapses. Therefore, we defined a memory consumption metric as the total required memory divided by the number of recurrent synapses (see section 2.4).

Each static synapse needs 12 bytes (see **Table 1**). As described in section 2.1, during network initialization synapses are generated by the process storing the source neuron. Subsequently, their values are communicated to the process containing the target neuron. The result is that, only during the initialization phase, each synapse is represented at both source and target processes. Therefore, we expect a minimum of 24 bytes/(recurrent synapse) to be allocated during initialization, which is the moment of peak memory usage.

The measured total memory consumption is between 25 and 32 bytes per recurrent synapse, for all simulations performed, from 1 to 1,024 MPI processes (**Figure 5**). From **Figure 5A** we observe a different trend of the memory cost vs. number of cores, below and above the threshold of single-node platforms (16 cores, see section 2.6). Beyond a single node, additional memory is mainly required by the buffers allocated by MPI interconnect libraries that adopt different strategies for communications over shared memory (inside a node) or among multiple nodes (**Figure 5A**). The memory occupation per synapse has been measured for different network sizes. **Figure 5A** shows that, for a given number of cores, the memory overhead typically decreases for increasing size of the network, as expected.

**Figure 5B** shows the impact of the spatial scale of connectivity decay on the memory footprint. Here the network has size  $96 \times 96$  columns (for a total of 17.6G synapses), and  $\lambda = 0.4, 0.6$  imd. As expected the memory footprint, mainly dependent on the total number of synapses, essentially remains constant.





Note that the relative differences for  $\lambda = 0.4, 0.6$  imd decrease as the number of cores increases; this is consistent with the fact that, given the network size, for larger core

numbers more columns get distributed on processes residing in different cores. This dilutes the effect of  $\lambda$ , which goes in the same direction.

### 3.3. Simulation Speed and Its Scaling

#### 3.3.1. Impact of Simulated State (SW or AW)

DPSNN execution times for SW and AW state simulations are reported in **Figure 6**. **Figures 6A,B** show example snapshots of the simulated network activity in the SW and AW states, respectively; **Figures 6C,D** show corresponding power spectra of the network activity, confirming the main features predicted by theory for such states (such as the low-frequency power increase for SW, the high-frequency asymptote proportional to the average firing rate, the spectral resonances related to SFA, and delays of the recurrent synaptic interaction). **Figures 6E,F** represent strong scaling for SW and AW: the wall-clock time required to simulate 1 s of activity, vs. number of processes, for four network sizes (see **Table 2**). In the same plots, weak scaling behavior is measured by joining points referring to a 4-fold increase in both the network size and the number of processes.

Simulation speeds are measured using the equivalent synaptic-events-per-second metric, defined in section 2.4. Departures from the ideal scaling behavior (linearly decreasing strong scaling plots, horizontal weak scaling plots) are globally captured in **Figure 7A**, for both SW and AW states and for all the problem sizes. For the simulations reported in this study, the simplified synaptic-events-per-second metric is a good approximation of a more complex metric separating recurrent and external events (Poisson noise), as demonstrated by **Figure 7B**: looking in more detail, the simulation of recurrent synaptic events is slower than that of external events that are locally generated by the routine that computes the dynamics of individual neurons. In our configurations there is about one order magnitude more recurrent synaptic event than external.

As numbers representative of the good scalability of the simulator, we remark that in the worst cases ( $1,024$  processes;  $96 \times 96$  for SW,  $192 \times 192$  for AW), the speedup with respect to the reference case ( $24 \times 24$  on a single process) is about 590 for AW and 460 for SW, compared with the ideal speedup of 1,024.

We also notice that performance scaling is slightly better for AW than SW, which is understood in terms of workload balance (due to the more homogeneous use of resources in AW); it is in any case remarkable that, although SW and AW dynamic states imply very different distributions of activity in space and time, the simulator provides comparable performance scaling figures.

Starting from the simulation speed expressed in terms of equivalent synaptic events per wall-clock second, DPSNN simulations presents a slow-down factor with respect to the real time, variable with the simulated network state and the available hw resources used for simulation. For example, considering a commodity cluster as the one used in this paper, able to allocate up to 64 processes, a network of size  $24 \times 24$  is about 12 times slower than real-time when simulating a SW state and about 23 times slower for the AW state. For larger simulations the slow-down factor increases, due to the not perfect scaling of the simulator performances. In this case, a  $96 \times 96$  network distributed on 1,024 processes presents a slow-down factor of about 23 for a SW simulation and of about 31 for the AW state.

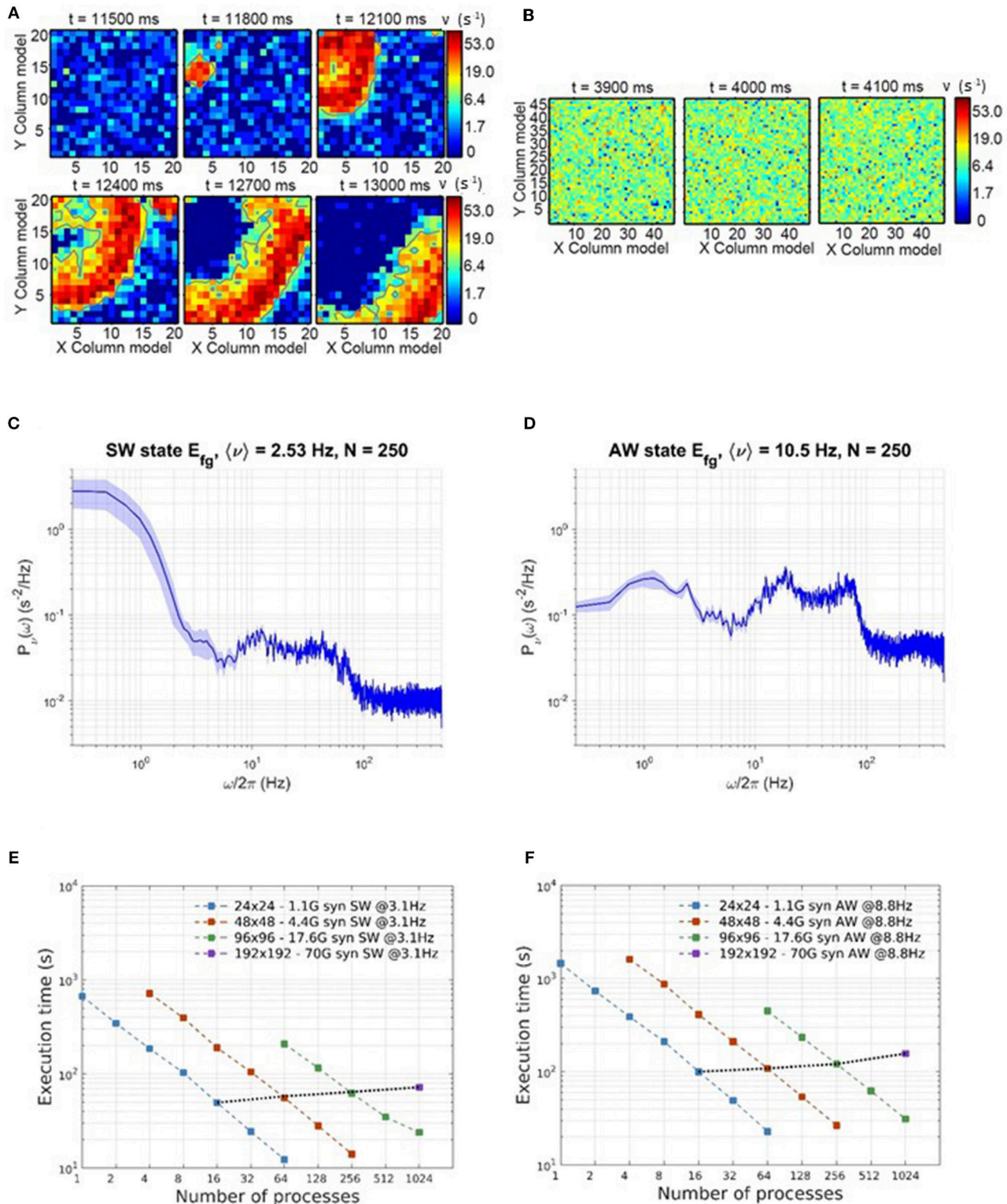
#### 3.3.2. Impact of Communication

Simula et al. (2019) studied the relative impact of computation and communication on the performance of DPSNN applied to simulations of AW states. For neural network sizes and number of processes in the range of those reported in this paper, the time spent in computation is still dominant, while communication grows to be the dominant factor when the number of neurons and synapses per process is reduced. In this paper, we evaluated the impact of communications on DPSNN performances both on SW and AW state simulations, using two different approaches. In the case of SW simulations, the analysis has been carried out with varying  $\lambda$  (therefore varying the ratio of local vs. remote excitatory synaptic connections, at a fixed total number of synapses per neuron):  $\lambda = 0.4$  imd (60% local connectivity),  $\lambda = 0.6$  imd (35% local connectivity); clearly, higher  $\lambda$  results in higher payload in communication between processes. Also, **Figure 8A** shows the known linear dependence of the slow-wave speed on  $\lambda$  (Coombes, 2005; Capone and Mattia, 2017). As the wave speed increases, the duration of the Up states stays approximately constant (not shown), so that an increasing portion of the network is simultaneously activated, which in turn may impact the simulation performance. In physical units, for an inter-modular distance (imd) of 0.4 mm,  $\lambda = 0.6$  imd implies a spatial decay scale of 0.24 mm, and the corresponding speed is  $\sim 15$  mm/s, which is in the range of biologically plausible values (Sanchez-Vives and McCormick, 2000; Ruiz-Mejias et al., 2011; Wester and Contreras, 2012; Stroh et al., 2013; Capone et al., 2019b).

**Figure 8B** shows that the impact of  $\lambda$  on SW simulations is almost negligible. **Figure 9** shows, instead, the impact of different mean firing rates on AW simulations. Higher firing rates imply higher payloads. DPSNN also demonstrates good scaling behavior in this case, with a slight performance increase for systems with a higher communication payload; this latter feature is due to communication costs being typically dominated by latencies and not bandwidth in spiking network simulations.

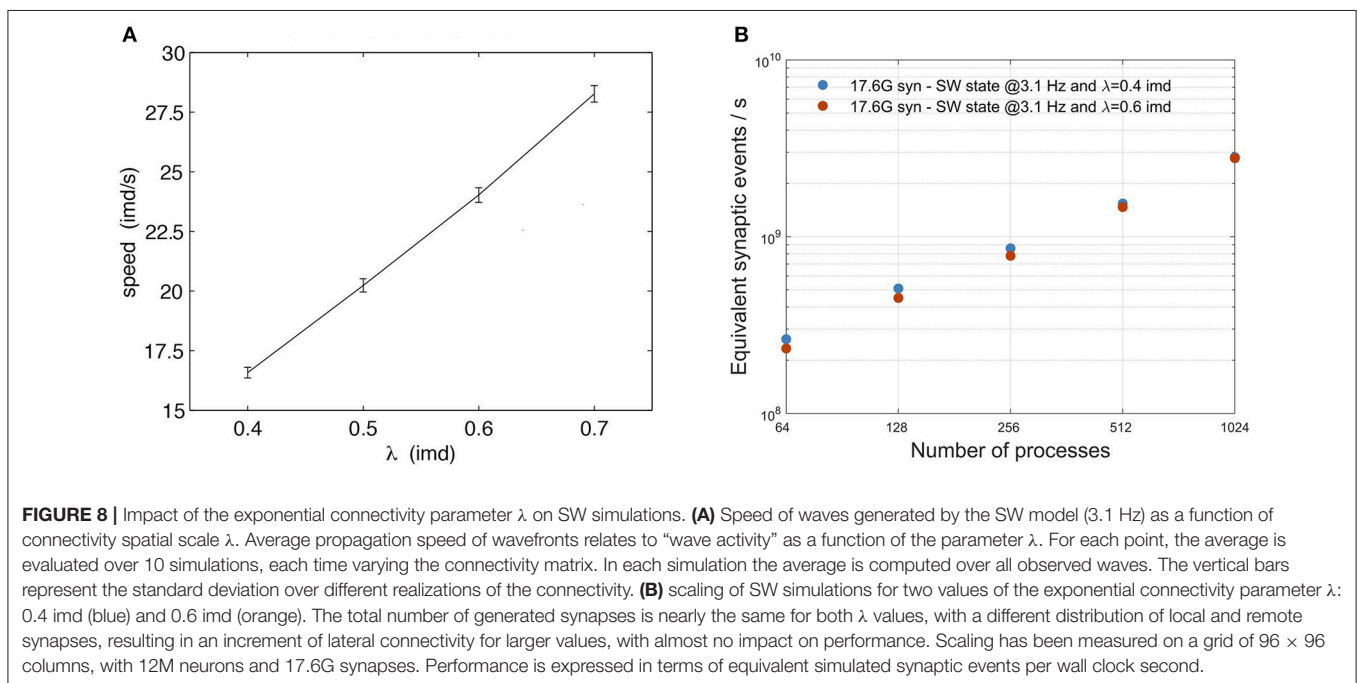
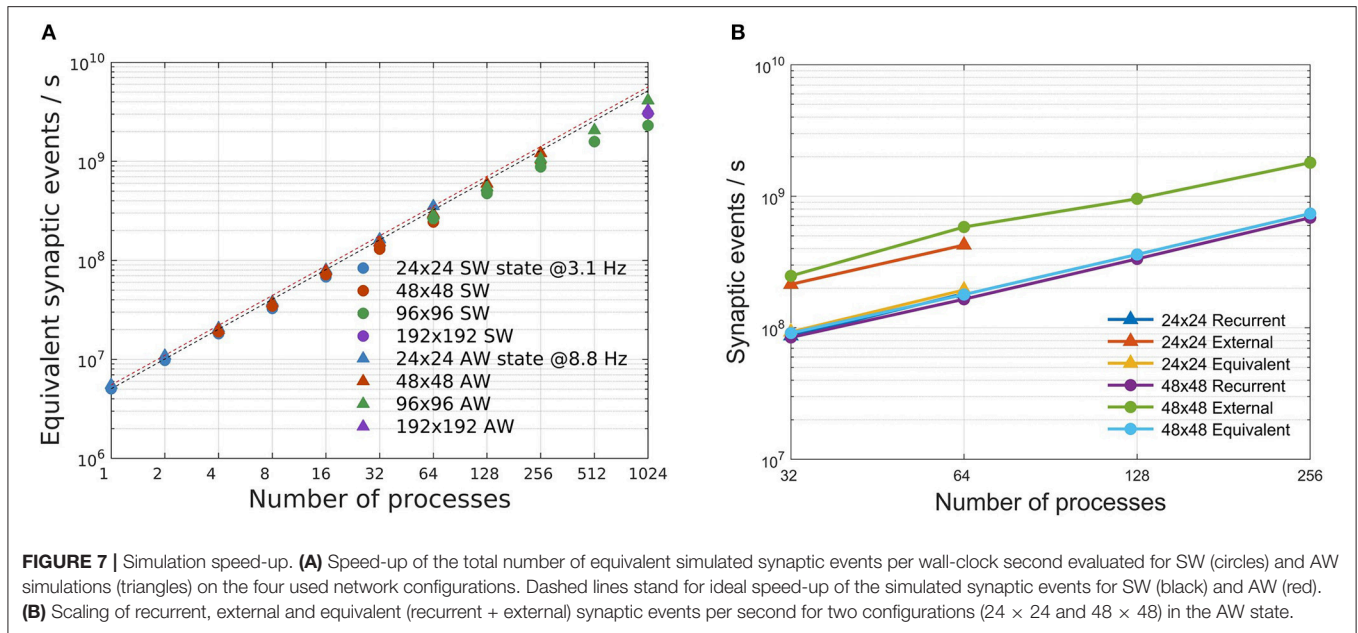
### 3.4. DPSNN and NEST: Comparison of Performances

**Table 3** reports a performance comparison between DPSNN and NEST using the configuration, described in section 2.5. DPSNN is about 130 times faster than NEST for SW simulations and about 80 times faster for AW cases, and its initialization is about 19 times faster. The memory footprint of DPSNN is about 2.5 times lower due to the decisions of representing the identities of presynaptic and post-synaptic neurons with only 4 bytes per neuron and the storage of weights using only 2 bytes per synapse (see **Table 1**). For the comparison of execution speed, we selected the best execution time for each simulator, for a fixed amount of used hardware resources; that is, the number of nodes of the cluster server. On the NEST simulator, we explored the space of all possible combination of MPI processes and number of threads during the AW simulations, in order to find the configuration performing better on a fixed number of hardware resources. In the same configuration, we also compared the initialization phase, the memory usage, and the SW execution times.



**FIGURE 6 |** Simulation time scaling and phenomenological behavior. **(A)** Time consecutive snapshots of the activity distribution in space, showing the wavefront propagation during a simulation expressing SW states. **(B)** Consecutive snapshots of the whole network activity in an asynchronous state. **(C,D)** Power spectra of network activity, respectively in SW **(C)** and AW **(D)** states. **(E,F)** Scaling of wall-clock execution time for 1 s of SW **(E)** and AW **(F)** simulated activity. In both SW and AW states, the scaling has been measured on four different network sizes, as in Table 2.





## 4. DISCUSSION

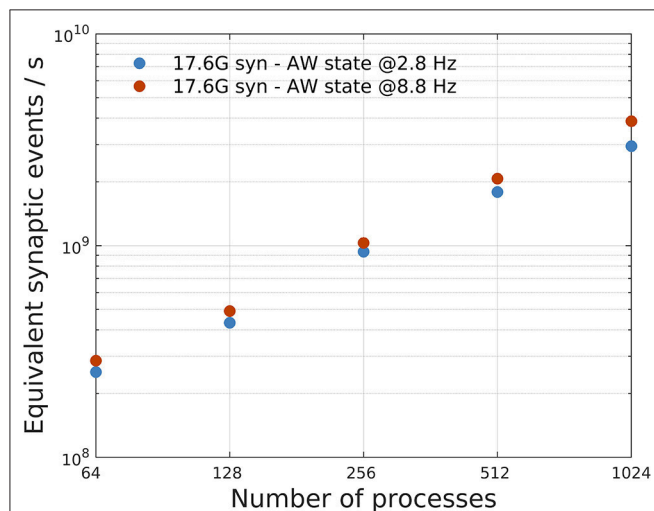
We presented a parallel distributed neural simulator, with emphasis on the robustness of its performance and scaling with respect to quite different collective dynamical regimes. This mixed time- and event-driven simulation engine (**Figure 1**) has been used to simulate large-scale networks including up to 46 million point-like spiking neurons interconnected by 70 billion instantaneous current synapses.

The development of DPSNN originated from the need for a simple, yet representative benchmark (i.e., a mini-application) developed to support the hardware/software co-design of distributed and parallel neural simulators. Early versions of DPSNN (Paolucci et al., 2013) have been used to drive the development of the EURETILE system (Paolucci et al., 2016) in which a custom parallelization environment and the APEnet hardware interconnect were tested. DPSNN was then extended to incorporate the event-driven approach



of Mattia and Del Giudice (2000), implementing a mixed time-driven and event-driven strategy inspired by Morrison et al. (2005). This simulator version is also currently included among the mini-applications driving the development of the interconnect system of the EXANEST ARM-based HPC architecture (Katevenis et al., 2016). In the framework of the Human Brain Project (<https://www.humanbrainproject.eu>), DPSNN is used to develop high-speed simulation of SW and AW states in multi-modular neural architectures. The modularity results from the organization of the network into densely connected modules mimicking the known modular structure of cortex. In this modeling framework, the inter-modular synaptic connectivity decays exponentially with the distance.

In this section, we first discuss the main strengths of the simulation engine and then put our work into perspective by comparing the utility and performances of such a specialized engine with those of a widely used general-purpose neural simulator (NEST).



**FIGURE 9 |** Impact of different mean firing rates on AW state simulations. A network with 17.6G synapses ( $96 \times 96$  grid of columns) has been simulated with two different values of mean firing rates, 2.8 Hz (blue) and 8.8 Hz (orange), corresponding to a different payload in communications. The plot shows a slight performance increase, in terms of equivalent simulated synaptic events per second, for systems with higher firing rate (and higher communication payload).

#### 4.1. Speed, Scaling, and Memory Footprint at Realistic Neural and Synaptic Densities

DPSNN is a high-speed simulator. The speed ranged from  $3 \times 10^9$  to  $4.1 \times 10^9$  equivalent synaptic events per wall-clock second, depending on the network state and the connectivity range, on commodity clusters including up to 1,024 hardware cores. As an order of magnitude, the simulation of a square cortical centimeter ( $\sim 27 \times 10^9$  synapses) at realistic neural and synaptic densities is about 30 times slower than real time on 1024 cores (Figure 7).

DPSNN is memory parsimonious: the memory required for the above square cortical centimeter is 837 GB (31 byte/synapses, including all library overheads), which is in the range of commodity clusters with few nodes. The total memory consumption ranges between 25 and 32 byte per recurrent synapse for the whole set of simulated neural networks and all configurations of the execution platform (from 1 to 1,024 MPI processes, Figure 5). The choice of representing with 4 bytes the identities of presynaptic and post-synaptic neurons limits the total number of neurons in the network to 2 billions. However, this is not yet a serious limitation for neural networks including thousands of synapses per neuron to be simulated on execution platforms including few hundreds of multi-core nodes. The size of neural ID representation will have to be enlarged for simulation of systems at the scale of human brains. Concerning synaptic weights, as already discussed, static synapses are stored with only two bytes of precision, but injection of current and neural dynamics is performed with double precision arithmetic. When plasticity is turned on, single precision floating-point storage of LTP and LTD contributions is adopted.

The engine has very low initialization times. DPSNN takes about 4 s to set up a network with  $\sim 17 \times 10^9$  synapses (Figure 4). We note that the initialization time is relevant, especially when many relatively short simulations are needed to explore a large parameter space.

Its performance is robust: good weak and strong scaling have been measured in the observed range of hardware resources and for all sizes of simulated cortical grids. The simulation speed was nearly independent from the mean firing rate (Figure 9), the range of connectivity (Figure 8), and from the cortical dynamic state (AW/SW) (Figure 7).

#### 4.2. Key Design Guidelines of the Simulation Engine

A few design guidelines contribute to the speed and scaling of the simulation engine. We kept as driving criteria the

**TABLE 3 |** Memory footprint, initialization, and execution times for 10 s of activity in SW and AW states required by DPSNN and NEST to simulate a neural network made of 2.8M neurons and 4.4G synapses ( $48 \times 48$  grid of cortical columns).

Execution platform		Init time (s)		SW Exec time (s)		AW Exec time (s)		Memory (GB)	
Nodes	Cores/Procs.	DPSNN	NEST	DPSNN	NEST	DPSNN	NEST	DPSNN	NEST
2	48	29.2	603.4	1,143.3	164,555.3	2,371.3	164,008.8	78.8	208.0
4	96	16.3	299.3	620.0	83,305.0	1,257.7	85,025.2	78.4	200.8
8	192	8.3	149.1	360.0	42,601.7	684.9	42,425.2	96.3	197.6

The second column reports the number of processes over which simulations are distributed. In case of DPSNN it corresponds to the number of pure MPI processes, while for NEST it is the number of Virtual Processes (VP). Each VP is calculated as MPI processes  $\times$  number of threads, where six threads are used for all NEST simulations.

goals of increasing the locality in memory accesses, the reduction of interprocess communication, an ordered traversal of lists and a reduction of backward searches and random accesses, the minimization of the number of layers in the calling stack, and a complete distribution of computation and storage among the cooperating processes with no need for centralized structures.

We stored in the memory of a process a set of spatially neighboring neurons, incoming synapses, and outgoing axons. For the kind of spatially organized neural networks described in this paper, this reduces the size of the payload and the required number of interprocess communications. Indeed, many spiking events will need to reach target neurons (and synapses) stored in the same process of the spiking neuron itself.

An explicit ordering in memory is adopted for outgoing and ingoing communication channels (representing the first branches in an axonal like arborizations). Explicit ordering in memory is also adopted for other data structures, like: the lists of incoming spiking events and recurrent synapses, the set of incoming event queues associated to different synaptic delays. Lists are implemented as arrays, without internal pointers. Also explicit ordering is used for the set of target neurons. Neurons are progressively numbered with lower bits in their identifiers associated to their local identity and higher bits conveying the id of the hosting process. The explicit ordering of neurons and synapses reduces the time spent during the sequence of memory accesses that will be followed during the simulation steps. In particular, lists are traveled only once per communication step, e.g., first searching for target synapses that are targets of spiking events and then through an ordered list of target neurons.

A third design criteria has been to keep the stack of nested function calls short, following the scheme described in **Figure 1**, and attempting to group at each level multiple calls to computational methods accessing the same memory structure. As an example, both the generation of external synaptic events (e.g., Poissonian stimulus) and the temporal reordering of recurrent and external synaptic events targeting a specific neuron, is deferred to the computation of the individual neural dynamics. The execution of this routine is supported by local queues storing all the synaptic currents targeting the individual neuron. This queue of events is accessed during a single call of the routine computing the dynamics of the neuron. In this case the data structure supports both locality in memory and in computation.

Similar design guidelines would be problematic to follow for general purpose simulation engines that are supposed to support maximum flexibility in the description and simulation of the data structures and of the dynamics of neurons, axonal arborizations and synapses. Moreover, higher abstraction requires separating the functionality of the simulation engine into independent modules. This would dictate a higher number of layers in the calling stack, more complex interfaces and data structures that hide details like their internal memory ordering.

### 4.3. Comparison With a State-of-the-Art User-Friendly Simulator and Motivations for Specialized Engines

There is a widely felt need for versatile, general-purpose neural simulators that offer a user-friendly interface for designing complex numerical experiments and provide the user with a wide set of models of proven scientific value. This boosted a number of initiatives (notably the NEST initiative, now central to the European Human Brain Project). However, such flexibility comes at a price. Performance-oriented engines, missing all the layers required for offering user generality and flexibility, contain the bare minimum code. In the case of DPSNN, this resulted in higher simulation speed, reduced memory footprint, and diminished initialization times (see section 3.4 and **Table 3**). In addition, optimization techniques developed for such engines on use-cases of proven scientific value can offer a template for future releases of general-purpose simulators. Indeed, this is what is happening in the current framework of cooperation with the NEST development team. Finally, engines stripped down to essential kernels constitute more easily manageable mini-application benchmarks for the hardware/software co-design of specialized simulation systems, because of easier profiling and simplified customizations on system software environments and hardware targets under development.

### 4.4. Future Works

The present implementation of DPSNN demonstrated to be efficient for homogeneous bidimensional grids of neural columns and for their mapping of up to 1,024 processes, and this facilitates a set of interesting scientific applications. However, further optimization could improve DPSNN performance, either in the perspective of moving simulations toward million-core exascale platforms or targeting real-time simulations at smaller scale (Simula et al., 2019), in particular addressing sleep-induced optimizations in cognitive tasks like classification (Capone et al., 2019a). For instance, we expect that the delivery of spiking messages will be a key element to be further optimized (e.g., using a hierarchical communication strategy). This will also be beneficial for an efficient support of white-matter long-range connectivity (brain connectomes) between multiple cortical areas.

A full exploitation of the model requires parameters tuned exploiting information provided by experimental data. Addressing this goal our team is improving tools for the analysis of both electrophysiological (De Bonis et al., 2019) and optical (Celotto et al., 2018) recordings (micro-ECoG and wide-field Calcium Imaging). The main aim is the spatio-temporal characterization of SWA. We also plan to apply inference methods to obtain refined maps of connectivity and excitability for insertion in the simulated model.

## DATA AVAILABILITY

The source code of the DPSNN engine and the data that support the findings of this study are openly available in GitHub at <https://github.com/APE-group/201903LargeScaleSimScaling>.

The DPSNN code also corresponds to the internal svn release 961 of the APE group repository.

## AUTHOR CONTRIBUTIONS

EP, PP, and FS improved the simulation engine, measured and analyzed the scaling. CC, MM, and PD provided the cortical model. MS-V provided the experimental data for calibration of simulations. All authors contributed to the final version of the manuscript and approved it for publication.

## FUNDING

This research has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreements No. 785907 (HBP SGA2), No. 720270 (HBP SGA1), and No. 671553 (EXANEST).

## REFERENCES

- Bazhenov, M., Timofeev, I., Steriade, M., and Sejnowski, T. J. (2002). Model of thalamocortical slow-wave sleep oscillations and transitions to activated states. *J. Neurosci.* 22, 8691–8704. doi: 10.1523/JNEUROSCI.22-19-08691.2002
- Brunel, N., and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.* 11, 1621–1671.
- Capone, C., and Mattia, M. (2017). Speed hysteresis and noise shaping of traveling fronts in neural fields: role of local circuitry and nonlocal connectivity. *Sci. Rep.* 7:39611. doi: 10.1038/srep39611
- Capone, C., Pastorelli, E., Golosio, B., and Paolucci, P. S. (2019a). Sleep-like slow oscillations improve visual classification through synaptic homeostasis and memory association in a thalamo-cortical model. *Sci. Rep.* 9. doi: 10.1038/s41598-019-45525-0
- Capone, C., Rebollo, B., Muñoz, A., Illa, X., Del Giudice, P., Sanchez-Vives, M. V., et al. (2019b). Slow waves in cortical slices: how spontaneous activity is shaped by laminar structure. *Cereb. Cortex* 29, 319–335. doi: 10.1093/cercor/bhx326
- Carnevale, N. T., and Hines, M. L. (2006). *The NEURON Book*. Cambridge, UK: Cambridge University Press.
- Celotto, M., De Luca, C., Muratore, P., Resta, F., Allegra Mascaro, A. L., Pavone, F. S., et al. (2018). Paolucci, analysis and model of cortical slow waves acquired with optical techniques. *arXiv:1811.11687*.
- Coombes, S. (2005). Waves, bumps, and patterns in neural field theories. *Biol. Cybern.* 93, 91–108. doi: 10.1007/s00422-005-0574-y
- Curto, C., Sakata, S., Marguet, S., Itskov, V., and Harris, K. D. (2009). A simple model of cortical dynamics explains variability and state dependence of sensory responses in urethane-anesthetized auditory cortex. *J. Neurosci.* 29, 10600–10612. doi: 10.1523/JNEUROSCI.2053-09.2009
- De Bonis, G., Dasilva, M., Pazienti, A., Sanchez-Vives, M. V., Mattia, M., and Paolucci, P. S. (2019). Slow waves analysis pipeline for extracting features of slow oscillations from the cerebral cortex of anesthetized mice. *arXiv:1902.08599*.
- Destexhe, A., and Contreras, D. (2011). “The fine structure of slow-wave sleep oscillations: from single neurons to large networks,” in *Sleep and Anesthesia*, Chapter 4, ed A. Hutt (New York, NY: Springer New York), 69–105.
- Furber, S. B., Lester, D. R., Plana, L. A., Garside, J. D., Painkras, E., Temple, S., et al. (2013). Overview of the SpiNNaker system architecture. *IEEE Trans. Comput.* 62, 2454–2467. doi: 10.1109/TC.2012.142
- Gewaltig, M.-O., and Diesmann, M. (2007). NEST (NEural Simulation Tool). *Scholarpedia* 2:1430. doi: 10.4249/scholarpedia.1430
- Gigante, G., Mattia, M., and Giudice, P. D. (2007). Diverse population-bursting modes of adapting spiking neurons. *Phys. Rev. Lett.* 98:148101. doi: 10.1103/PhysRevLett.98.148101

## ACKNOWLEDGMENTS

Large-scale simulations have been performed on the Galileo platform, provided by the CINECA in the frameworks of HBP SGA 1 and 2 and of the INFN-CINECA collaboration on the Computational Theoretical Physics Initiative. We acknowledge Hans Ekkehard Plesser and Dimitri Plotnikov for their support in setting up NEST simulations. We are grateful to the members of the INFN APE Parallel/Distributed Computing laboratory for their strenuous support. This manuscript has been released as a pre-print at Pastorelli et al. (2019).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnsys.2019.00033/full#supplementary-material>

- Goodman, D., and Brette, R. (2009). The brian simulator. *Front. Neurosci.* 3:26. doi: 10.3389/neuro.01.026.2009
- Han, F., Caporale, N., and Dan, Y. (2008). Reverberation of recent visual experience in spontaneous cortical waves. *Neuron* 60, 321–327. doi: 10.1016/j.neuron.2008.08.026
- Hill, S. L., and Tononi, G. (2005). Modeling sleep and wakefulness in the thalamocortical system. *J. Neurophysiol.* 93, 1671–1698. doi: 10.1152/jn.00915.2004
- Hines, M. L., and Carnevale, N. T. (1997). The neuron simulation environment. *Neural Comput.* 9, 1179–1209.
- Hobson, J. A., and Pace-Schott, E. F. (2002). The cognitive neuroscience of sleep: neuronal systems, consciousness and learning. *Nat. Rev. Neurosci.* 3, 679–693. doi: 10.1038/nrn915
- Izhikevich, E. M., and Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3593–3598. doi: 10.1073/pnas.0712231105
- Jordan, J., Ippen, T., Helias, M., Kitayama, I., Sato, M., Igarashi, J., et al. (2018). Extremely scalable spiking neuronal network simulation code: from laptops to exascale computers. *Front. Neuroinform.* 12:2. doi: 10.3389/fninf.2018.00002
- Katevenis, M., Chrysos, N., Marazakis, M., Mavroidis, I., Chaix, F., Kallimanis, N., et al. (2016). “The exanest project: interconnects, storage, and packaging for exascale systems,” in *2016 Euromicro Conference on Digital System Design (DSD)*, 60–67.
- Krishnan, G. P., Chauvette, S., Shamie, I., Soltani, S., Timofeev, I., Cash, S. S., et al. (2016). Cellular and neurochemical basis of sleep stages in the thalamocortical network. *eLife* 5, 1–29. doi: 10.7554/eLife.18607
- Lazzaro, J., Wawrzyniec, J., Mahowald, M., Sivillotti, M., and Gillespie, D. (1993). Silicon auditory processors as computer peripherals. *IEEE Trans. Neural Netw.* 4, 523–528.
- Luczak, A., Barthó, P., and Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* 62, 413–425. doi: 10.1016/j.neuron.2009.03.014
- Mattia, M., and Del Giudice, P. (2000). Efficient event-driven simulation of large networks of spiking neurons and dynamical synapses. *Neural Comput.* 12, 2305–2329. doi: 10.1162/089976600300014953
- Mattia, M., and Del Giudice, P. (2002). Population dynamics of interacting spiking neurons. *Phys. Rev. E* 66:051917. doi: 10.1103/PhysRevE.66.051917
- Mattia, M., and Sanchez-Vives, M. V. (2012). Exploring the spectrum of dynamical regimes and timescales in spontaneous cortical activity. *Cognit. Neurodyn.* 6, 239–250. doi: 10.1007/s11571-011-9179-4
- Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., et al. (2014). A million spiking-neuron integrated circuit with

- a scalable communication network and interface. *Science* 345, 668–673. doi: 10.1126/science.1254642
- Modha, D. S., Ananthanarayanan, R., Esser, S. K., Ndirango, A., Sherbondy, A. J., and Singh, R. (2011). Cognitive computing. *Commun. ACM* 54, 62–71. doi: 10.1145/1978542.1978559
- Morrison, A., Mehring, C., Geisel, T., Aertsen, A., and Diesmann, M. (2005). Advancing the boundaries of high-connectivity network simulation with distributed computing. *Neural Comput.* 17, 1776–1801. doi: 10.1162/0899766054026648
- Nageswaran, J. M., Dutt, N., Krichmar, J. L., Nicolau, A., and Veidenbaum, A. V. (2009). A configurable simulation environment for the efficient simulation of large-scale spiking neural networks on graphics processors. *Neural Netw.* 22, 791–800. doi: 10.1016/j.neunet.2009.06.028
- Paolucci, P. S., Ammendola, R., Biagioni, A., Frezza, O., Lo Cicero, F., Lonardo, A., et al. (2013). Distributed simulation of polychronous and plastic spiking neural networks: strong and weak scaling of a representative mini-application benchmark executed on a small-scale commodity cluster. *arXiv:1310.8478*.
- Paolucci, P. S., Biagioni, A., Murillo, L. G., Rousseau, F., Schor, L., Tosoratto, L., et al. (2016). Dynamic many-process applications on many-tile embedded systems and HPC clusters: the EURETILE programming environment and execution platforms. *J. Syst. Archit.* 69, 29–53. doi: 10.1016/j.sysarc.2015.11.008
- Pastorelli, E., Capone, C., Simula, F., Sanchez-Vives, M. V., Del Giudice, P., Mattia, M., et al. (2019). Scaling of a large-scale simulation of synchronous slow-wave and asynchronous awake-like activity of a cortical model with long-range interconnections. *arXiv:1902.08410*.
- Potjans, T. C., and Diesmann, M. (2014). The cell-type specific cortical microcircuit: Relating structure and activity in a full-scale spiking network model. *Cereb. Cortex* 24, 785–806. doi: 10.1093/cercor/bhs358
- Reyes-Puerta, V., Yang, J.-W., Siwek, M. E., Kilb, W., Sun, J.-J., and Luhmann, H. J. (2016). Propagation of spontaneous slow-wave activity across columns and layers of the adult rat barrel cortex *in vivo*. *Brain Struct. Funct.* 221, 4429–4449. doi: 10.1007/s00429-015-1173-x
- Ricciardi, L. M. (1977). *Diffusion Processes and Related Topics in Biology*. Berlin; Heidelberg; New York, NY: Springer-Verlag Berlin Heidelberg.
- Ruiz-Mejias, M., Ciria-Suarez, L., Mattia, M., and Sanchez-Vives, M. V. (2011). Slow and fast rhythms generated in the cerebral cortex of the anesthetized mouse. *J. Neurophysiol.* 106, 2910–2921. doi: 10.1152/jn.00440.2011
- Sanchez-Vives, M., and Mattia, M. (2014). Slow wave activity as the default mode of the cerebral cortex. *Arch. Ital. Biol.* 152, 147–155. doi: 10.12871/000298292014239
- Sanchez-Vives, M. V., and McCormick, D. A. (2000). Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nat. Neurosci.* 3:1027. doi: 10.1038/79848
- Schmitt, S., Klähn, J., Bellec, G., Grübl, A., Güttler, M., Harte, A., et al. (2017). “Neuromorphic hardware in the loop: training a deep spiking network on the BrainScaleS wafer-scale system,” in *2017 International Joint Conference on Neural Networks (IJCNN)* (Anchorage, AK: IEEE), 2227–2234. doi: 10.1109/IJCNN.2017.7966125
- Simula, F., Pastorelli, E., Paolucci, P. S., Martinelli, M., Lonardo, A., Biagioni, A., et al. (2019). “Real-time cortical simulations: energy and interconnect scaling on distributed systems,” in *2019 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)* (Pavia), 283–290.
- Solovey, G., Alonso, L. M., Yanagawa, T., Fujii, N., Magnasco, M. O., Cecchi, G. A., et al. (2015). Loss of consciousness is associated with stabilization of cortical activity. *J. Neurosci.* 35, 10866–10877. doi: 10.1523/JNEUROSCI.4895-14.2015
- Steyn-Ross, M. L., Steyn-Ross, D. A., and Sleight, J. W. (2013). Interacting Turing-Hopf instabilities drive symmetry-breaking transitions in a mean-field model of the cortex: a mechanism for the slow oscillation. *Phys. Rev. X* 3:21005. doi: 10.1103/PhysRevX.3.021005
- Stimberg, M., Goodman, D., Benichoux, V., and Brette, R. (2014). Equation-oriented specification of neural models for simulations. *Front. Neuroinform.* 8:6. doi: 10.3389/fninf.2014.00006
- Strogatz, S. H. (2018). *Nonlinear Dynamics and Chaos With Student Solutions Manual: With Applications to Physics, Biology, Chemistry, and Engineering*. Boca Raton, FL: CRC Press.
- Stroh, A., Adelsberger, H., Groh, A., Rühlmann, C., Fischer, S., Schierloh, A., et al. (2013). Making waves: initiation and propagation of corticothalamic  $\text{Ca}^{2+}$  waves *in vivo*. *Neuron* 77, 1136–1150. doi: 10.1016/j.neuron.2013.01.031
- Wester, J. C., and Contreras, D. (2012). Columnar interactions determine horizontal propagation of recurrent network activity in neocortex. *J. Neurosci.* 32, 5454–5471. doi: 10.1523/JNEUROSCI.5006-11.2012
- Wilson, M. A., Bhalla, U. S., Uhley, J. D., and Bower, J. M. (1989). “Genesis: a system for simulating neural networks,” in *Advances in Neural Information Processing Systems 1*, ed D. S. Touretzky (San Francisco, CA: Morgan-Kaufmann), 485–492.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Pastorelli, Capone, Simula, Sanchez-Vives, Del Giudice, Mattia and Paolucci. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# The Role of Top-Down Modulation in Shaping Sensory Processing Across Brain States: Implications for Consciousness

Tom Sikkens<sup>1,2</sup>, Conrado A. Bosman<sup>1,2</sup> and Umberto Olcese<sup>1,2\*</sup>

<sup>1</sup>Cognitive and Systems Neuroscience Group, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands, <sup>2</sup>Research Priority Area Brain and Cognition, University of Amsterdam, Amsterdam, Netherlands

## OPEN ACCESS

### Edited by:

Maria V. Sanchez-Vives,  
August Pi i Sunyer Biomedical  
Research Institute (IDIBAPS), Spain

### Reviewed by:

Roman Shusterman,  
University of Oregon, United States  
Paul Miller,  
Brandeis University, United States

### \*Correspondence:

Umberto Olcese  
u.olcese@uva.nl

**Received:** 01 February 2019

**Accepted:** 05 July 2019

**Published:** 24 July 2019

### Citation:

Sikkens T, Bosman CA and Olcese U  
(2019) The Role of Top-Down  
Modulation in Shaping Sensory  
Processing Across Brain States:  
Implications for Consciousness.  
*Front. Syst. Neurosci.* 13:31.  
doi: 10.3389/fnsys.2019.00031

Top-down, feedback projections account for a large portion of all connections between neurons in the thalamocortical system, yet their precise role remains the subject of much discussion. A large number of studies has focused on investigating how sensory information is transformed across hierarchically-distributed processing stages in a feedforward fashion, and computational models have shown that purely feedforward artificial neural networks can even outperform humans in pattern classification tasks. What is then the functional role of feedback connections? Several key roles have been identified, ranging from attentional modulation to, crucially, conscious perception. Specifically, most of the major theories on consciousness postulate that feedback connections would play an essential role in enabling sensory information to be consciously perceived. Consequently, it follows that their efficacy in modulating target regions should drastically decrease in nonconscious brain states [non-rapid eye movement (REM) sleep, anesthesia] compared to conscious ones (wakefulness), and also in instances when a given sensory stimulus is not perceived compared to when it is. Until recently, however, this prediction could only be tested with correlative experiments, due to the lack of techniques to selectively manipulate and measure the activity of feedback pathways. In this article, we will review the most recent literature on the functions of feedback connections across brain states and based on the presence or absence of perception. We will focus on experiments studying mismatch negativity, a phenomenon which has been hypothesized to rely on top-down modulation but which persists during nonconscious states. While feedback modulation is generally dampened in nonconscious states and enhanced when perception occurs, there are clear deviations from this rule. As we will discuss, this may pose a challenge to most theories of consciousness, and possibly require a change in how the level of consciousness in supposedly nonconscious states is assessed.

**Keywords:** brain states, sensory processing, top-down modulation, feedback projections, consciousness, mismatch negativity

## INTRODUCTION

Our brain is capable of sustaining all the functions necessary for life, from the most basic ones (breathing, autonomic regulation) to the most complex ones (language, social behavior). One function which stands out for being at the same time extremely complex to grasp, yet seamless in its presence, is consciousness. Even defining consciousness is extremely difficult, and has been a long-lasting subject of debate for philosophers and cognitive scientists (Dennett, 1991, 2018; Chalmers, 1995; Crick and Koch, 2003; Tononi and Koch, 2015; Dehaene et al., 2017; Storm et al., 2017; Lamme, 2018). In spite of this, in the last two decades significant progress has been made in the scientific study of consciousness, and in particular in the search of the neural correlates of consciousness (NCC, see **Table 1** for a list of abbreviations): the neural signature of brain processes underlying consciousness (Aru et al., 2012; Koch et al., 2016). For scientists to be able to measure the NCC, however, an operational definition of consciousness is necessary.

Chalmers (1995) famously proposed an ontology of cognitive phenomena associated with consciousness—the easy problems of consciousness: mechanisms controlling wakefulness and sleep, the ability to report mental states, the control of behavior, etc. These phenomena—termed easy because experiments to study them may be technically challenging but pose no conceptual difficulty—are inextricably associated with conscious processing, but are not consciousness. What really defines consciousness is the subjective experience that is inextricably associated with all the cognitive phenomena mentioned above. This has been defined by Chalmers as the hard problem of consciousness

because of the so-called explanatory gap (Levine, 1983) between neuron-level mechanisms and subjective experience.

The ontology first proposed by Chalmers has inspired the development of theories of consciousness aimed at addressing the nature of conscious experience, such as those proposed by Crick and Koch (1990, 2003) and Tononi and Edelman (1998a,b), and, more recently, the Integrated Information Theory (IIT; Tononi, 2004; Tononi et al., 2016). Nevertheless, Chalmers' proposal remains highly controversial and is firmly rejected by some philosophers, chief among them Daniel Dennett (Dennett, 1991, 2018; Cohen and Dennett, 2011). According to Dennett, there is no hard problem of consciousness, but rather consciousness can be understood by studying the functions associated with it. Dennett's approach can be seen as a theoretical foundation for another of the most influential theories of consciousness, the Global Neural Workspace (GNW) theory (Sergent and Dehaene, 2004; Baars, 2005; Dehaene et al., 2017). GNW is a theory of conscious access (Lamme, 2018; Naccache, 2018). What the theory aims to explain are the neuronal mechanisms which allow the brain to access (and subsequently report) information (Baars, 2002). This ultimately corresponds to an attempt to understand consciousness *via* addressing one of Chalmers' easy problems, which in Dennett's framework is all that is needed to uncover the mystery of consciousness.

The debate between these (and other) philosophical frameworks to explain consciousness, and between the different theoretical models to explain why the brain—and possibly artificial systems—is conscious is lively and at times heated, as attested by the vast recent literature on the topic (Boly et al., 2017; Odegaard et al., 2017; Dennett, 2018; Lamme, 2018; Naccache, 2018; Olcese, 2018; Olcese et al., 2018b). Nevertheless, in spite of the fundamental differences between the various frameworks to study consciousness, most theories aimed at explaining the neural mechanisms of consciousness agree on a key ingredient which makes our brain conscious: feedback connectivity. Feedforward processing consists in a progressive processing and transmission of sensory information from sensory organs up to higher-order cortical and motor regions (Lamme and Roelfsema, 2000; Lamme, 2018). While very complex forms of processing can be achieved by purely feedforward networks, as exemplified by the performance of deep artificial neural networks in the field of computer vision (LeCun et al., 2015), this is commonly agreed to occur non-consciously (Lamme, 2018). For example, both IIT and GNW concur that feedforward processing is *per se* not conscious. In IIT this can even be quantified, and purely feedforward networks (such as deep artificial neural networks) achieve a  $\Phi$  value of 0 (Oizumi et al., 2014), where  $\Phi$  quantifies the “level” of consciousness in IIT. Recurrent processing is, conversely, seen as a pre-requisite for consciousness by most if not all modern theories of consciousness. In IIT, as previously said, only systems with some level of integration (i.e., recursive connections) possess  $\Phi > 0$ , and therefore can be considered conscious (Oizumi et al., 2014). In GNW, the feedback flow of information from frontal to posterior brain regions gives rise to the so-called global ignition which is seen as essential for consciousness

**TABLE 1** | List of abbreviations.

Abbreviation	Definition
Cg1	Anterior Cingulate Cortex
DD	Deviance Detection
ECoG	Electrocorticogram
EEG	Electroencephalography
ERP	Evoked Response Potential
fMRI	Functional Magnetic Resonance Imaging
GABA	$\gamma$ -Aminobutyric Acid
GNW	Global Neural Workspace
IIT	Integrated Information Theory
L1–6	Cortical layers 1–6
LFP	Local Field Potential
MCS	Minimally-conscious state
MMN	Mismatch Negativity
MMNr	Mismatch Negativity response
N1	Negative peak in visual ERPs occurring 100 ms after stimulus onset and typically associated to the MMNr
NCC	Neural Correlates of Consciousness
P300	Typical ERP elicited within a decision-making process and during oddball paradigms, with a peak occurring about 300 ms after stimulus onset
P3b	Subcomponent of the P300 ERP which has been linked to aware processing
PC	Predictive Coding
REM	Rapid Eye Movement Sleep
SSA	Stimulus-Specific Adaptation
V1–V4	Visual cortical areas 1, 4
VS	Vegetative State

(Sergent and Dehaene, 2004; Baars, 2005). Similarly, other theories, such as the recurrent processing hypothesis (Lamme, 2006, 2018), and the Predictive Coding (PC) framework (Pennartz, 2015) indicate feedback processing as a key ingredient for consciousness.

The neuroscientific evidence underlying the link between recurrent processing and consciousness has been until recently primarily limited to correlational studies done at the mesoscopic level in human subjects, *via* techniques such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI; van Gaal et al., 2011; Fahrenfort et al., 2012). These techniques, albeit powerful, lack cell-level resolution. Thus, while experiments done using EEG and fMRI can provide evidence about a generalized increase or decrease in patterns of neural activity which are compatible with modulation in feedback/recurrent coupling, they cannot discriminate what individual neurons and neuronal populations do. As an example, EEG experiments have shown that, during brain states characterized by the loss of consciousness [Non-rapid eye movement (REM) sleep, anesthesia, coma], cortical effective connectivity drops markedly (Massimini et al., 2005, 2012; Ferrarelli et al., 2010; Casarotto et al., 2016). This was thought to reflect a generalized drop in the communication between cortical areas. Recently, however, studies performed at cellular resolution in rats showed that specific forms of long-range connectivity are preserved or even enhanced in Non-REM sleep compared to wakefulness (Olcese et al., 2016, 2018a). Therefore, human studies can only give limited insight into the role of recurrent connections in conscious processing, as only average, area-level dynamics can be assessed.

Our aim is to provide an overview of the existing evidence on the role of feedback processing as a key constituent of the NCC, as provided by studies with neuron-level resolution performed in animal models. In particular, we will focus on one of the easy problems of consciousness, as defined by Chalmers (1995): the difference between wakefulness and sleep/anesthesia. We will specifically address how top-down, intra-cortical feedback varies between states of consciousness (wakefulness vs. sleep and anesthesia) and thus investigate whether the presence or absence of this form of neural dynamics can be considered a valid NCC. This will allow us to dwell into the vast literature on cortical processing in animal models, and to focus on the neocortex, i.e., the brain region which is thought to be crucial for consciousness (Koch et al., 2016). By primarily centering on differences between conscious and non-conscious brain states, we will be able to address to what extent feedback communication decreases when consciousness fades. In the first section of this manuscript, we will discuss the circuit-level mechanisms of feedforward and feedback cortical communication, and what has been reported in terms of variation across behavioral states. Next, we will zoom in on studies investigating the genesis of mismatch responses. This strongly preserved phenomenon has been shown to occur during both conscious and non-conscious states and is composed of two different components: stimulus-specific adaptation (SSA) and deviance detection (DD; Näätänen et al., 2007; Garrido et al., 2009b; Hamm and Yuste, 2016; Harms et al., 2016). These two

components are, respectively, classically thought to represent feedforward and feedback forms of processing. The analysis of how these components vary across behavioral state will allow us to assess if and how feedback processing is disrupted during loss of consciousness.

## THE ROLE OF FEEDBACK PROCESSING ACROSS DIFFERENT BRAIN STATES

Brain state transitions can be described as the result of dynamic changes in functional long-range neuronal networks across time. Local and global connectivity greatly differ in their dynamics during the establishment of different behavioral states. While local connectivity is less affected by brain state transitions (Townsend et al., 2015; Olcese et al., 2016, 2018b; Siclari and Tononi, 2017), behavioral states exert great influences over the modulation of long-range connections (Destexhe et al., 1999; Massimini et al., 2005). In particular, a reduction in long-range connectivity has been associated with the loss in consciousness occurring in NREM sleep and anesthesia (Olcese et al., 2016; Storm et al., 2017). During wakefulness, local activity elicited within cortical microcircuits is usually broadcasted towards different brain areas through large-scale networks interconnected through feedforward and feedback projections, enabling sustained activity. This recurrent activity—necessarily supported by feedback, reentrant connections—is acknowledged to be at the basis of the NCC by several theories of consciousness (Koch et al., 2016; Lamme, 2018). In this section, we review the contribution of large-scale feedback networks in the emergence of conscious brain states, and how these networks influence the local activity of cortical microcircuits across different states.

In contrast with the predominant role recurrent connectivity plays in many of the aforementioned theories of consciousness, classical theories about sensory processing have mostly relied on the role of feedforward projections to explain how sensory stimulation can be processed (Hubel and Wiesel, 1977). Under this framework, simple stimulus features are encoded by neurons located in early sensory areas. Subsequently, this information is transferred *via* feedforward projections towards hierarchically higher areas. In these higher areas, the cumulative processing of incoming stimuli can be finally integrated into one coherent perception (Riesenhuber and Poggio, 1999). However, several cognitive processes, such as sensory perception, attention and goal-oriented behavior, have been shown to require feedback modulation—see e.g., Gazzaley and Nobre (2012), Gilbert and Li (2013), Manita et al. (2015) and Kwon et al. (2016). Furthermore, a pure feedforward approach falls short to explain how percepts—processed in a hierarchical feedforward fashion—can be accessible to conscious experience (Lamme and Roelfsema, 2000). This has been extensively investigated at the mesoscopic scale in humans (Fahrenfort et al., 2007, 2008, 2012), but the underlying circuit-level dynamics are less well understood. In a recent study investigating the neuronal responses in V1, V4, and prefrontal cortex of awake macaques during near-threshold stimulation, strong stimuli—those invariably leading to a conscious report—elicited a strong neuronal response progressing through all recorded

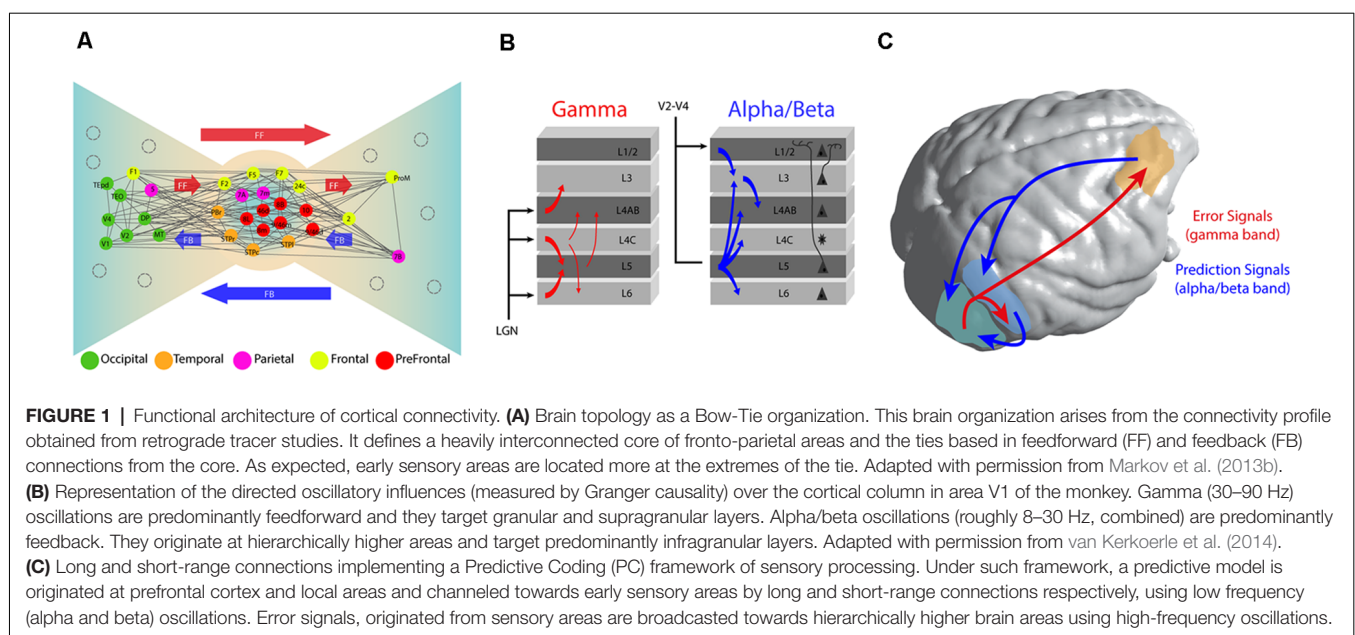
areas, in a way compatible with the notion of inter-areal feedforward propagation (van Vugt et al., 2018). However, in case of weaker, near-threshold stimuli, signal propagation across multiple areas was highly influenced by the global pre-stimulus brain state of the animal. In detail, a computational modeling approach was used by the authors of the study to show that recurrent connectivity between hierarchically-organized areas and feedback connectivity towards sensory cortices are necessary to be able to predict neuronal responses occurring in early processing stages when stimulus detection is reported (van Vugt et al., 2018). These findings strongly suggest that feedforward activity *per se*—albeit important for perceptual processing—is not sufficient to elicit a conscious experience of a perceived stimulus. Conversely, recurrent connectivity seems to be required to support conscious responses—see also Boly et al. (2011). Nevertheless, the precise contribution of this feedback connectivity to the establishment of these networks and the mechanisms underlying local cortical circuitry modulation remains to be elucidated.

## Anatomical Organization of Recurrent Neuronal Networks

To understand what precise feedback connectivity mechanisms are contributing to consciousness, it is important to revisit some anatomical principles that govern brain connectivity in general. Different anatomical studies have estimated that a vast majority of cortical connections are essentially local (Markov et al., 2014a), organized in evolutionarily conserved microcircuits (Bosman and Aboitiz, 2015). This local architecture represents a basic organizational unit for cortical computations (Douglas and Martin, 2004; Womelsdorf et al., 2014) and, as we discuss below, the way distinct local circuits communicate with each other is tightly linked to the emergence of different brain states.

Cortical microcircuits follow a columnar organization spread across all cortical layers. The prototypical microcircuit in sensory cortices features a central granular layer (L4), receiving projections from the thalamus and other cortical layers—but see Constantinople and Bruno (2013). Supragranular layers 2 and 3 (L2/3) receive presynaptic inputs from L4, while layer 1 (L1) displays cortico-cortical fibers connecting neighborhood columns. Infragranular layers include layers 5 and 6 (L5 and L6). Neurons in these layers receive local inputs from collateral projections of L2/3 neurons and provide feedback and feedforward connectivity to different thalamic structures (Hubel and Wiesel, 1977; Shepherd, 2011; Harris and Shepherd, 2015). Both feedforward and feedback projections have a strong preference for targeting areas that are close to their origin (Markov et al., 2014b). Supporting this highly redundant local connectivity, an estimated amount of 20% of the total connectivity is organized through long-range interareal connections (Markov et al., 2014a). These projections follow a well-defined pattern of connectivity (Felleman and Van Essen, 1991; Markov et al., 2014a,b). Feedforward projections preferentially originate in supragranular layers and target granular layers. Conversely, feedback connections arise from infragranular layers and avoid targeting granular layers (Felleman and Van Essen, 1991).

The quantification of this pattern of connectivity (Ercsey-Ravasz et al., 2013; Markov et al., 2014b) has suggested an important structural heterogeneity across brain areas, arranged in what Kennedy and colleagues have denominated a Bow-Tie organization (Markov et al., 2013b). Under such an organization, a subset of association areas including frontal, parietal, and temporal cortices appear to be heavily interconnected *via* recurrent connections, forming a central core of brain areas (Figure 1A). This highly interconnected core is characterized by a high prevalence of long-distance connections departing from





this core to other brain regions (Markov et al., 2013a). At the periphery, two different sets of regions are linked to the core through feedback and feedforward connections: a first group composed of primary and secondary sensory regions, and a second group consisting of premotor areas. These two groups are interconnected from the first to the second group *via* direct feedforward connections, and from the second to the first one *via* direct feedback connections (Ercsey-Ravasz et al., 2013).

This organizational principle encompasses several advantages. It reduces brain wiring and volumetric surface, yet increases the efficiency of information transfer and computational speed (Markov et al., 2013b). It reveals the importance of a fronto-parietal network, fundamentally characterized by a high prevalence of long-range connections, which some authors have suggested to be compatible with a GNW organization (Ercsey-Ravasz et al., 2013; Markov et al., 2013a), and provides an evolutionary advantage for long-range communication and cortical coordination of brain dynamics (Bosman and Aboitiz, 2015).

## Functional Dynamics of Feedback and Feedforward Connections: A Possible Role of Brain Oscillations

A bow-tie brain topology strongly relies on the functionality of long-range feedback connections to support sensory processing. As stated before, classical (and mainly feedforward) theories about perception are insufficient to explain sensory processing under such organizational principles. Conversely, the PC framework offers important insights into sensory perception in a view that is compatible with the notion of neuronal entrainment. PC is rooted in the tradition of inferential models of brain perception (Friston, 2010; Clark, 2013). Under this formulation, feedback projections transmit to hierarchically lower areas a generative model of sensory perception. In turn, feedforward projections transfer a signal error from the model, derived from the comparison between the existing model and the incoming sensory signals (Rao and Ballard, 1999; Bastos et al., 2012). Accordingly, recurrent (i.e., local) interactions between models and error signals at the level of cortical microcircuits are thought to improve the generation of statistical inferences about sensory perception. It has been argued that cortical microcircuit architectures can effectively implement PC computations. Under such scenario, local cortical microcircuits effectively integrate data and models originated from long-range feedforward and feedback networks, respectively (Bastos et al., 2012, 2015; Fontolan et al., 2014; Chao et al., 2018). Nevertheless, it is important to note that a PC framework can infer statistical regularities, but it cannot specify why data from different types of sensors would be consciously experienced (Pennartz, 2009). In other words, a PC framework helps to explain the importance of feedback connectivity but does not offer a solution to the hard problem of consciousness.

Experimentally, the study of neuronal dynamics underlying cortical computations across areas requires the utilization of techniques able to record several areas and spatial levels

simultaneously. From a myriad of emerging techniques (Adesnik and Naka, 2018), ensemble recordings and surface local field potentials (LFPs), recorded—respectively—by high-density laminar probes and electrocorticograms (ECoGs), are considered a primary choice due to their capacity to simultaneously record multiple local neuronal assemblies through cortical layers (Lewis et al., 2015; Pesaran et al., 2018). Laminar recordings in early visual areas have shown that LFP oscillations—a prominent feature of field recordings which has been associated with several perceptual and cognitive functions (Bosman et al., 2014; Fries, 2015)—are compartmentalized across cortical layers (**Figure 1B**). High-frequency oscillations (e.g., in the gamma frequency band, between 30–90 Hz) are observed mostly in supragranular layers. Conversely, lower frequency bands, such as alpha (8–12 Hz) and beta (13–30 Hz), are observed mostly in infragranular layers (von Stein et al., 2000; Buffalo et al., 2011; van Kerkoerle et al., 2014).

Brain rhythm compartmentalization appears to have a functional role during inter-areal communication (Fries, 2015). Recent studies enabling simultaneous recordings across multiple brain areas have demonstrated that directed long-range interactions can be exerted across different frequency band channels, a scenario compatible with the PC framework (Bastos et al., 2012, 2015; Fontolan et al., 2014; Michalareas et al., 2016; Chao et al., 2018). A study using high-density laminar profile in V1 of awake monkeys has shown that, while gamma band responses initiate in L4 and propagate through supra and infragranular layers of higher visual areas, alpha oscillations travel in the opposite direction (van Kerkoerle et al., 2014, **Figure 1B**). These relationships between frequencies and cortical layers are consistently observed through the cortical hierarchy. Using ECoGs in awake monkeys, Bastos et al. (2015) showed that, while feedforward directed influences (measured in terms of Granger causality) are observed through gamma oscillations across eight cortical regions recorded simultaneously, feedback influences across the visual hierarchy are consistently carried out by beta oscillations, and a similar functional connectivity pattern has been observed in humans (Michalareas et al., 2016). Furthermore, new studies have shown that feedback modulation can be updated continuously using error signals broadcasted from sensory areas to prefrontal cortex (Chao et al., 2018; **Figure 1C**). Neurons located in prefrontal cortex have the ability to integrate such error signals and continuously submit updated versions of the model through feedback projections to temporal areas (Chao et al., 2018). In early sensory cortices, the efficiency of feedforward frequency coupling—measured as an increment of gamma frequency band coherence between V1 and V4—increases after augmented top-down beta frequency band modulation (Richter et al., 2017). Importantly, optimized feedforward efficiency improves conscious access of perceived stimuli (Lamme and Roelfsema, 2000; Boly et al., 2011), and behavioral responses during attentional tasks (Rohenkohl et al., 2018), indicating that the extent of long-range feedback modulation correlates with conscious accessibility and enhanced behavioral performance.

Thus, simultaneous recordings across multiple areas using LFP rhythmic fluctuations have consistently shown that long-range feedback signals, possibly originating from

highly interconnected hubs comprising anterior regions of the brain, effectively modulate the activity of early sensory regions. This feedback modulation ultimately facilitates feedforward communication, behavioral performance and conscious reporting of perceived stimuli. A PC architecture, implemented across local-to-global anatomical connections and dynamical LFP oscillatory phase relationships, may support this organization (Olcese et al., 2018b).

However, while several studies have shown the importance of PC architectures and their oscillatory dynamics during stimulus processing, the exact contribution of these processes during different brain states still needs to be elucidated (Olcese et al., 2018b). The observation of local and global network dynamics across wide-brain areas across different behavioral states is a crucial step to understand what the mechanisms underlying these behavioral and network transitions are. Feedforward gamma connectivity globally decreases during deep sleep states, as compared to awake states, in intracranial recordings in epilepsy patients (Mikulan et al., 2018), but there are no consistent reports about low-frequency (and putatively feedback) phase relationships during sleep, in spite of the reduced role of feedback during non-conscious states (Boly et al., 2011). During anesthesia, a pharmacologically-induced brain state, quantitative EEG studies have shown that sedation with propofol is accompanied by a decreased posterior alpha and increased frontal/central beta power (Gugino et al., 2001; Akeju and Brown, 2017). This shift has been extensively studied and is thought to emerge from a disruption of prefrontal circuits created by a strong low-frequency thalamocortical synchronization (Vijayan et al., 2013; Flores et al., 2017). Yet, the detailed effects of anesthesia on cortico-cortical synchronization remain elusive. Finally, a promising line of research about the effects of feedback connectivity in patients with consciousness disorders has been developed in recent years. Feedback influences, but not feedforward ones, seem to be compromised in these patients (Boly et al., 2011), in agreement with the notion that feedback connections are important to sustain conscious activity. Intriguingly, a temporary recovery in the consciousness state of a subgroup of patients has been observed following the administration of the GABA modulator Zolpidem (Hall et al., 2010; Williams et al., 2013). Zolpidem administration reduces EEG power and coherence at 6–10 Hz frequencies (Williams et al., 2013). Surprisingly, however, Zolpidem seems to elevate beta frequency power at anterior regions following its administration. These effects seem to be mediated by thalamocortical interactions but is not clear whether direct cortico-cortical communication is also affected (Hall et al., 2010; Williams et al., 2013). Future studies might help to unveil the exact nature of these effects.

In conclusion, long-range feedback connections are an essential component of conscious brain states and determine how sensory information is processed at the local level (*via* both feedforward and recurrent mechanisms). Yet, the underlying circuit-level mechanisms remain poorly understood. Several studies have outlined the organizational

principles—both anatomical and functional—of feedback connections in the context of sensory processing. However, contrasting results have been presented on how feedforward and feedback functional connectivity vary as a function of the state of consciousness. Oscillatory dynamics during sensory processing may provide a framework to understand the role of feedback projections, but these phenomena involve a wide range of brain circuits, operating both at feedforward and feedback level and both at local and global scales. Consequently, one possible solution to understand the role of feedback projections and the underlying circuit mechanisms in conscious processing is to focus on processes that are purely feedback in nature, and which are thought to be exclusively dependent on long-range, top-down mechanisms.

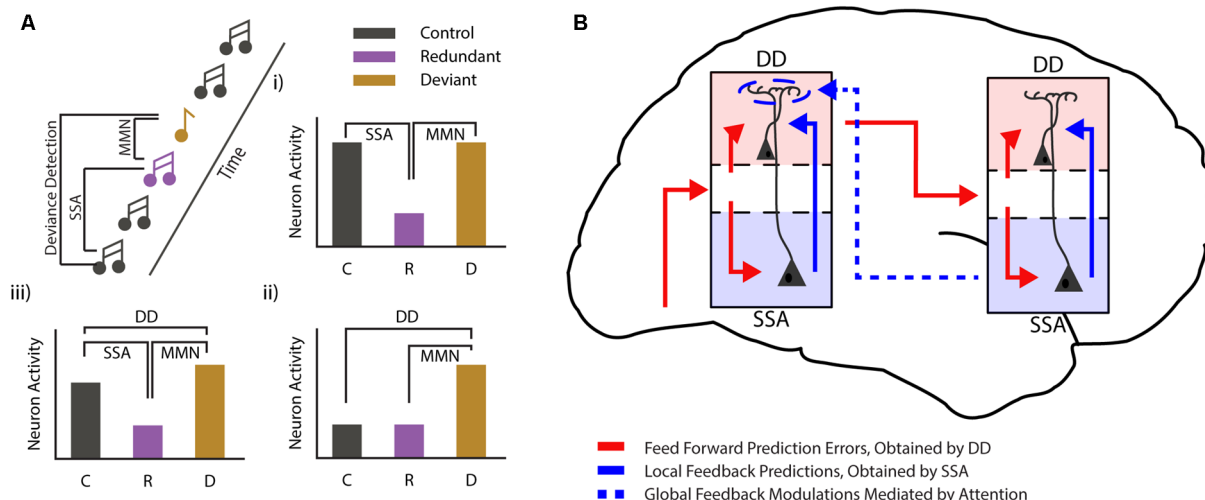
## MISMATCH NEGATIVITY: PRESERVED TOP-DOWN MODULATION DURING NON-CONSCIOUS BRAIN STATES?

The mismatch negativity response (MMNr) is a well-studied electrophysiological phenomenon that occurs in the human brain after violation of a rule, which is established by a sequence of repeated stimuli (Näätänen et al., 2007; Garrido et al., 2009b). The MMNr reflects the brain's capacity to automatically detect unpredicted sensory changes in our environment, without the need for attention (Tiitinen et al., 1994; Näätänen et al., 2001; Stefanics et al., 2014), and is impaired in several psychological afflictions such as schizophrenia, attention-deficit hyperactive disorder and psychosis (Erickson et al., 2016; Näätänen et al., 2016). In recent years the interest for the MMNr has grown considerably, especially in relation to the hierarchical PC framework, bringing a new plethora of insights from both human and rodent experiments.

The MMNr is often studied using an oddball paradigm, in which a particular stimulus (standard) is repeatedly presented to a subject, but sometimes unpredictably changed to a different, unexpected stimulus (deviant, see **Figure 2A**). The idea behind this paradigm is that the repetition of a particular stimulus or event leads to the formation of a prediction related to the frequency or probability at which this event occurs. Stimuli that deviate from the frequently presented one elicit a strong mismatch response. Thus, the MMNr can be interpreted as a prediction error to a stimulus that does not match the statistical regularities of sensory stimuli being perceived, thus updating our internal representation of the world (Friston, 2005; Garrido et al., 2009b; Stefanics et al., 2014).

## Different Components of the Mismatch Negativity Response

The MMNr can be split into two functionally distinct components (Hamm and Yuste, 2016; Harms et al., 2016). The first component arises as an effect of stimulus repetition. Repetition of a particular event leads to a decreased neuronal response to that event, a phenomenon often referred to as repetition suppression or SSA. SSA has been found in many different neural systems and at various levels: from single



**FIGURE 2 |** Mismatch negativity during different states of consciousness. **(A)** Cartoon showing the commonly used auditory oddball paradigm and the different comparisons that can be used to differentiate between stimulus-specific adaptation (SSA) and deviance detection (DD). Three bar graphs (i–iii) show the hypothetical cases in which classical MMN can be observed. In one case (i) MMN can be fully explained by SSA. Another example (ii) shows a pure deviance detecting neuron, while the final example (iii) shows MMN as it is most likely found in awake subjects, with both SSA and DD. **(B)** Schematic showing how local DD may be maintained during loss of consciousness, while global predictions are lost. Global, long-range feedback projections are reduced during loss of consciousness, while local connectivity is maintained. This suggests that low-level predictions might arise from local connectivity (for example from deep to superficial layers), while more complex, global prediction requires top-down modulation. These top-down projections mostly target interneurons in the most superficial layer of the cortex, where they are in a prime position to modulate neurons from both the superficial and the deep layers of the cortex.

neurons in the cortex of primates (Miller and Desimone, 1994), cats (Ulanovsky et al., 2003) and rodents (Taaseh et al., 2011), to field and imaging recordings in rodents (Chen et al., 2015; Hamm and Yuste, 2016; Parras et al., 2017; Hamm et al., 2018) and humans (Garrido et al., 2009a). While SSA closely resembles the MMNr originally described in humans (Ulanovsky et al., 2003; Nelken and Ulanovsky, 2007) it does not inherently have the second crucial component of the MMNr, true deviance detection (DD). For DD to occur the increased response to the deviant compared to the standard stimulus should not be fully explained by the decrease in neuronal firing caused by SSA (i.e., it should not simply be equal to the response elicited when the standard stimulus is not presented in a sequence). Rather, the increase in this response should have added value in detecting the occurrence of a deviant or irregular event (Hamm and Yuste, 2016; Harms et al., 2016). In this paragraph, we will discuss the relationship between SSA and DD, and how they are related to the MMNr.

Repetition of a particular stimulus leads to neuronal adaptation in the response to this stimulus, which can be characterized by the decrease of the neuronal response over repetitions. There are several hypotheses on which neuronal mechanisms could underlie this adaptation. Some models suggest that relatively simple feedforward mechanisms, such as neuronal fatigue, may fully explain the effects of stimulus repetition (Grill-Spector et al., 2006). Indeed, it is possible that SSA can be explained, at least in part, by purely feedforward mechanisms (Garrido et al., 2009a; Farley et al., 2010; May and Tiitinen, 2010), where (synapses of) neurons that are

responsive to the standard stimulus adapt over time, resulting in responsive depression, while the neurons responding to the deviant stimulus are “fresh” and are activated in full. A more likely case, however, is that multiple mechanisms contribute to repetition suppression under different conditions (Grill-Spector et al., 2006).

Another possible view on SSA comes from the PC framework. According to this framework, the decrease in neuronal activation upon repeated stimulation reflects the brain’s ability to predict that stimulus (Friston, 2005, 2008, 2010). This is achieved by neuronal processes optimized to probabilistically represent causes of sensory inputs. These processes can be seen as the building of an internal model representing our external environment. Keeping this model up-to-date requires a constant interaction between top-down predictions and bottom-up prediction errors (Bastos et al., 2012). An increased ability of this generative model to predict sensory inputs is reflected in a decrease of neuronal activity necessary to update the model. In other words, SSA can be considered as a product of perceptual learning where the predicted part of a sensory input is “explained away.”

DD, on the other hand, is an increased neuronal response to deviant stimuli, and can be considered a neural correlate of error signaling or memory update, where the prediction does not match the actual input and an adjustment of the generative model is called for.

Though SSA has been extensively studied in animal models, until recently it remained unclear whether MMN and more specifically DD could also be found in rodents (Taaseh et al., 2011; Chen et al., 2015; Hamm and Yuste, 2016;

Parras et al., 2017; Hamm et al., 2018). These studies, paving the way for investigating the microcircuit-level mechanisms of DD, not only found that true DD is present in the rodent neocortex, but that SSA and DD can be observed within different time-windows of the whole MMNr (Chen et al., 2015; Hamm and Yuste, 2016). While SSA is mainly found in earlier time components (40–80 ms after stimulus onset), DD is predominantly visible in a later time window (120–240 ms after stimulus onset). Furthermore, these separate components possibly involve distinct cortical networks and neuronal populations (Chen et al., 2015; Natan et al., 2015; Hamm and Yuste, 2016; Parras et al., 2017; Hamm et al., 2018). Though questions remain on how experimental results regarding SSA and DD in animal models compare to MMNr in human subjects, it can be assumed that they are all part of the same PC process.

## MMN and the Canonical Microcircuits for Predictive Coding

The generation of the MMNr is a hierarchical process. While forms of SSA can already be found in the early stages of sensory processing such as the thalamus (Natan et al., 2015; Parras et al., 2017; Hamm et al., 2018), DD seems to be absent in these early processing steps and appears later in the hierarchy, starting from (primary) cortical areas. Moreover, the proportion of neurons showing true DD increases when going up in the cortical hierarchy (Parras et al., 2017; Hamm et al., 2018). This has been taken as an indication that, while SSA can arise from purely feedforward processing streams, DD is more likely dependent on recurrent or feedback processing. This notion fits very well with the connectivity scheme proposed by the PC theory (Bastos et al., 2012; Aukstulewicz and Friston, 2016)—see “The Role of Feedback Processing Across Different Brain States” section.

In the PC framework, top-down predictions play an important role in eliciting mismatch or error-signals. Excitatory connections from (higher-order) cortical areas send predictions to predominantly inhibitory neuronal population in earlier sensory areas (Bastos et al., 2012). These inhibitory neurons enable local pyramidal cells to compare sensory inputs with these predictions, resulting in processes such as SSA (Bastos et al., 2012; Aukstulewicz and Friston, 2016; Yarden and Nelken, 2017). Indeed, it has been shown that deactivating auditory cortex reduces SSA in the superior colliculus of rats, though SSA is not completely extinguished (Anderson and Malmierca, 2013). Furthermore, inhibiting projections from the anterior cingulate cortex (Cg1)—a high-order cortical area involved in modulating visual responses (Zhang et al., 2014; Fiser et al., 2016)—to the primary visual cortex (V1) abolishes DD at a population level in V1 itself (Hamm et al., 2018). This would support a primary role for long-range feedback projections—which are also thought to be essential for conscious processing—in enabling DD. It is important to note, however, that, when splitting neuronal activity into populations that are “adapting,” “deviance detecting” or “non-modulated,” the activity of “deviance detecting” neurons in V1 remains intact (Hamm et al., 2018). In other words, while DD seems to disappear

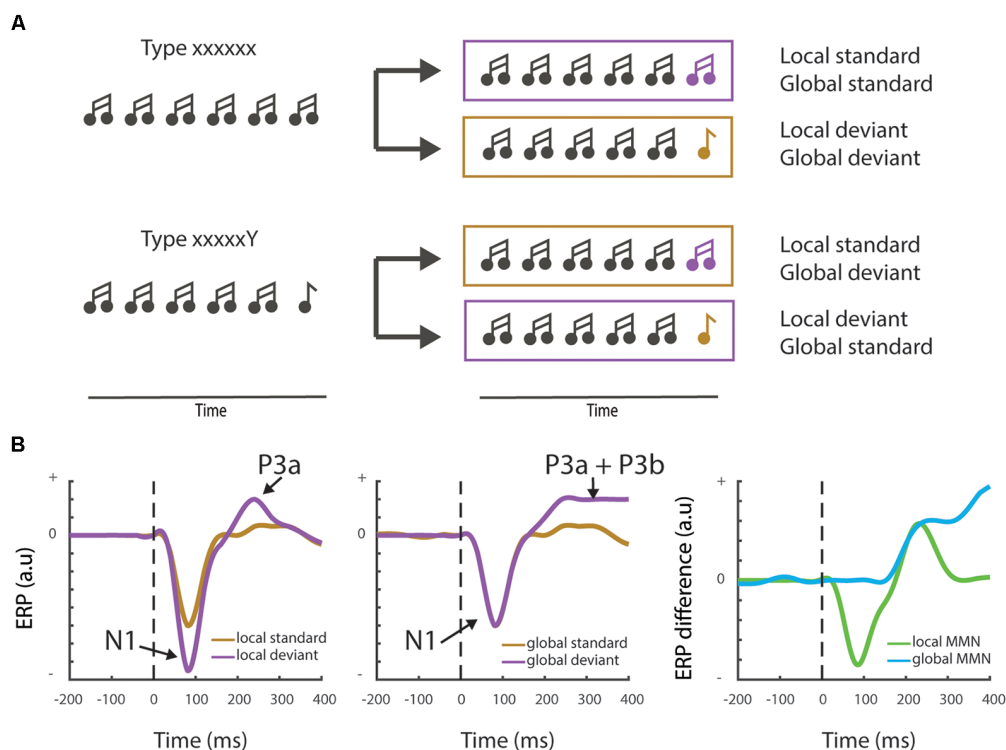
when looking at the gross activity of V1, it is preserved at the single-neuron level. This raises the question of how important long-range feedback connections are for facilitating mismatch responses in early sensory cortices, and what their exact role is. Is it possible for example that mismatches caused by low-complexity stimulus features such as visual orientation can be solved within the visual cortex through local (recurrent) connectivity?

To answer this question, we should make a distinction between local/low-level expectations that can operate independently of active cognitive processing (the pre-attentive part of the MMNr), and global/higher-order generalizations that require cognitive or attentional modulation. One way to make this dissociation is by using an adapted version of the oddball paradigm. The local-global paradigm dissociates between two types of predictions, based on local probabilities vs. global rules (Figure 3A). It uses blocks of stimuli where a sequence of for instance five stimuli (“xxxxx”) is presented interspersed with infrequent blocks in which the final stimulus is changed to a rare/deviant stimulus (“xxxxY”; Bekinschtein et al., 2009; Wacongne et al., 2011). Importantly, in other blocks, the frequent sequence is of the “xxxxY” type, while the deviant sequence consists of five identical stimuli (“xxxxx”). Evoked response potential (ERP) recordings reveal that the fifth, locally deviant tone of the standard “xxxxY” type, although fully predictable, still elicits a MMNr. However, only the rare violation sequence (or global deviant), which contains the five identical tones “xxxxx,” elicits a distinct and later novelty response, the “late positive complex” or P3b wave (Bekinschtein et al., 2009, Figure 3B). This P3b component has been associated with consciousness through processes such as attention and awareness—see Chennu and Bekinschtein (2012) for an overview. In the next paragraph, we will discuss how the MMNr and related ERP components, such as the P3b wave, are influenced by different states of awareness and how these changes are supported by (long-range) feedback connections.

## Mismatch Negativity in Relation to Consciousness

The MMNr has been investigated extensively in relation to different states of consciousness. It has been of interest in cognitive research because of its robustness: MMN can be observed in different states of consciousness, be it awake, asleep or anesthetized (Atienza et al., 2002; Koelsch et al., 2006). This has led to the belief that MMN is an automated response not dependent on active, conscious processing (Näätänen et al., 2001; Stefanics et al., 2014). However, ERP components related to the MMNr, especially the P3b wave, can be altered by conscious processes such as attention and awareness (Woldorff et al., 1998; Näätänen et al., 2007). Consequently, it has been hypothesized that, as the MMNr may be a key NCC, changes in its properties can be used to track different states of awareness, such as sedation or loss of consciousness. In a clinical setting, the estimation of a patient’s level of consciousness may be of paramount importance for the correct diagnosis of disorders of consciousness.





**FIGURE 3 |** Local vs. global mismatch responses. **(A)** Cartoon showing the local-global oddball paradigm and the distinct stimulus trains that differentiate between local and global deviants. Note that in the “xxxxxY” block the change of the final “Y” to an “x” stimulus induces a global deviant (“xxxxxx”), whereas it remains a standard stimulus at the local level. **(B)** Schematic examples (not actual data—see e.g., Bekinschtein et al., 2009 or Strauss et al., 2015 for measured electrophysiological traces) showing the typically observed field responses during a local-global mismatch paradigm. The left panel shows the classical Mismatch negativity response (MMNr) where the N1 component is more negative for the deviant response compared to the standard response. The local MMNr also generally show an early P3a response. Global deviants (middle panel) may not elicit a change in the N1 response, but induce a maintained sensory novelty or P3b response. The right panel depicts the difference waves (deviant—standard) generally used to visualize the MMNr.

In cognitive electrophysiology, the P3b wave is a part of the most widely studied P300 ERP component. It has been proposed as a marker of conscious perception of salient events or stimuli (Sutton et al., 1965) and more specifically as a sign of the “top-down” deployment of selective attention to task-relevant stimuli. While the P3b response, similarly to the MMNr, is often studied using a form of the oddball paradigm, the processes underlying the two responses may be very different. For example, the processing of statistical irregularities that elicits the MMNr is distinct from, and may not necessarily result in, conscious awareness of a stimulus (Chennu and Bekinschtein, 2012). Furthermore, the MMNr has been reported in humans over various states of consciousness (Atienza et al., 2002; Koelsch et al., 2006; Bekinschtein et al., 2009), whereas the P3b component seems dependent on aware processing of the stimulus (Sergent et al., 2005) and is diminished in states of impaired consciousness (Boly et al., 2011; Strauss et al., 2015; Nourski et al., 2018). Nevertheless, although the P3b response is usually considered an event separate from the MMNr, the two are often temporally overlapping. For example, recent studies in mice assess oddball and mismatch responses in temporal windows up to 400 ms after the onset of sensory stimuli (Chen et al., 2015; Hamm and Yuste,

2016), beyond the typical N1 response which peaks 100 ms after stimulus onsets and is associated with the MMNr in human EEG studies. Here, we will consider the P3b as an ERP component related to the MMNr. Future experiment will need to assess whether the P3b can be considered as a long-latency DD component of the MMNr or rather a separate neuronal event.

Upon loss of consciousness, detection of local deviants remains, at least partially, present in the human brain (Strauss et al., 2015; Nourski et al., 2018). Detection of global deviants is, on the other hand, absent during states of unconsciousness and deep sleep (Strauss et al., 2015; Nourski et al., 2018). Thus, global MMNr and the related P3b component are dependent on cognitive processes that are unavailable during loss of consciousness, whereas local DD remains partially present. While both the local DD component of the MMNr and the P3b response have been hypothesized to be dependent on temporal/frontal feedback connections (Friston, 2005, 2010; Garrido et al., 2009b; Buschman and Kastner, 2015), these differences indicate that distinct forms of feedback projections may be involved in the detection of local vs. global deviants.

As mentioned before (local) MMN has been reported under different states of awareness in both human and animal

experiments (Atienza et al., 2002; Koelsch et al., 2006; Parras et al., 2017). However, questions have been raised on whether these reports truly reflect the whole MMN process, including its deviance detecting component, or whether MMN under low conscious conditions can be fully explained by SSA. In a recent study, Strauss et al. (2015) showed disruption of the local MMNr at intermediate time components during non-REM sleep. These mid-range components have previously been associated with the error-signaling or deviance detecting processes underlying the MMNr (Wacongne et al., 2012). These results indicate that the feedback modulation that is necessary to support not only global but also local DD might be missing during non-conscious brain states. Indeed, another study showed that patients in a vegetative state (VS) display reductions in feedback connectivity, while feedforward connections are maintained (Boly et al., 2011)—but see King et al. (2011). Moreover, intermediate (DD-related) components of the MMNr are not present in these patients, whereas they are still reported in patients in a minimally conscious state (MCS). Thus, the main difference between MCS and VS patients seems to be the reduction of feedback connectivity in VS patients, leading to a disruption in DD. Finally, a recent animal study that supports the previous findings showed that inactivation of feedback connections from frontal cortex to visual cortex disrupts local DD processes in the latter (Hamm et al., 2018).

Altogether, the studies mentioned in the previous paragraphs indicate that: (1) feedback connections are necessary to maintain the (local) DD component of MMN, and (2) feedback connectivity is strongly reduced during non-conscious brain states. Interestingly, however, the same study showed that, while DD disappears upon inhibiting feedback projections when looking at the population level (i.e., in terms of multi-unit activity), at the single-neuron level DD is unchanged (Hamm et al., 2018). Moreover, other studies have reported both MMNr and (local) DD at the single-neuron level in anesthetized animals (Taaseh et al., 2011; Parras et al., 2017). Thus, while DD seems to disappear in population-level recordings during non-conscious states or when prefrontal cortex is inhibited (Boly et al., 2011; Strauss et al., 2015; Hamm et al., 2018), it is preserved in the cortex when looking at the single-neuron level (Parras et al., 2017; Hamm et al., 2018). These discrepancies highlight some of the knowledge gaps in our current understanding of the role of feedback projections in MMN generation and, potentially, their involvement in conscious responses.

Several distinct mechanisms may explain the aforementioned discrepancies. First, local DD might primarily depend—as classically hypothesized and similarly to global DD—on long-range feedback projections. During anesthesia/loss of consciousness such feedback is reduced (but not completely eliminated), and this dampens the extent of local DD in such a way that it is no longer visible at the population level. A second possibility is that, at least for local mismatches, DD is primarily due to feedforward or local (recurrent) mechanisms (Figure 2B). Feedback projections, in a manner similar to top-down attention, amplify and synchronize DD across neurons. This makes DD visible at the population level and

allows the “error” signal to be more effectively transmitted to higher-order areas. Current experimental evidence does not allow us to disambiguate between these two scenarios. In fact, no experiment has yet been able to either monitor or abolish (or even identify) all sources of feedback that might play a role in the generation of the local DD response. Consequently, only hypotheses can be made on its genesis. Addressing this problem is essential not only to provide an understanding of the mechanisms underlying MMN (which is clinically relevant for the diagnosis of several neuropsychiatric disorders), but also to better understand the neural basis of consciousness.

## DISCUSSION: ARE FEEDBACK CONNECTIONS ESSENTIAL FOR CONSCIOUSNESS?

Long-range feedback projections have been hypothesized to play an essential role in the generation of consciousness, yet conclusive evidence for this is lacking. First (“The Role of Feedback Processing Across Different Brain States” section), while it has been shown that long-range feedback is dampened in non-conscious brain states (Boly et al., 2011), some studies found exceptions to this general trend (Chennu et al., 2014), even for cortico-cortical top-down projections. Other studies even showed that, during Non-REM sleep, some forms of inter-areal communication may even be enhanced compared to wakefulness (Olcese et al., 2018a). One possible solution to this discrepancy between theoretical frameworks and experimental evidence may be that not all long-range, cortico-cortical feedback projections are equally important in the generation of consciousness. Identifying which feedback pathways support conscious processing is crucial to develop better markers for the level of consciousness. State-of-the-art diagnostic tools for linking brain dynamics to the level of consciousness, such as the perturbational complexity index (Casali et al., 2013), are able to characterize inter-areal communication in general but lack the specificity to probe specific connections. Second (“Mismatch Negativity: Preserved Top-Down Modulation During Non-conscious Brain States?” section), the role of feedback is being questioned even for fundamental brain mechanisms which have long been held to depend on it and to be hallmarks of conscious states, such as the DD component of the MMNr. Not only may the DD component still be present in non-conscious states, but also it may not be fully dependent on feedback projections (Hamm et al., 2018). In particular, a key question is whether feedback projections contribute to the generation of the DD response, or only facilitate it *via* synchronizing neuronal activity. In other words, is the DD response an error signal computed in higher-order areas (as the PC framework suggests), or is it generated locally and top-down modulation only amplifies it? And, if the latter proves to be the case, what are the local/recurrent mechanisms underlying DD, and is this consciously perceived? Finally, it remains to be addressed whether the feedback functional dynamics responsible for the

DD component of the MMNr are merely a correlate of consciousness (NCC) or rather represent a key mechanism underlying the emergence of consciousness. Addressing these questions is crucial for us to understand the neural bases of consciousness, yet remains challenging and requires an interdisciplinary approach.

First, an experimental paradigm in which DD responses are preserved in non-conscious states at the single-neuron—but not population—level needs to be developed for both humans and animals. Correlational human experiments involving intracranial recordings are needed to investigate whether single-neuron DD is consciously perceived even in the absence of long-range feedback and population-level signals. This can be done *via* retrospective reports (Koch et al., 2016; Siclari et al., 2017), and will allow us to understand if feedback projections are involved in generating conscious percepts, or rather enable to report those. The answer to this question is critical to reveal which cortical regions are involved in the genesis of conscious perception (Koch et al., 2016; Boly et al., 2017; Lamme, 2018). Second, causal explanations provided by animal experiments are subsequently crucial to go beyond the correlational level and reveal what the exact role of feedback projections is in the generation of DD responses. By leveraging and integrating techniques such as high-channel count electrophysiology, two-photon calcium imaging and optogenetics, it will be possible to uncover whether (and which) feedback projections play a role in the computation of the DD response, or whether they only have a modulatory nature. One additional intriguing outcome of such experiments (which can be only achieved by combining human and animal experiments) would also be to finally reveal the differential role of top-down projection from frontal to posterior cortices, or within the posterior cortex (parietal, temporal and occipital regions), and address the debate on which cortical regions support conscious processing (Koch et al., 2016; Boly et al., 2017; Odegaard et al., 2017; Lamme, 2018).

## CONCLUSIONS

How does the brain transform sensory stimuli into perceptual experiences? And how does it generate an internal model of the world? These are questions that have puzzled philosophers since the dawn of civilization. Neuroscientists have in the past decades started to formulate theories on how the brain might generate consciousness, yet all theories are lacking in terms of specific, circuit-level mechanisms. Theories focused on addressing the hard problem of consciousness (e.g., IIT) provide a foundational framework of the properties a conscious system needs to possess, but do not specify how this is reflected in terms of neuron-level (micro)circuits. The same is true for neurophysiology-based theories (recurrent processing, GNW, PC framework). General, mesoscopic-scale principles have been outlined, yet very limited insight is provided about the corresponding single-neuron correlates. Irrespective of one's philosophical or theoretical stance (see the debate between Chalmers and Dennett), this is a key question which was until recently not

addressable, but that developments in neurotechnology make within reach.

While functional long-range feedback projections are generally agreed to be closely associated to the presence of consciousness, recent studies have begun to indicate that not all feedback (not even within the cortex) may play the same role. The precise function of feedback has proven to be elusive even for phenomena (such as DD) that have historically been considered dependent on feedback. In order to eventually understand consciousness, therefore, two questions must be addressed: which forms of feedback support consciousness—and are not just a correlate of it? And how do they operate, at the microcircuit level?

Answering these questions will make it possible to eventually test the mechanistic role of feedback connections in generating consciousness. Is proper feedback sufficient to generate consciousness, or only necessary, provided that other conditions are met? If proper feedback were sufficient to generate consciousness in information processing systems, any biological or artificial device would have the potential for being conscious (although so far only brains may have achieved the right form of feedback). Although certain configurations of feedback projections are most likely essential to achieve functions normally associated with consciousness—Chalmers' easy problems, but see also Dehaene et al. (2017)—we concur with Tononi and Koch (2015) that merely reproducing a function does not imply possessing consciousness (Tononi and Koch, 2015; Tononi et al., 2016). Thus, while a computer simulation with a proper, yet virtual, form of feedback would likely be able to match conscious systems from a functional point of view, we surmise that a proper physical substrate is a pre-requisite to achieve real consciousness, i.e., experience—see Tononi and Koch (2015) for an in-depth discussion on the topic. Rather, we support the hypothesis that proper feedback projections are necessary to achieve consciousness in computational systems (neuron-based, artificial, etc.), but only if these possess an appropriate physical substrate (Tononi and Koch, 2015; Tononi et al., 2016). Nevertheless, these fascinating questions can only be achieved after a deeper understanding of the role of feedback projections in conscious processing. This will finally allow, we believe, to unveil the mystery which the mechanisms of consciousness still represent.

## AUTHOR CONTRIBUTIONS

TS, CB and UO wrote the manuscript.

## FUNDING

This study was supported by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2) to CB and UO, and by the FLAG-ERA JTC 2015 project CANON [co-financed by the Netherlands Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NWO)] to CB and UO.

## REFERENCES

- Adesnik, H., and Naka, A. (2018). Cracking the function of layers in the sensory cortex. *Neuron* 100, 1028–1043. doi: 10.1016/j.neuron.2018.10.032
- Akeju, O., and Brown, E. N. (2017). Neural oscillations demonstrate that general anesthesia and sedative states are neurophysiologically distinct from sleep. *Curr. Opin. Neurobiol.* 44, 178–185. doi: 10.1016/j.conb.2017.04.011
- Anderson, L. A., and Malmierca, M. S. (2013). The effect of auditory cortex deactivation on stimulus-specific adaptation in the inferior colliculus of the rat. *Eur. J. Neurosci.* 37, 52–62. doi: 10.1111/ejn.12018
- Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746. doi: 10.1016/j.neubiorev.2011.12.003
- Atienza, M., Cantero, J. L., and Dominguez-Marín, E. (2002). Mismatch negativity (MMN): an objective measure of sensory memory and long-lasting memories during sleep. *Int. J. Psychophysiol.* 46, 215–225. doi: 10.1016/s0167-8760(02)00113-7
- Auksztulewicz, R., and Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex* 80, 125–140. doi: 10.1016/j.cortex.2015.11.024
- Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends Cogn. Sci.* 6, 47–52. doi: 10.1016/s1364-6613(00)01819-2
- Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog. Brain Res.* 150, 45–53. doi: 10.1016/s0079-6123(05)50004-9
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038
- Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., et al. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85, 390–401. doi: 10.1016/j.neuron.2014.12.018
- Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. U S A* 106, 1672–1677. doi: 10.1073/pnas.0809667106
- Boly, M., Garrido, M. L., Gosseries, O., Bruno, M.-A., Boveroux, P., Schnakers, C., et al. (2011). Preserved feedforward but impaired top-down processes in the vegetative state. *Science* 332, 858–862. doi: 10.1126/science.1202043
- Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., and Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *J. Neurosci.* 37, 9603–9613. doi: 10.1523/JNEUROSCI.3218-16.2017
- Bosman, C. A., and Aboitiz, F. (2015). Functional constraints in the evolution of brain circuits. *Front. Neurosci.* 9:303. doi: 10.3389/fnins.2015.00303
- Bosman, C. A., Lansink, C. S., and Pennartz, C. M. A. (2014). Functions of gamma-band synchronization in cognition: from single circuits to functional diversity across cortical and subcortical systems. *Eur. J. Neurosci.* 39, 1982–1999. doi: 10.1111/ejn.12606
- Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J., and Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. *Proc. Natl. Acad. Sci. U S A* 108, 11262–11267. doi: 10.1073/pnas.1011284108
- Buschman, T. J., and Kastner, S. (2015). From behavior to neural dynamics: an integrated theory of attention. *Neuron* 88, 127–144. doi: 10.1016/j.neuron.2015.09.017
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* 5:198ra105. doi: 10.1126/scitranslmed.3006294
- Casarotto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fecchio, M., Napolitani, M., et al. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Ann. Neurol.* 80, 718–729. doi: 10.1002/ana.24779
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219.
- Chao, Z. C., Takaura, K., Wang, L., Fujii, N., and Dehaene, S. (2018). Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. *Neuron* 100, 1252.e3–1266.e3. doi: 10.1016/j.neuron.2018.10.004
- Chen, I.-W., Helmchen, F., and Lütcke, H. (2015). Specific early and late oddball-evoked responses in excitatory and inhibitory neurons of mouse auditory cortex. *J. Neurosci.* 35, 12560–12573. doi: 10.1523/JNEUROSCI.2240-15.2015
- Chennu, S., and Bekinschtein, T. A. (2012). Arousal modulates auditory attention and awareness: insights from sleep, sedation and disorders of consciousness. *Front. Psychol.* 3:65. doi: 10.3389/fpsyg.2012.00065
- Chennu, S., Finoia, P., Kamau, E., Allanson, J., Williams, G. B., Monti, M. M., et al. (2014). Spectral signatures of reorganised brain networks in disorders of consciousness. *PLoS Comput. Biol.* 10:e1003887. doi: 10.1371/journal.pcbi.1003887
- Clark, A. (2013). Whatever next? Predictive brains, situated agents and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/s0140525x12000477
- Cohen, M. A., and Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends Cogn. Sci.* 15, 358–364. doi: 10.1016/j.tics.2011.06.008
- Constantinople, C. M., and Bruno, R. M. (2013). Deep cortical layers are activated directly by thalamus. *Science* 340, 1591–1594. doi: 10.1126/science.1236425
- Crick, F., and Koch, C. (1990). Towards a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275.
- Crick, F., and Koch, C. (2003). A framework for consciousness. *Nat. Neurosci.* 6, 119–126. doi: 10.1038/nn0203-119
- Dehaene, S., Lau, H., and Kouider, S. (2017). What is consciousness and could machines have it? *Science* 358, 486–492. doi: 10.1126/science.aan8871
- Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little, Brown and Co.
- Dennett, D. C. (2018). Facing up to the hard question of consciousness. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373:20170342. doi: 10.1098/rstb.2017.0342
- Destexhe, A., Contreras, D., and Steriade, M. (1999). Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. *J. Neurosci.* 19, 4595–4608. doi: 10.1523/jneurosci.19-11-04595.1999
- Douglas, R. J., and Martin, K. A. C. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27, 419–451. doi: 10.1146/annurev.neuro.27.070203.144152
- Ercsey-Ravasz, M., Markov, N. T., Lamy, C., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., et al. (2013). A predictive network model of cerebral cortical connectivity based on a distance rule. *Neuron* 80, 184–197. doi: 10.1016/j.neuron.2013.07.036
- Erickson, M. A., Ruffle, A., and Gold, J. M. (2016). A meta-analysis of mismatch negativity in schizophrenia: from clinical risk to disease specificity and progression. *Biol. Psychiatry* 79, 980–987. doi: 10.1016/j.biopsych.2015.08.025
- Fahrenfort, J. J., Scholte, H. S., and Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *J. Cogn. Neurosci.* 19, 1488–1497. doi: 10.1162/jocn.2007.19.9.1488
- Fahrenfort, J. J., Scholte, H. S., and Lamme, V. A. F. (2008). The spatiotemporal profile of cortical processing leading up to visual perception. *J. Vis.* 8:12. doi: 10.1167/8.1.12
- Fahrenfort, J. J., Snijders, T. M., Heinen, K., van Gaal, S., Scholte, H. S., and Lamme, V. A. F. (2012). Neuronal integration in visual cortex elevates face category tuning to conscious face perception. *Proc. Natl. Acad. Sci. U S A* 109, 21504–21509. doi: 10.1073/pnas.1207414110
- Farley, B. J., Quirk, M. C., Doherty, J. J., and Christian, E. P. (2010). Stimulus-specific adaptation in auditory cortex is an NMDA-independent process distinct from the sensory novelty encoded by the mismatch negativity. *J. Neurosci.* 30, 16475–16484. doi: 10.1523/jneurosci.2793-10.2010
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1
- Ferrarelli, F., Massimini, M., Sarasso, S., Casali, A., Riedner, B. A., Angelini, G., et al. (2010). Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proc. Natl. Acad. Sci. U S A* 107, 2681–2686. doi: 10.1073/pnas.0913008107
- Fiser, A., Mahringer, D., Oyibo, H. K., Petersen, A. V., Leinweber, M., and Keller, G. B. (2016). Experience-dependent spatial expectations in mouse visual cortex. *Nat. Neurosci.* 19, 1658–1664. doi: 10.1038/nn.4385
- Flores, F. J., Hartnack, K. E., Fath, A. B., Kim, S.-E., Wilson, M. A., Brown, E. N., et al. (2017). Thalamocortical synchronization during induction



- and emergence from propofol-induced unconsciousness. *Proc. Natl. Acad. Sci. U S A* 114, E6660–E6668. doi: 10.1073/pnas.1700148114
- Fontolan, L., Morillon, B., Liegeois-Chauvel, C., and Giraud, A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat. Commun.* 5:4694. doi: 10.1038/ncomms5694
- Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron* 88, 220–235. doi: 10.1016/j.neuron.2015.09.034
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol.* 4:e1000211. doi: 10.1371/journal.pcbi.1000211
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Garrido, M. I., Kilner, J. M., Kiebel, S. J., Stephan, K. E., Baldeweg, T., and Friston, K. J. (2009a). Repetition suppression and plasticity in the human brain. *Neuroimage* 48, 269–279. doi: 10.1016/j.neuroimage.2009.06.034
- Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009b). The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120, 453–463. doi: 10.1016/j.clinph.2008.11.029
- Gazzaley, A., and Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends Cogn. Sci.* 16, 129–135. doi: 10.1016/j.tics.2011.11.014
- Gilbert, C. D., and Li, W. (2013). Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14, 350–363. doi: 10.1038/nrn3476
- Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23. doi: 10.1016/j.tics.2005.11.006
- Gugino, L. D., Chabot, R. J., Prichep, L. S., John, E. R., Formanek, V., and Aglio, L. S. (2001). Quantitative EEG changes associated with loss and return of consciousness in healthy adult volunteers anaesthetized with propofol or sevoflurane. *Br. J. Anaesth.* 87, 421–428. doi: 10.1093/bja/87.3.421
- Hall, S. D., Yamawaki, N., Fisher, A. E., Clauss, R. P., Woodhall, G. L., and Stanford, I. M. (2010). GABA(A) alpha-1 subunit mediated desynchronization of elevated low frequency oscillations alleviates specific dysfunction in stroke—a case report. *Clin. Neurophysiol.* 121, 549–555. doi: 10.1016/j.clinph.2009.11.084
- Hamm, J. P., and Yuste, R. (2016). Somatostatin interneurons control a key component of mismatch negativity in mouse visual cortex. *Cell Rep.* 16, 597–604. doi: 10.1016/j.celrep.2016.06.037
- Hamm, J. P., Shymkiv, Y., Han, S., Yang, W., and Yuste, R. (2018). Cortical subnetworks encode context of a visual stimulus. *bioRxiv* [Preprint]. doi: 10.1101/452219
- Harms, L., Michie, P. T., and Näätänen, R. (2016). Criteria for determining whether mismatch responses exist in animal models: focus on rodents. *Biol. Psychol.* 116, 28–35. doi: 10.1016/j.biopsycho.2015.07.006
- Harris, K. D., and Shepherd, G. M. G. (2015). The neocortical circuit: themes and variations. *Nat. Neurosci.* 18, 170–181. doi: 10.1038/nn.3917
- Hubel, D. H., and Wiesel, T. N. (1977). Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.* 198, 1–59. doi: 10.1098/rspb.1977.0085
- King, J.-R., Bekinschtein, T., and Dehaene, S. (2011). Comment on “Preserved feedforward but impaired top-down processes in the vegetative state.” *Science* 334:1203. doi: 10.1126/science.1210012
- Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22
- Koelsch, S., Heinke, W., Sammler, D., and Olthoff, D. (2006). Auditory processing during deep propofol sedation and recovery from unconsciousness. *Clin. Neurophysiol.* 117, 1746–1759. doi: 10.1016/j.clinph.2006.05.009
- Kwon, S. E., Yang, H., Minamisawa, G., and O'Connor, D. H. (2016). Sensory and decision-related activity propagate in a cortical feedback loop during touch perception. *Nat. Neurosci.* 19, 1243–1249. doi: 10.1038/nn.4356
- Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501. doi: 10.1016/j.tics.2006.09.001
- Lamme, V. A. F. (2018). Challenges for theories of consciousness: seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373:20170344. doi: 10.1098/rstb.2017.0344
- Lamme, V. A., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/s0166-2236(00)01657-x
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pac. Philos. Q.* 64, 354–361. doi: 10.1111/j.1468-0114.1983.tb00207.x
- Lewis, C. M., Bosman, C. A., and Fries, P. (2015). Recording of brain activity across spatial scales. *Curr. Opin. Neurobiol.* 32, 68–77. doi: 10.1016/j.conb.2014.12.007
- Manita, S., Suzuki, T., Homma, C., Matsumoto, T., Odagawa, M., Yamada, K., et al. (2015). A top-down cortical circuit for accurate sensory perception. *Neuron* 86, 1304–1316. doi: 10.1016/j.neuron.2015.05.006
- Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., et al. (2014a). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb. Cortex* 24, 17–36. doi: 10.1093/cercor/bhs270
- Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., et al. (2014b). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.* 522, 225–259. doi: 10.1002/cne.23458
- Markov, N. T., Ercsey-Ravasz, M., Lamy, C., Ribeiro Gomes, A. R., Magrou, L., Misery, P., et al. (2013a). The role of long-range connections on the specificity of the macaque interareal cortical network. *Proc. Natl. Acad. Sci. U S A* 110, 5187–5192. doi: 10.1073/pnas.1218972110
- Markov, N. T., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., and Kennedy, H. (2013b). Cortical high-density counterstream architectures. *Science* 342:1238406. doi: 10.1126/science.1238406
- Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. doi: 10.1126/science.1117256
- Massimini, M., Ferrarelli, F., Sarasso, S., and Tononi, G. (2012). Cortical mechanisms of loss of consciousness: insight from TMS/EEG studies. *Arch. Ital. Biol.* 150, 44–55. doi: 10.4449/aib.v150i2.1361
- May, P. J. C., and Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology* 47, 66–122. doi: 10.1111/j.1469-8986.2009.00856.x
- Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.-M., Kennedy, H., and Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* 89, 384–397. doi: 10.1016/j.neuron.2015.12.018
- Mikulan, E., Hesse, E., Sedeño, L., Bekinschtein, T., Sigman, M., García, M. D. C., et al. (2018). Intracranial high- $\gamma$  connectivity distinguishes wakefulness from sleep. *Neuroimage* 169, 265–277. doi: 10.1016/j.neuroimage.2017.12.015
- Miller, E. K., and Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science* 263, 520–522. doi: 10.1126/science.8290960
- Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (2001). “Primitive intelligence” in the auditory cortex. *Trends Neurosci.* 24, 283–288. doi: 10.1016/s0166-2236(00)01790-2
- Näätänen, R., Todd, J., and Schall, U. (2016). Mismatch negativity (MMN) as biomarker predicting psychosis in clinically at-risk individuals. *Biol. Psychol.* 116, 36–40. doi: 10.1016/j.biopsycho.2015.10.010
- Naccache, L. (2018). Why and how access consciousness can account for phenomenal consciousness. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373:20170357. doi: 10.1098/rstb.2017.0357
- Natan, R. G., Briguglio, J. J., Mwilambwe-Tshilobo, L., Jones, S. I., Aizenberg, M., Goldberg, E. M., et al. (2015). Complementary control of sensory adaptation by two types of cortical interneurons. *Elife* 4:e09868. doi: 10.7554/elif.09868
- Nelken, I., and Ulanovsky, N. (2007). Mismatch negativity and stimulus-specific adaptation in animal models. *J. Psychophysiol.* 21, 214–223. doi: 10.1027/0269-8803.21.34.214
- Nourski, K. V., Steinschneider, M., Rhone, A. E., Kawasaki, H., Howard, M. A., and Banks, M. I. (2018). Auditory predictive coding across awareness states under anesthesia: an intracranial electrophysiology study. *J. Neurosci.* 38, 8441–8452. doi: 10.1523/JNEUROSCI.0967-18.2018

- Odegaard, B., Knight, R. T., and Lau, H. (2017). Should a few null findings falsify prefrontal theories of conscious perception? *J. Neurosci.* 37, 9593–9602. doi: 10.1523/JNEUROSCI.3217-16.2017
- Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput. Biol.* 10:e1003588. doi: 10.1371/journal.pcbi.1003588
- Olcese, U. (2018). Non-REM sleep and the neural correlates of consciousness: more than meets the eye. *Arch. Ital. Biol.* 156, 137–148. doi: 10.12871/00039829201835
- Olcese, U., Bos, J. J., Vinck, M., and Pennartz, C. M. A. (2018a). Functional determinants of enhanced and depressed interareal information flow in nonrapid eye movement sleep between neuronal ensembles in rat cortex and hippocampus. *Sleep* 41, 1–18. doi: 10.1093/sleep/zsy167
- Olcese, U., Oude Lohuis, M. N., and Pennartz, C. M. A. (2018b). Sensory processing across conscious and nonconscious brain states: from single neurons to distributed networks for inferential representation. *Front. Syst. Neurosci.* 12:49. doi: 10.3389/fnsys.2018.00049
- Olcese, U., Bos, J. J., Vinck, M., Lankelma, J. V., van Mourik-Donga, L. B., Schlumm, F., et al. (2016). Spike-based functional connectivity in cerebral cortex and hippocampus: loss of global connectivity is coupled to preservation of local connectivity during non-REM sleep. *J. Neurosci.* 36, 7676–7692. doi: 10.1523/JNEUROSCI.4201-15.2016
- Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., and Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat. Commun.* 8:2148. doi: 10.1038/s41467-017-02038-6
- Pennartz, C. M. A. (2009). Identification and integration of sensory modalities: neural basis and relation to consciousness. *Conscious. Cogn.* 18, 718–739. doi: 10.1016/j.concog.2009.03.003
- Pennartz, C. M. A. (2015). *The Brain's Representational Power*. The MIT Press Available online at: <https://mitpress.mit.edu/books/brains-representational-power>. Accessed November 24, 2017.
- Pesaran, B., Vinck, M., Einevoll, G. T., Sirota, A., Fries, P., Siegel, M., et al. (2018). Investigating large-scale brain dynamics using field potential recordings: analysis and interpretation. *Nat. Neurosci.* 21, 903–919. doi: 10.1038/s41593-018-0171-8
- Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580
- Richter, C. G., Thompson, W. H., Bosman, C. A., and Fries, P. (2017). Top-down beta enhances bottom-up gamma. *J. Neurosci.* 37, 6698–6711. doi: 10.1523/JNEUROSCI.3771-16.2017
- Riesenhuber, M., and Poggio, T. (1999). Are cortical models really bound by the “binding problem”? *Neuron* 24, 87–93, 111–125. doi: 10.1016/S0896-6273(00)80824-7
- Rohenkohl, G., Bosman, C. A., and Fries, P. (2018). Gamma synchronization between V1 and V4 improves behavioral performance. *Neuron* 100, 953.e3–963.e3. doi: 10.1016/j.neuron.2018.09.019
- Sergent, C., and Dehaene, S. (2004). Neural processes underlying conscious perception: experimental findings and a global neuronal workspace framework. *J. Physiol. Paris* 98, 374–384. doi: 10.1016/j.jphysparis.2005.09.006
- Sergent, C., Baillet, S., and Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nat. Neurosci.* 8, 1391–1400. doi: 10.1038/nn1549
- Shepherd, G. M. (2011). The microcircuit concept applied to cortical evolution: from three-layer to six-layer cortex. *Front. Neuroanat.* 5:30. doi: 10.3389/fnana.2011.00030
- Siclari, F., and Tononi, G. (2017). Local aspects of sleep and wakefulness. *Curr. Opin. Neurobiol.* 44, 222–227. doi: 10.1016/j.conb.2017.05.008
- Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J. J., Riedner, B., et al. (2017). The neural correlates of dreaming. *Nat. Neurosci.* 20, 872–878. doi: 10.1038/nn.4545
- Stefanics, G., Kremláček, J., and Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Front. Hum. Neurosci.* 8:666. doi: 10.3389/fnhum.2014.00666
- Storm, J. F., Boly, M., Casali, A. G., Massimini, M., Olcese, U., Pennartz, C. M. A., et al. (2017). Consciousness regained: disentangling mechanisms, brain systems and behavioral responses. *J. Neurosci.* 37, 10882–10893. doi: 10.1523/JNEUROSCI.1838-17.2017
- Strauss, M., Sitt, J. D., King, J.-R., Elbaz, M., Azizi, L., Buiatti, M., et al. (2015). Disruption of hierarchical predictive coding during sleep. *Proc. Natl. Acad. Sci. U S A* 112, E1353–E1362. doi: 10.1073/pnas.1501026112
- Sutton, S., Braren, M., Zubin, J., and John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science* 150, 1187–1188. doi: 10.1126/science.150.3700.1187
- Taaseh, N., Yaron, A., and Nelken, I. (2011). Stimulus-specific adaptation and deviance detection in the rat auditory cortex. *PloS One* 6:e23369. doi: 10.1371/journal.pone.0023369
- Tiitinen, H., May, P., Reinikainen, K., and Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* 372, 90–92. doi: 10.1038/372090a0
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.* 5:42. doi: 10.1186/1471-2202-5-42
- Tononi, G., and Edelman, G. M. (1998a). Consciousness and complexity. *Science* 282, 1846–1851. doi: 10.1126/science.282.5395.1846
- Tononi, G., and Edelman, G. M. (1998b). Consciousness and the integration of information in the brain. *Adv. Neurol.* 77, 245–279.
- Tononi, G., and Koch, C. (2015). Consciousness: here, there and everywhere? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140167. doi: 10.1098/rstb.2014.0167
- Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461. doi: 10.1038/nrn.2016.44
- Townsend, R. G., Solomon, S. S., Chen, S. C., Pietersen, A. N. J., Martin, P. R., Solomon, S. G., et al. (2015). Emergence of complex wave patterns in primate cerebral cortex. *J. Neurosci.* 35, 4657–4662. doi: 10.1523/JNEUROSCI.4509-14.2015
- Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* 6, 391–398. doi: 10.1038/nn1032
- van Gaal, S., Lamme, V. A. F., Fahnenfort, J. J., and Ridderinkhof, K. R. (2011). Dissociable brain mechanisms underlying the conscious and unconscious control of behavior. *J. Cogn. Neurosci.* 23, 91–105. doi: 10.1162/jocn.2010.21431
- van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., et al. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. U S A* 111, 14332–14341. doi: 10.1073/pnas.1402773111
- van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., et al. (2018). The threshold for conscious report: signal loss and response bias in visual and frontal cortex. *Science* 360, 537–542. doi: 10.1126/science.aar7186
- Vijayan, S., Ching, S., Purdon, P. L., Brown, E. N., and Kopell, N. J. (2013). Thalamocortical mechanisms for the anteriorization of  $\alpha$  rhythms during propofol-induced unconsciousness. *J. Neurosci.* 33, 11070–11075. doi: 10.1523/JNEUROSCI.5670-12.2013
- von Stein, A., Chiang, C., and König, P. (2000). Top-down processing mediated by interareal synchronization. *Proc. Natl. Acad. Sci. U S A* 97, 14748–14753. doi: 10.1073/pnas.97.26.14748
- Wacongne, C., Changeux, J.-P., and Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *J. Neurosci.* 32, 3665–3678. doi: 10.1523/JNEUROSCI.5003-11.2012
- Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., and Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci. U S A* 108, 20754–20759. doi: 10.1073/pnas.1117807108
- Williams, S. T., Conte, M. M., Goldfine, A. M., Noirhomme, Q., Gosseries, O., Thonnard, M., et al. (2013). Common resting brain dynamics indicate a possible mechanism underlying zolpidem response in severe brain injury. *eLife* 2:e01157. doi: 10.7554/eLife.01157
- Woldorff, M. G., Hillyard, S. A., Gallen, C. C., Hampson, S. R., and Bloom, F. E. (1998). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology* 35, 283–292. doi: 10.1017/s0048577298961601

- Womelsdorf, T., Valiante, T. A., Sahin, N. T., Miller, K. J., and Tiesinga, P. (2014). Dynamic circuit motifs underlying rhythmic gain control, gating and integration. *Nat. Neurosci.* 17, 1031–1039. doi: 10.1038/nn.3764
- Yarden, T. S., and Nelken, I. (2017). Stimulus-specific adaptation in a recurrent network model of primary auditory cortex. *PLoS Comput. Biol.* 13:e1005437. doi: 10.1371/journal.pcbi.1005437
- Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J. P., Chang, W.-C., Jenvay, S., et al. (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. *Science* 345, 660–665. doi: 10.1126/science.1254126

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Sikkens, Bosman and Olcese. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms



# General Anesthesia: A Probe to Explore Consciousness

Vincent Bonhomme<sup>1,2,3\*</sup>, Cécile Staquet<sup>1,3</sup>, Javier Montupil<sup>1,2,3</sup>, Aline Defresne<sup>1,2,3</sup>, Murielle Kirsch<sup>1,3</sup>, Charlotte Martial<sup>4</sup>, Audrey Vanhaudenhuyse<sup>5</sup>, Camille Chatelle<sup>4</sup>, Stephen Karl Larroque<sup>4</sup>, Federico Raimondo<sup>4</sup>, Athena Demertzi<sup>6</sup>, Olivier Bodart<sup>4</sup>, Steven Laureys<sup>4</sup> and Olivia Gosseries<sup>4</sup>

<sup>1</sup>Anesthesia and Intensive Care Laboratory, GIGA-Consciousness, GIGA Institute, University of Liege, Liege, Belgium,

<sup>2</sup>University Department of Anesthesia and Intensive Care Medicine, Centre Hospitalier Régional de la Citadelle

(CHR Citadelle), Liege, Belgium, <sup>3</sup>Department of Anesthesia and Intensive Care Medicine, Centre Hospitalier Universitaire de

Liège (CHU Liège), Liege, Belgium, <sup>4</sup>Coma Science Group, GIGA-Consciousness, GIGA Institute, University of Liege, Liege,

Belgium, <sup>5</sup>Sensation & Perception Research Group, GIGA-Consciousness, Department of Algology, GIGA Institute, University

of Liege, Centre Hospitalier Universitaire de Liège (CHU Liège), Liege, Belgium, <sup>6</sup>Physiology of Cognition Research Lab,

GIGA-Consciousness, GIGA Institute, University of Liege, Liege, Belgium

## OPEN ACCESS

### Edited by:

James W. Grau,  
Texas A&M University, United States

### Reviewed by:

Harry Scheinin,  
University of Turku, Finland  
Uncheol Lee,  
University of Michigan, United States  
Seun Akeju,  
Massachusetts General Hospital and  
Harvard Medical School,  
United States

### \*Correspondence:

Vincent Bonhomme  
vincent.bonhomme@chuliege.be

**Received:** 04 April 2019

**Accepted:** 24 July 2019

**Published:** 14 August 2019

### Citation:

Bonhomme V, Staquet C, Montupil J, Defresne A, Kirsch M, Martial C, Vanhaudenhuyse A, Chatelle C, Larroque SK, Raimondo F, Demertzi A, Bodart O, Laureys S and Gosseries O (2019) General Anesthesia: A Probe to Explore Consciousness. *Front. Syst. Neurosci.* 13:36. doi: 10.3389/fnsys.2019.00036

General anesthesia reversibly alters consciousness, without shutting down the brain globally. Depending on the anesthetic agent and dose, it may produce different consciousness states including a complete absence of subjective experience (unconsciousness), a conscious experience without perception of the environment (disconnected consciousness, like during dreaming), or episodes of oriented consciousness with awareness of the environment (connected consciousness). Each consciousness state may potentially be followed by explicit or implicit memories after the procedure. In this respect, anesthesia can be considered as a proxy to explore consciousness. During the recent years, progress in the exploration of brain function has allowed a better understanding of the neural correlates of consciousness, and of their alterations during anesthesia. Several changes in functional and effective between-region brain connectivity, consciousness network topology, and spatio-temporal dynamics of between-region interactions have been evidenced during anesthesia. Despite a set of effects that are common to many anesthetic agents, it is still uneasy to draw a comprehensive picture of the precise cascades during general anesthesia. Several questions remain unsolved, including the exact identification of the neural substrate of consciousness and its components, the detection of specific consciousness states in unresponsive patients and their associated memory processes, the processing of sensory information during anesthesia, the pharmacodynamic interactions between anesthetic agents, the direction-dependent hysteresis phenomenon during the transitions between consciousness states, the mechanisms of cognitive alterations that follow an anesthetic procedure, the identification of an eventual unitary mechanism of anesthesia-induced alteration of consciousness, the relationship between network effects and the biochemical or sleep-wake cycle targets of anesthetic agents, as well as the vast between-studies variations in dose and administration mode, leading



to difficulties in between-studies comparisons. In this narrative review, we draw the picture of the current state of knowledge in anesthesia-induced unconsciousness, from insights gathered on propofol, halogenated vapors, ketamine, dexmedetomidine, benzodiazepines and xenon. We also describe how anesthesia can help understanding consciousness, we develop the above-mentioned unresolved questions, and propose tracks for future research.

**Keywords:** general anesthesia, consciousness, mechanisms, brain function, brain networks

## INTRODUCTION: GENERAL ANESTHESIA IS MORE COMPLEX THAN SIMPLY “ABSENCE OF CONSCIOUSNESS”

General anesthesia aims at providing patients with a state where they can tolerate unpleasant and/or noxious interventions, usually during a surgical procedure. Routinely, this involves a cocktail of medications ensuring an alteration of consciousness (pharmacological hypnosis) with absence of awareness of the surrounding environment, explicit recall of undercurrent events, a limitation of the stress response to nociception (anti-nociception), as well as immobility (muscle relaxation). General anesthesia does not shut down the brain globally and does not always produce a complete absence of consciousness. Similarly to disorders of consciousness in patients suffering from a severe brain insult (Aubinet et al., 2018), consciousness may be altered to various degrees, and its alteration may concern different consciousness elements as a function of the equilibrium between the inherent pharmacodynamic properties of anesthetic agents, their concentration in the body, and the intensity of the underlying stimulation by surgery.

Hence, one may not really speak about the depth of anesthetic hypnosis (Bayne et al., 2016), but rather about the presence or absence of one consciousness element or the other. As opposed to concepts that prevailed previously proposing that the brain was simply switched off by anesthesia, it is now clear that subjects may retain several higher-order brain functions until high concentrations of anesthetic agents are attained (Sleigh et al., 2018). If consciousness is defined as reflecting subjective experience/selfhood, or alternatively the sense of being a distinct entity capable of agency, sentience, narrative identity in time, and other higher-order components, anesthesia is capable of suppressing some of these components while keeping others functional (Sleigh et al., 2018). From a more basic and operational point of view, consciousness states that can be observed during general anesthesia are: (i) unconsciousness; (ii) disconnected consciousness; and (iii) connected consciousness (Table 1, Sanders et al., 2012). Unconsciousness can be defined as the inability to achieve any subjective experience and is expected to be the most common anesthetic state. During unconsciousness, reflex motor responses to stimulation may occur, such as arm withdrawal in response to noxious stimulation, but they are not purposeful and voluntary and do not imply a conscious connection with

the environment. Disconnected consciousness is characterized by the presence of a mental content, but no conscious perception of the environment. In that case, the mental content can be similar to dreaming (Radek et al., 2018), or more distorted like ketamine-induced psychedelic subjective experiences. Connected consciousness in anesthesia refers to the subjective experience of self, and perception of information from the environment, which may happen in episodes of variable duration, and are not that rare (Sanders et al., 2017; Linassi et al., 2018; Radek et al., 2018). Such episodes can be observed immediately after laryngoscopy and tracheal intubation in approximately 5% of patients (Sanders et al., 2017). Whereas disconnected consciousness refers to internal awareness, connected consciousness refers to both external and internal awareness during anesthesia. External awareness is defined as the perception of environmental sensory stimuli (e.g., auditory, visual, olfactory, or somesthetic), and internal awareness refers to all environmental stimuli-independent thoughts (e.g., inner speech, autobiographical memories, or mind-wandering; Vanhaudenhuyse et al., 2011).

The classic method to know in which states patients were during anesthesia is to ask them after recovery, but this delayed assessment misses a lot of events that are not followed by explicit memories. In anesthesia, both disconnected and connected consciousness is rarely followed by explicit recall after the procedure, but the possibility of implicit memories exists. The isolated forearm technique allows assessing connected consciousness “online” during general anesthesia with muscle relaxation, when the patient is unable to manifest consciousness because being paralyzed. This technique consists of isolating the patient’s forearm from the systemic blood circulation through a cuffed upper arm tourniquet, which is inflated before the administration of the neuromuscular blocking agent. In such a setup, the isolated forearm remains non-paralyzed, and the anesthesiologist may ask verbal instructions, such as squeezing the hand and observe the patient’s response (Sanders et al., 2017). The same assessment can be done without the need of a tourniquet when the procedure does not require the use of neuromuscular blocking agents. A debate exists regarding the real incidence and risk factors of connected consciousness episodes during routine anesthesia practice. Although higher incidences have already been reported, the recently published incidence of 5% (and higher in younger patients) immediately after tracheal intubation was claimed to be conservative by the authors, because unclear responses were

**TABLE 1** | Possible consciousness states during general anesthesia and their cognitive and mnemonic characteristics.

Consciousness state	Mental content		Sensory processing	Purposeful response to command	Explicit memory	Implicit memory
	External awareness (perception of environmental sensory stimuli)	Internal awareness (thoughts independent from the environment) and sense of self (agency, sentience, identity, ...)				
Unconsciousness	No	No	Possible (not accessible from the conscious field)	No	No	Possible
Disconnected consciousness	No	Yes	Yes (not related to external stimulation, e.g., seeing or smelling something during a dream)	No	Possible	Possible
Connected consciousness	Yes	Yes	Yes	Yes	Possible (probably rare)	Possible

Inspired from Vanhaudenhuyse et al. (2011), Sanders et al. (2012) and Sleight et al. (2018).

not counted and because assessment of connected consciousness did not occur during the remaining procedure (Sanders et al., 2017). During experimental studies, it remains important to gather subjective experiences immediately after recovery, to increase the likelihood of catching connectedness during unresponsive periods.

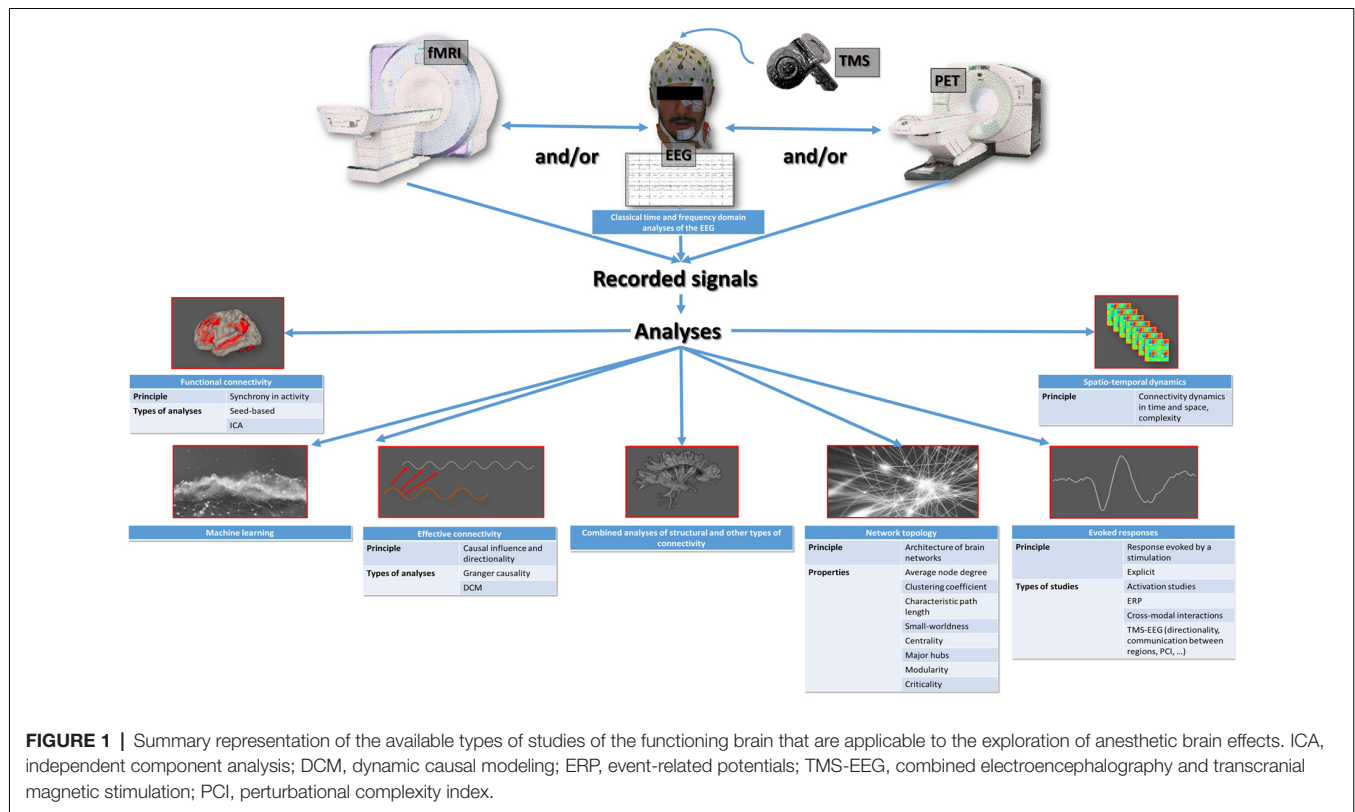
Classical processed electroencephalogram (EEG) indexes of the depth of anesthesia such as the Bispectral Index are not sensitive and specific enough to distinguish between the possible consciousness states of anesthesia (Gao et al., 2018). When asked to cite the consciousness states they would consider acceptable during anesthesia, subjects diversely appreciate one situation or the other. The eventuality of recall or feeling pain are major determinants of their appreciation (Rowley et al., 2017). By reversibly splitting and selectively altering some components of consciousness and selfhood, anesthetic agents are unique tools to explore the associated functional correlates.

Aside from the prerequisite of cortical arousal, which is controlled by sub-cortical sleep-wake cycle regulating systems, the complex phenomena of consciousness and subjective experience/selfhood mainly entail within and between neural networks interactions to generate and integrate information (Tononi, 2004). These activities translate into recordable electrical and metabolic complex signals that can be analyzed using sophisticated techniques. They have been and continue to be used in mostly single anesthetic drug studies in humans, and will ultimately shed light on the specific drug-related functional changes occurring during the different consciousness states of anesthesia, the hysteresis occurring during state transitions in one direction or the other (Kim et al., 2018), sensory processing, explicit or implicit memorization, and the frequently observed functional brain disturbances during the recovery period (Numan et al., 2017). We hereby narratively review the current state of the art in this domain, including studies on propofol, halogenated vapors, ketamine, dexmedetomidine, benzodiazepines and xenon, and underline questions that are not resolved to date.

## SEVERAL WAYS OF EXPLORING THE EFFECTS OF ANESTHETIC MEDICATIONS ON THE BRAIN

Currently, the exploration of brain function in humans mainly involves EEG recordings, combined with transcranial magnetic stimulation (TMS) or not, and functional brain imaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). The EEG signal corresponds to the underlying neuronal activities, and the blood oxygen level-dependent (BOLD) signal of fMRI (or the emitted gamma-ray energy in PET) corresponds to changes in regional cerebral blood flow in response to changes in activity. In addition to classical time and frequency domain analyses of the EEG (Marchant et al., 2014) and to classical activation studies of fMRI and PET (Bonhomme et al., 2001), sophisticated analyses can be applied to the recorded signals, with the primary aim to characterize the interactions between different brain regions. These interactions are thought to be the core mechanism of sensory processing and mental content generation (Lee and Mashour, 2018a). The number of possible analysis techniques is high, and each of them addresses a specific aspect of within-brain communication (**Figure 1**; Mashour and Hudetz, 2018; Staquet et al., 2018). It is therefore important to know exactly what they look at, in order to correctly interpret the results.

The most basic witness of brain communication is between-region synchrony in activity, or functional connectivity. When using fMRI, functional connectivity may be evidenced using hypothesis-based approaches, where brain regions show synchronized (as well as anti-correlated) activity with a predefined seed region (Boveroux et al., 2010; Bonhomme et al., 2016). Hypothesis-free data-driven analyses, for example, Independent Component Analysis (ICA), where signal synchronization is sought globally around the brain (Ribeiro de Paula et al., 2017), can also be used. Effective connectivity refers to the causal influence and directionality



of the activity of a brain region on the activity of another region, and may be evidenced using Granger causality (GC) and its derivatives (Nicolaou and Georgiou, 2014), or using dynamic causal modeling (DCM; Gómez et al., 2013; Crone et al., 2017). Another approach, based on graph theory, consists in depicting the architecture or topology of brain networks (Lee and Mashour, 2018a), which are defined as assemblies of nodes (specific brain regions) that are linked by edges (the communication route between two regions). For each network, a set of topological properties can be described, including the average node degree (average number of edges for each node), the clustering coefficient (a measure of segregation, which estimates the isolation of node groups from other groups, a high clustering coefficient corresponds to a high degree of functional specialization), the characteristic path length (average of the minimum number of edges needed to link two nodes, a short path length denotes high integration capacities or efficiency), the participation coefficient (also a measure of integration), the small-worldness (best compromise between functional specialization and efficiency, a small-world network has a high clustering coefficient and short path length), the centrality (index of the central character of a node), the major hubs (most influential nodes, with a high number of edges and a high degree of centrality), the modularity (estimates the separation of a network into modules, which are groups of nodes separated by a hub), and criticality (or metastability, a state between order and disorder resulting from a scale-free organization, allowing the maximum number of

possible configurations; Lee et al., 2017). Another mode of investigating brain communication relates to the evolution of the between-region interrelations over time and space, or spatio-temporal dynamics, and their complexity (Untergehrer et al., 2014; Wang et al., 2017; Cavanna et al., 2018; Huang et al., 2018; Muthukumaraswamy and Liley, 2018; Thiery et al., 2018), which is linked to criticality (Barttfeld et al., 2015; Lee et al., 2018). Finally, the analysis of evoked responses to stimulation offers other insights. The oldest studies in that respect were event-related potential EEG studies (Uhl et al., 1980; Plourde and Boylan, 1991), and the fMRI or PET activation studies (Bonhomme et al., 2008). The most recent ones look into how the brain handles sensory information (Lichtner et al., 2018a,b; Nourski et al., 2018) and between-region communication (Darracq et al., 2018a), as well as the directionality of information transfer (Sanders et al., 2018), and sensory cross-modal interactions (Bekinschtein et al., 2009). Some mixed approaches exist, melting one mode of analysis with another, such as those measuring the spatio-temporal complexity of TMS-evoked cortical responses (Casali et al., 2013; Bodart et al., 2017). Most recent emerging techniques use machine learning/decoding approaches through, for example, multivariate pattern analyses, but these have not much been applied to the anesthesia paradigm, yet (Liang et al., 2018). It is also possible to combine structural connectivity analyses (i.e., the exploration of anatomical connections through the white matter), and other types of connectivity analyses such as functional connectivity (Agarwal et al., 2016).

## CURRENT STATE OF THE ART—INTEGRATING THE AVAILABLE DATA

The process of characterizing changes in brain function during general anesthesia is still ongoing. To date, information has been obtained through studies investigating one single anesthetic agent, mostly during the installing/induction and maintenance phase of sedation, and not all analysis modalities have been applied to each of them. The studies looking at other moments in the anesthesia process (such as the specific aspects of the direction of anesthetic state transitions, recovery as compared to induction of anesthesia, or post-operative delirium), at sensory processing, at memory processes, and at less frequent consciousness states (such as connected consciousness without memorization) are still scarce. A summary of the currently available information follows for each commonly used anesthetic agent (Table 2).

### Propofol

Thanks to its safety and ease of use, the  $\gamma$ -amino-butyric acid (GABA) neurotransmission promoting agent propofol has been the most widely studied anesthetic agent with respect to brain mechanisms in primates and healthy humans. Years after the pioneering works of Fiset and Alkire, who were the first to demonstrate region-specific and dose-dependent effect of propofol on brain activity (Alkire et al., 1995; Fiset et al., 1999), it came out that this agent diminishes the randomness of the spontaneous and evoked EEG signal (Wang et al., 2017; Darracq et al., 2018b), alters long-distance cortical communication (Gómez et al., 2013) and reduces its complexity (Sarasso et al., 2015), disrupts within- and between-network functional connectivity in large-scale brain networks that sustain consciousness (Boveroux et al., 2010), and particularly functional fronto-parietal connectivity. Recent findings also suggest the suppression of the complexity of regions sparsely connected with large-scale brain networks as a mechanism of propofol-induced alteration in oriented reactivity to stimulation (Pappas et al., 2019). Effective connectivity, its dynamics, and its directionality are altered by propofol (Lee et al., 2009, 2013; Boly et al., 2012; Untergerhrer et al., 2014; Guldenmund et al., 2016; Sanders et al., 2018). Contrarily, propofol relatively preserves lower-order sensory networks connectivity but impedes the effective connectivity within them (Gómez et al., 2013). When looking at network topological properties, propofol-induced unresponsiveness is associated with increased local efficiency, particularly in parietal regions (Lee et al., 2017), reduced global efficiency (Kim et al., 2016), disturbance of the posterior parietal hub activity (Lee et al., 2013), fewer small-world properties (Barttfeld et al., 2015), a limitation of the repertoire of possible connectivity configurations (Barttfeld et al., 2015; Cavanna et al., 2018) by constraining the traffic of information to inflexible patterns (Mashour, 2018; Uhrig et al., 2018), and remoteness from criticality (Tagliazucchi et al., 2016) with preserved scale-free organization of networks (preservation of node size and node degree distribution; Liu et al., 2014). Under

propofol, the dynamics of within-brain interactions become more stable (Alonso et al., 2019). Propofol also synchronizes local activity and prolongs the timescales of long-distance communications (Huang et al., 2018). The consequences of propofol infusion on sensory processing involve a generalized alteration in short-latency evoked EEG responses to auditory novelty (which are thought to reflect pre-attentive processing) within higher-order cortical areas, but outside the auditory cortex. Long-latency responses to novelty (which may reflect conscious processing) are markedly suppressed by propofol in all regions (Nourski et al., 2018). Noxious information still reaches the cortex through the spinal cord, even under high propofol concentrations (Lichtner et al., 2018a), and in the presence of remifentanyl (Lichtner et al., 2018b), but the cortical functional connectivity networks usually involved in nociception are reconfigured (Lichtner et al., 2018a).

### Halogenated Vapors

Similarly to propofol, the inhaled halogenated vapors such as sevoflurane have GABAergic properties in addition to other biochemical targets such as potassium channels (Bonhomme et al., 2011). They break down functional connectivity in higher-order resting-state large-scale networks such as the default-mode network, and the salience network, as well as the thalamo-cortical connectivity within them (Palanca et al., 2015), with a preservation of connectivity within sensory networks (Ranft et al., 2016). The fronto-parietal anterior to posterior effective connectivity is also reduced by sevoflurane at anesthetic doses (Lee et al., 2013; Ranft et al., 2016). Halogenates put networks aside from criticality (Lee et al., 2018), and limit the repertoire of possible network configurations (Cavanna et al., 2018; Uhrig et al., 2018), possibly through an augmentation of temporal persistence in neuronal oscillation amplitude (Thiery et al., 2018). The intermediate strength spatio-temporal patterns of functional connectivity are disrupted within and between networks, while higher strength patterns are preserved within networks (Kafashan et al., 2016), and fewer transitions in between-networks connectivity patterns occur (Golkowski et al., 2019).

### Ketamine

Among anesthetic agents, ketamine can be considered as the black sheep, because it induces distinct behavioral and functional changes as compared to other agents. The N-methyl-D-aspartate (NMDA) glutamate receptor antagonist ketamine produces a dissociative state with disconnected consciousness, through an isolation of the individual from the environment, and the emergence of intense dreaming with hallucinations and distorted self-perceptions. Ketamine increases functional connectivity globally (Driesen et al., 2013b). This occurs through a reconfiguration of between-region communication that becomes ineffective for some networks. Indeed, ketamine disrupts functional connectivity in all higher-order networks but the executive control network (Bonhomme et al., 2016). Also, connectivity is further preserved in sensory networks. A disruption of the fronto-parietal effective connectivity, and particularly the connectivity going from the anterior to the



**TABLE 2 |** Summary of the known effects of anesthetic agents on brain function.

	Functional connectivity	Effective connectivity	Topological properties	Evoked responses—sensory processing	Spatio-temporal dynamics
Propofol	Disruption of within- and between-network functional connectivity in large-scale brain networks (particularly fronto-parietal connectivity; Boveroux et al., 2010)	Reduced amplitude and complexity of long-distance cortical communication (Gómez et al., 2013; Sarasso et al., 2015)	Increased local efficiency (parietal regions; Kim et al., 2016; Lee et al., 2017)	Generalized alteration in short-latency evoked electrocorticographic responses to auditory novelty within higher-order cortical areas, outside the auditory cortex (Nourski et al., 2018)	Alteration of dynamics and directionality of effective connectivity (Lee et al., 2009; Untergerhrer et al., 2014; Sanders et al., 2018)
	Suppression of the complexity of regions sparsely connected with large-scale brain networks (Pappas et al., 2019)	Disruption of effective connectivity in large-scale brain networks (Lee et al., 2009; Boly et al., 2012; Lee et al., 2018; Untergerhrer et al., 2014; Guldenmund et al., 2016; Sanders et al., 2018)	Fewer small-world properties (Barttfeld et al., 2015)	Suppression of long-latency responses to novelty (Nourski et al., 2018)	Reduced complexity and randomness of the electroencephalographic signal (Wang et al., 2017; Darracq et al., 2018b)
	Disruption of thalamo-cortical connectivity within higher-order networks (Boveroux et al., 2010)	Alteration of effective connectivity in lower-order sensory networks (Gómez et al., 2013)	Limitation of connectivity configuration repertoire (Barttfeld et al., 2015; Cavanna et al., 2018)	Reconfiguration of cortical functional connectivity networks involved in nociception, despite activation of spinal cord and cortex by noxious stimulation (Lichtner et al., 2018a,b)	Synchronization of local activity (Huang et al., 2018)
	Preservation of connectivity in lower-order sensory networks (Boveroux et al., 2010)		Traffic of information constrained to inflexible patterns (Mashour, 2018; Uhrig et al., 2018)		Prolongation of long-distance communication timescales (Gómez et al., 2013; Huang et al., 2018)
Halogenated vapors	Breakdown of functional connectivity in higher-order resting-state consciousness networks (Palanca et al., 2015)	Disruption of fronto-parietal anterior to posterior effective connectivity (Lee et al., 2013)	Remoteness from criticality, with preserved scale-free organization (Liu et al., 2014; Tagliazucchi et al., 2016; Alonso et al., 2019)	No information	Augmentation of temporal persistence in neuronal oscillation amplitude (Thierry et al., 2018)
	Disruption of thalamo-cortical connectivity within higher-order networks (Palanca et al., 2015)		Disturbance of posterior parietal hub activity (Lee et al., 2013)		Disruption of intermediate strength spatio-temporal patterns of functional connectivity within and between consciousness networks (Kafashan et al., 2016)
	Preservation of connectivity in lower-order sensory networks (Ranft et al., 2016)		Limitation of connectivity configuration repertoire (Cavanna et al., 2018; Uhrig et al., 2018)		Preservation of higher strength spatio-temporal patterns within networks (Kafashan et al., 2016)
			Remoteness from criticality (Lee et al., 2018)		

(Continued)

**TABLE 2 |** (Continued)

	Functional connectivity	Effective connectivity	Topological properties	Evoked responses—sensory processing	Spatio-temporal dynamics
Ketamine	<p>Global increase in functional connectivity, with network reorganization (Driesen et al., 2013b)</p> <p>Disruption of functional connectivity in all higher-order consciousness networks but not in the executive control network (Bonhomme et al., 2016)</p> <p>Preservation of functional connectivity in sensory networks (Bonhomme et al., 2016)</p> <p>Long-term effect on the interactions between the default mode network and networks involved in depression? and restoration of the abnormal connectivity of depressed patients (Li et al., 2018; Vutskits, 2018)</p> <p>Transient effect on working memory network (Driesen et al., 2013a)</p>	<p>Disruption of fronto-parietal anterior to posterior effective connectivity (Lee et al., 2013; Vlisides et al., 2017)</p> <p>Reduced alpha power in the precuneus and temporo-parietal junction (possibly related to disconnected consciousness; Vlisides et al., 2018; Darraçq et al., 2018a)</p>	No information	TMS-evoked communication complexity close to the waking state (Sarasso et al., 2015)	No information
Dexmedetomidine	<p>Reduced within-network and thalamic connectivity in higher-order consciousness networks (Guldenmund et al., 2017)</p> <p>Preservation of lower-order sensory networks functional connectivity (Guldenmund et al., 2017)</p> <p>Better preservation of functional connectivity between thalamus, medial anterior cingulate cortex, and mesopontine area as compared to sleep and propofol unresponsiveness (Guldenmund et al., 2017)</p>	No information	<p>Reduced local and global large-scale network efficiency (Hashmi et al., 2017)</p> <p>Reduced large-scale network connectivity strength (Hashmi et al., 2017)</p> <p>No impairment in node degree (Hashmi et al., 2017)</p>	No information	No information
Benzodiazepines	<p>Disruption of higher-order consciousness networks (Greicius et al., 2008; Liang et al., 2018)</p> <p>Preservation of lower-order sensory networks (Frolich et al., 2017)</p>	<p>Disruption of effective connectivity in large-scale brain networks (Greicius et al., 2008; Ferrarelli et al., 2010; Liang et al., 2018)</p>	No information	<p>Reduced auditory cortex activation by sounds (Frolich et al., 2017)</p> <p>Reduced duration and propagation of evoked TMS cortical response (Ferrarelli et al., 2010)</p>	No information
Xenon	No information	No information	Remoteness from criticality (Colombo et al., 2019)	<p>TMS-evoked high amplitude slow waves with low complexity (Sarasso et al., 2015)</p>	Slowing down and smoothing of the temporal profile of the EEG signal (Colombo et al., 2019)

posterior part of the brain, has been observed during ketamine anesthesia (Lee et al., 2013; Vlisides et al., 2017). It induces a degree of TMS-evoked communication complexity that is close to the waking state and is very different from other agents such as propofol or xenon in that respect (Sarasso et al., 2015). Indeed, ketamine increases the randomness of the EEG signal (Wang et al., 2017). Reduced alpha power in the precuneus and temporal-parietal junction, both regions involved in multisensory integration and body representation, has been proposed as a mechanism for the ketamine-induced dissociative altered consciousness state and disconnected consciousness (Darracq et al., 2018a; Vlisides et al., 2018). Effects of ketamine on other functional systems have also been evidenced, such as its long-term effect on the interactions between the default-mode network and other networks involved in depression pathophysiology (Li et al., 2018), the restoration of abnormal connectivity observed in depressed patients (Scheidegger et al., 2012; Vutsits, 2018), and a transient effect on a network involved in working memory (Driesen et al., 2013a).

### Dexmedetomidine

At first glance, the brain functional connectivity profile of the alteration of consciousness induced by the  $\alpha_2$ -adrenoceptor agonist dexmedetomidine appears similar to the one induced by physiological sleep and propofol, with a reduced within-network and thalamic connectivity in the higher-order consciousness networks, and a preservation of lower-order sensory networks (Guldenmund et al., 2017). However, dexmedetomidine better preserves functional connectivity between the thalamus, the anterior cingulate cortex, and the mesopontine area than sleep or propofol sedation (Guldenmund et al., 2017), all regions that are parts of the salience network. This may be in relation with the ability of dexmedetomidine to induce a state where the subject retains the capacity of rapidly recovering oriented responsiveness to external stimulation. From a topological point of view, dexmedetomidine reduces local and global large-scale network efficiency and connectivity strength, without impairing node degree (Hashmi et al., 2017).

### Benzodiazepines

Similarly to propofol, benzodiazepines, which are potent GABA receptor ligands, are known to preserve functional connectivity in lower-order sensory networks despite reduced direct auditory cortex activation by sounds (Frolich et al., 2017), but not in the higher-order consciousness networks (Greicius et al., 2008; Liang et al., 2015). They also reduce the duration and propagation of evoked TMS cortical responses (Ferrarelli et al., 2010). A long term and chronic administration of diazepam increase functional connectivity in areas of emotional processing (Pflanz et al., 2015).

### Xenon

Brain functional studies using the noble gas xenon are still scarce. This agent has anti-NMDA properties and interferes with potassium channels (Bonhomme et al., 2011). It reduces the activity of specific brain regions including the orbito- and mesiofrontal cortex, cingulate gyrus, thalamus, hippocampus and bilateral cerebellum (Rex et al., 2008). Similarly to

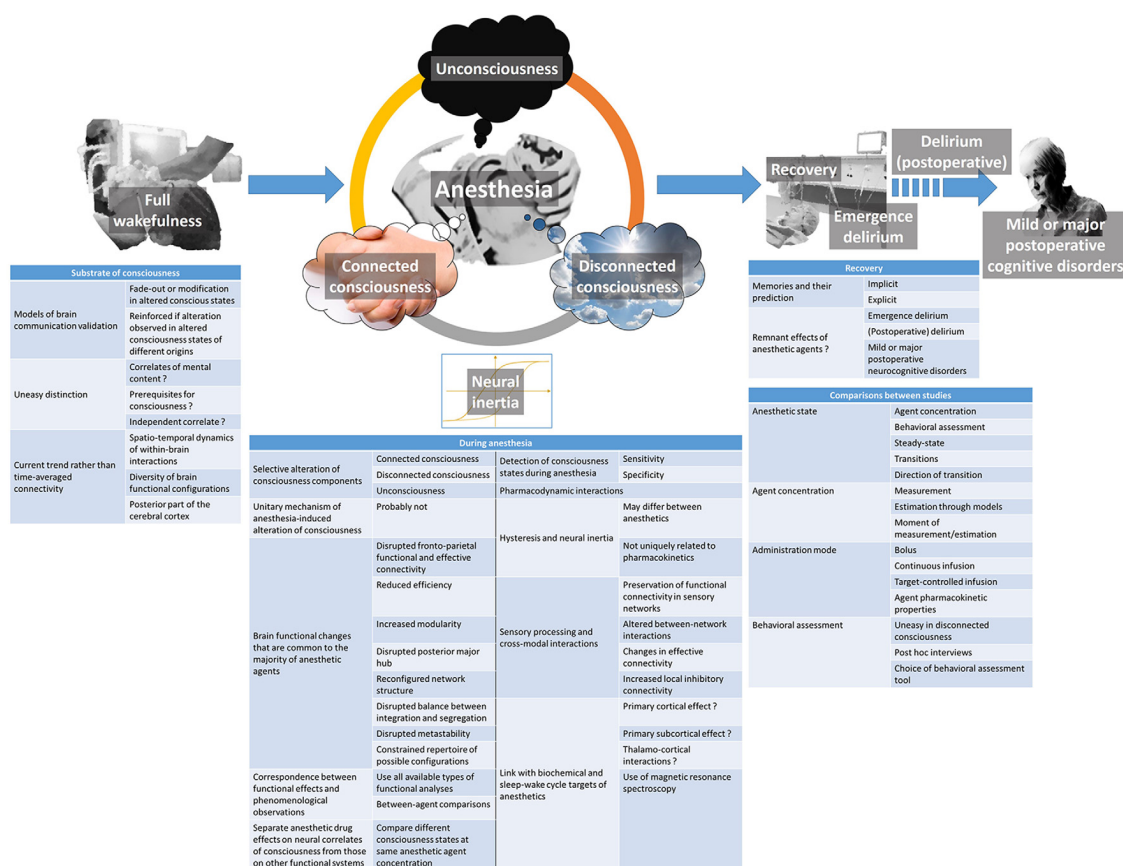
propofol, it slows down and smooths the temporal profile of the EEG signal, and slides the brain state away from criticality (Colombo et al., 2019). TMS-evoked cortical responses under xenon anesthesia correspond to high amplitude slow waves with low complexity as compared to the wake state (Sarasso et al., 2015).

## EMERGING ISSUES

Despite huge progress in unraveling the modalities of within-brain interactions, their implications in consciousness generation, and defining concepts that were not even imagined a decade ago, each discovery in the domain of consciousness physiology and brain effects of anesthetic agents leads to new questions. The ones of relevance for the understanding of anesthetic action, and whose resolution will allow making progress in the understanding of consciousness itself, are discussed hereafter (Figure 2).

### Substrate of Consciousness and Its Alteration

Different models of within-brain communication were proposed as being witnesses of conscious processing by the brain following their fade-out or modification during altered states of consciousness. The strength of certainty when proposing a model is always reinforced when similar modifications are observed during alterations of consciousness of different origins (Lee et al., 2018). However, making the distinction between the effective neural correlates of consciousness (those responsible for the mental content experience), the elements that may support consciousness but are not mandatory (such as the sub-cortical arousal systems), and the neural systems whose alteration is simply a consequence or independent correlate of unconsciousness is not always easy (Boly et al., 2017; Mashour and Hudetz, 2018). Some authors distinguish content-specific neural correlates of consciousness (related to a specific mental content but not necessarily involved in another one), and full neural correlates of consciousness (mandatory for a mental content to be present; Boly et al., 2017). Based on lesion data, electrical or magnetic stimulation data, and functional brain imaging data, they argue that both content-specific and full neural correlates of consciousness are mainly located in the posterior part of the brain, encompassing the parietal, occipital, and lateral temporal lobes (Boly et al., 2017). This view is opposed to the previously prevailing one that was attributing a major role to the prefrontal cortex in this respect. It seems, however, that the prefrontal cortex is important at regulating the level of consciousness, through its privileged reciprocal interactions with subcortical arousal systems, in addition to attention, working memory, verbal and motor report processes (Pal et al., 2018). Information integration would occur through between-brain regions correlation in activity or functional connectivity (Koch et al., 2016), while information differentiation would be linked to the spatio-temporal dynamics of within-brain interactions and the diversity of brain functional configurations (Mashour and Hudetz, 2018; Demertzi et al., 2019). During anesthesia, however, one may not consider



**FIGURE 2 |** Summary of the currently emerging issues regarding the brain effects of anesthetic agents and their relationship with the postulated neural mechanisms of consciousness.

consciousness as a whole, insofar as specific aspects of it can be altered separately.

Further, matching up the evidence coming from different study paradigms and consciousness alterations of different origins, and finding parallelisms in the modification of functional witnesses of within-brain interaction models according to a specified component of consciousness will certainly lead to progress in characterizing its neural correlates with precision. By precisely and reversibly modulating consciousness, anesthesia will continue to play a major role in this progress. However, when analyzing anesthesia studies, researchers should pay attention to not mixing up the specific anesthetic drug effects on these neural correlates and their non-specific effects on neural functional assemblies that sustain other brain functions. Some study designs may help to overcome the difficulty when they involve a modulation of the consciousness state without changing the anesthetic agent concentration. As suggested by Scheinin et al. (2018), this can be achieved using external stimulation or an additional pharmacological agent that promotes cortical arousal through the strengthening of some subcortical arousal neurotransmission systems. External stimulation is more likely to provoke a change in the consciousness state when sedation is provided by an  $\alpha_2$ -adrenoceptor agonist like

dexmedetomidine, but this can also occur under propofol sedation (Scheinin et al., 2018), or during routine general anesthesia when noxious stimulation elicits an episode of connected consciousness (Sanders et al., 2017). Similarly, physostigmine may change the level of arousal through an increase in cholinergic neurotransmission (Xie et al., 2011). When comparing the different brain states obtained at the same anesthetic agent concentration, it is important to precisely define their phenomenological characteristics, in order to pertinently correlate the functional changes with the behavioral ones.

## Specific Consciousness States of Anesthesia and the Associated Memory Processes

Dreaming and/or connectedness can occur frequently in apparently and behaviorally unresponsive subjects during anesthesia (Radek et al., 2018). The ability to sensitively and specifically distinguish between the different possible consciousness states of anesthesia, namely unconsciousness, disconnected consciousness, and connected consciousness, in addition to the ability of predicting implicit or explicit recall would be ideal. One must admit that this is currently not the



case with classical EEG analyses and commercially available depth of anesthesia monitors, as shown in a study by Gaskell et al. (2017). In this isolated forearm technique study, only the alpha-slow wave phase-amplitude coupling in the EEG was able to discriminate between patients with connected consciousness and those not responding to command. The first step before achieving such ability is to develop an understanding of the involved mechanisms, so as to define recordable correlates. Up to now, the memory processes of anesthesia have been poorly studied from a functional point of view. For example, a low dose of sevoflurane blocks emotional memory by reducing functional connectivity between the amygdala and the hippocampus (Alkire et al., 2008), and ketamine reversibly interferes with the circuits of working memory (Driesen et al., 2013a). Brief episodes of connected consciousness following intense noxious events during anesthesia are rarely—if not never—followed by explicit recall, but could potentially lead to implicit memories (Sanders et al., 2017). The proportion of true unexpected awareness episodes with explicit recall over the total number of detected and undetected connected consciousness episodes is currently not known. Nevertheless, very few elements are available to characterize the functional status of the brain at the very moment of connected consciousness, and its ability to initiate and consolidate memories at that time. Again, hints of mechanistic explanations to the anesthesia-induced consciousness states begin to be obtained, essentially by comparing altered consciousness states of different origins, but with the same phenomenological characteristics, and by correlating the observations with implicit and explicit memory data. Hence, when simultaneously measuring the effects of different anesthetic agents on brain activity, it was shown that a steeper decay-rate of the resting EEG power spectral density was characteristic of the absence of mental content in unresponsive subjects under propofol or xenon anesthesia. At the same time, a power spectral density decay similar to wakefulness was indicative of a mental content in unresponsive subjects under ketamine anesthesia (Colombo et al., 2019). TMS-evoked alpha power was found to be reduced during disconnected consciousness (dreaming) as compared to wakefulness, whenever it occurs during ketamine anesthesia or rapid-eye-movement sleep (Darracq et al., 2018a). Decreased TMS-evoked gamma power was observed in unconsciousness states induced by propofol anesthesia or non-rapid eye movement sleep (Darracq et al., 2018a) as compared to wakefulness and the other states. Slow-wave EEG activity saturation also seems to be pathognomonic of disconnectedness from the environment (Warnaby et al., 2017). All mechanistic information obtainable through this approach has not been gathered yet, but will certainly be in the future.

## Sensory Processing and Cross-modal Interactions During Anesthesia

Within the scope of defining the amount of environmental information that reaches the brain, and the extent of its processing during the different anesthetic consciousness states, the study of sensory processing and cross-modal interactions is important. An intriguing recurrent finding

of time-averaged connectivity studies is the preservation of functional connectivity within sensory networks, including thalamo-cortical connectivity, even during anesthesia-induced unresponsive states (Boveroux et al., 2010; Guldenmund et al., 2013; Bonhomme et al., 2016). However, sensory information handling by the brain appears to be altered during anesthesia through anomalies in between-network sensory modality crosstalk (Boveroux et al., 2010), changes in effective connectivity (Gómez et al., 2013), and increases in local inhibitory connectivity (Gómez et al., 2013). The use of more dynamic analysis paradigms and the study of evoked responses (Nourski et al., 2018) will certainly bring new information in this domain soon.

## Mixing Anesthetic Agents

As stated above, general anesthesia is rarely provided to patients using a single agent. Commonly, anesthetic agents with hypnotic properties are at least combined with opioids. It has been known for a long time that pharmacodynamic interactions between anesthetic agents exist, manifesting notably on routinely used processed EEG indexes of anesthetic depth (Bouillon et al., 2004). Very few information is currently available regarding those interactions at the level of within-brain communication, as well as information generation and handling by the brain. This topic merits further investigations using the existing functional brain imaging techniques.

## Direction-Dependent Mechanistic Differences in Anesthetic State Transitions

A phenomenon of hysteresis, sometimes referred to as neural inertia (Warnaby et al., 2017), occurs during anesthetic state transitions, meaning that the sequence of functional changes is distinct during forward and reverse transition from one consciousness state to the other. Although neural inertia typically refers to the pharmacokinetics of anesthetic agents, the observed differences are not uniquely related to them (Lee and Mashour, 2018b). This phenomenon may differ between anesthetic agents and is larger with the more potent ones (Kuizenga et al., 2018). Several EEG-based pharmacodynamic measures have been used to track the anesthesia recovery process (Purdon et al., 2013), and a phenomenon of hysteresis has been observed for the power spectrum, connectivity measures, structure and strength of networks (Kim et al., 2018), sensory-evoked EEG responses (Lewis et al., 2018), and slow-wave EEG activity saturation (Warnaby et al., 2017). The proposed involved mechanisms remain elusive. Specific study paradigms that precisely look at the dynamics of the transition loops and using different functional brain imaging modalities will help to elucidate them.

## Post-operative Remnant Effects of Anesthetic Agents

The effects of anesthetic agents on the brain do not cease when the syringe pump or the vaporizer are turned off. A series of perioperative neurocognitive disorders (NCD's) may be encountered in patients benefiting from surgery. The taxonomy of perioperative NCD's has recently been redefined by a group of experts (Evered et al., 2018). When considering the

NCD's that may have a link with anesthetic agents themselves, the following entities can be described. Immediately after anesthesia and the recovery of consciousness, cognition may remain altered for a limited period of time, a condition called emergence delirium. New NCD's that appear within the usually expected recovery period of 30 post-operative days are qualified as delirium or delayed neurocognitive recovery. The qualifier post-operative is added to delirium when new and persisting before discharge from the hospital, but not when appearing after discharge. After 30 days up to 12 months, persisting NCD's are named mild or major post-operative NCD's. They correspond to the previously used term of post-operative cognitive disorders (POCD). These entities may present over a wide variety of symptoms, ranging from an agitated and restless disorientation in time and space, to slight memory or other cognitive deficits (Mason et al., 2010). Very few information exists regarding their neural correlates and the underlying alterations in brain function. As compared to normal immediate post-operative recovery where functional and directed back-to-front connectivity is slowed down (Lee et al., 2013; van Dellen et al., 2014), hypoactive delirium is marked out by a less integrated network topology (Numan et al., 2017). Additional further studies are needed to fully understand the diversity of these cognitive disorders.

## Unitary Mechanism of Anesthesia-Induced Alteration of Consciousness

The question of a unitary mechanism of anesthesia-induced unconsciousness is a falsehood. Anesthetic agents produce different altered states of consciousness as a function of agent type and dose. This cannot occur through a common pathway. Several brain functional changes seem to be shared by the majority of anesthetic agents with hypnotic properties, including disrupted fronto-parietal functional and effective connectivity, reduced efficiency, increased modularity, disrupted posterior major hub, reconfigured network structure, disrupted balance between integration and segregation, disrupted metastability, and constrained repertoire of possible configurations (Lee and Mashour, 2018a). Among the observed functional effects, a good approach would be to establish a correspondence between them and phenomenological observations such as response to command when using the isolated forearm technique, *post hoc* report of dreaming with details, explicit recall of connected consciousness, standardized questionnaires for the detection of awareness, or specific paradigms for the detection of implicit memories. By submitting agent-related data to all available types of functional analyses, and making between-agent comparisons while taking into account the phenomenological characteristics of the consciousness states, substantial progress will be made in the understanding of the mechanism of each agent, and the functional correlate of consciousness components.

## Link Between Network Effects, Sleep-Wake Regulation, and Biochemical Targets

Anesthetic agents with hypnotic properties pertain to highly variable chemical families and have differing biochemical targets. They have also been shown to interfere with several subcortical

neurotransmission systems that are involved in the regulation of the sleep-wake cycle (Bonhomme et al., 2011), although anesthesia is clearly distinct from physiological sleep (Akeju and Brown, 2017). As a corollary, the crux of the matter remains to link those observations with the evidenced network effects. Some postulate a primary cortical effect, particularly for those agents mainly promoting the inhibitory GABA neurotransmission (Brown et al., 2011). Others favor a primary subcortical site of action and cortical dysfunction as mainly originating from thalamo-cortical interaction changes (Hutt et al., 2018). Dexmedetomidine, whose main target is the subcortical noradrenergic system, is an exception (Nelson et al., 2003). Nevertheless, each agent probably has its own sequential scheme of action, primarily cortical or subcortical, and we are still far from having drawn the complete picture for each of them. Magnetic resonance spectroscopy studies looking at *in vivo* neurotransmission will probably offer new insights soon (Abdallah et al., 2018).

## Dose, Administration Mode, Anesthetic State, and Comparisons

The published studies in the domain of anesthetic drug effects on brain function considerably vary in their design, sometimes rendering the comparison between attained sedation level (when comparing different anesthetic agents), anesthetic doses, and studied anesthetic states hazardous. The phenomenological assessment of behavioral changes can sometimes be uneasy to perform (like during disconnected consciousness) or must be made through *post hoc* interviews. Behavioral assessment tools may also differ between studies. In addition, the route and mode of administration, which depend on the nature (intravenous or inhaled) and pharmacokinetic properties of the anesthetic agent, may introduce confounders. For example, a single intravenous bolus produces a continuously evolving alteration of consciousness, while a computer-controlled target infusion allows steady-state conditions of recording. Looking at steady-state anesthetic conditions does not provide the same information as looking at transitions from one state to the other. As mentioned above, the direction of the transition is also important. Precisely measuring or estimating the attained plasma or effect-site concentration of the anesthetic agent at the moment of recording is not always possible or not always performed. Indeed, due to inter-individual variability, the same anesthetic agent concentration may produce a different effect from one subject to the other, and some authors prefer defining the studied anesthetic state behaviorally rather than in terms of attained concentration. As a consequence, and given the high variability in study protocols regarding dose, mode of administration, and achieved anesthetic states, it is not easy to have a clear idea of the exact dose-response relationships, either in terms of brain function modifications and correspondences with phenomenological observations. A continuum in the observed effects from light to very deep sedation is not necessarily a reality, these relationships may be different from one anesthetic agent to the other, and not all possible levels of sedation have been studied for all anesthetic agents. All types of knowledge derived from the published studies are informative, but caution

should be paid to not drawing erroneous conclusions when comparing them.

## CONCLUSION

The relationship between general anesthesia, brain function, and consciousness mechanisms is complex. As a matter of fact, anesthetic agents do not blunt out brain function globally but exert specific and dose-dependent effects on brain systems that sustain internal consciousness and perception of the environment. Each agent has its own mechanism of action, and dose-dependently induces distinct phenomenological altered consciousness states. Answering the questions that have recently emerged following recent discoveries on anesthetic brain effects will probably permit new insights into the specific diagnoses of anesthesia-induced altered states of consciousness, and into the understanding of the different aspects of consciousness itself. In that respect, general anesthesia can be considered as a flexible probe to explore consciousness.

## AUTHOR CONTRIBUTIONS

VB and OG wrote the first draft of this article, the other authors equally participated in the revision and editing of the manuscript.

## FUNDING

The writing of this article was financially supported by the University Department of Anesthesia and Intensive Care Medicine, Centre Hospitalier Régional de la Citadelle (CHR Citadelle), Liege, Belgium; the Department of Anesthesia

and Intensive Care Medicine, University Hospital Center of Liège (CHU Liège), Belgium; and the GIGA-Consciousness Thematic Unit, GIGA-Research, Liege University, Liege, Belgium. The studies performed by our group were funded by University Hospital of Liege; the Belgian National Funds for Scientific Research (FRS-FNRS); the European Union's Horizon 2020 Framework Program for Research and Innovation under the Specific Grant Agreement No. 720270 (Human Brain Project SGA1) and No. 785907 (Human Brain Project SGA2); the Luminous project (EU-H2020-fetopen-ga686764); the DOCMA project (EU-H2020-MSCA-RISE-778234); the Marie Skłodowska-Curie Actions (H2020-MSCA-IF-2016-ADOC-752686); the BIAL Foundation; the French Speaking Community Concerted Research Action (ARC 12-17/01); the James McDonnell Foundation; the Mind Science Foundation; the IAP research network P7/06 of the Belgian Government (Belgian Science Policy); the European Commission; the Public Utility Foundation "Université Européenne du Travail"; the "Fondazione Europea di Ricerca Biomedica"; Orion Pharma; the Society for Anesthesia and Resuscitation of Belgium; the Benoit Foundation; the Belgian Cancer Foundation (2017-064). SKL is a FRS-FNRS research fellow, AD is an associate researcher at FRS-FNRS and SL is research director at FRS-FNRS. The funders had no role in any study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## ACKNOWLEDGMENTS

We thank all the patients and volunteers who participated in our studies.

## REFERENCES

- Abdallah, C. G., De Feyter, H. M., Averill, L. A., Jiang, L., Averill, C. L., Chowdhury, G. M. L., et al. (2018). The effects of ketamine on prefrontal glutamate neurotransmission in healthy and depressed subjects. *Neuropsychopharmacology* 43, 2154–2160. doi: 10.1038/s41386-018-0136-3
- Agarwal, S., Stamatakis, E. A., Geva, S., and Warburton, E. A. (2016). Dominant hemisphere functional networks compensate for structural connectivity loss to preserve phonological retrieval with aging. *Brain Behav.* 6:e00495. doi: 10.1002/brb3.495
- Akeju, O., and Brown, E. N. (2017). Neural oscillations demonstrate that general anesthesia and sedative states are neurophysiologically distinct from sleep. *Curr. Opin. Neurobiol.* 44, 178–185. doi: 10.1016/j.conb.2017.04.011
- Alkire, M. T., Gruver, R., Miller, J., McReynolds, J. R., Hahn, E. L., and Cahill, L. (2008). Neuroimaging analysis of an anesthetic gas that blocks human emotional memory. *Proc. Natl. Acad. Sci. U S A* 105, 1722–1727. doi: 10.1073/pnas.0711651105
- Alkire, M. T., Haier, R. J., Barker, S. J., Shah, N. K., Wu, J. C., and Kao, Y. J. (1995). Cerebral metabolism during propofol anesthesia in humans studied with positron emission tomography. *Anesthesiology* 82, 393–403. doi: 10.1097/0000542-199502000-00010
- Alonso, L. M., Solovey, G., Yanagawa, T., Proekt, A., Cecchi, G. A., and Magnasco, M. O. (2019). Single-trial classification of awareness state during anesthesia by measuring critical dynamics of global brain activity. *Sci. Rep.* 9:4927. doi: 10.1038/s41598-019-41345-4
- Aubinet, C., Murphy, L., Bahri, M. A., Larroque, S. K., Cassol, H., Annen, J., et al. (2018). Brain, behavior, and cognitive interplay in disorders of consciousness: a multiple case study. *Front. Neurol.* 9:665. doi: 10.3389/fneur.2018.00665
- Barttfeld, P., Uhrig, L., Sitt, J. D., Sigman, M., Jarraya, B., and Dehaene, S. (2015). Signature of consciousness in the dynamics of resting-state brain activity. *Proc. Natl. Acad. Sci. U S A* 112, 887–892. doi: 10.1073/pnas.1418031112
- Bayne, T., Hohwy, J., and Owen, A. M. (2016). Are there levels of consciousness? *Trends Cogn. Sci.* 20, 405–413. doi: 10.1016/j.tics.2016.03.009
- Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. U S A* 106, 1672–1677. doi: 10.1073/pnas.0809667106
- Bodart, O., Gosseries, O., Wannez, S., Thibaut, A., Annen, J., Boly, M., et al. (2017). Measures of metabolism and complexity in the brain of patients with disorders of consciousness. *Neuroimage Clin.* 14, 354–362. doi: 10.1016/j.nicl.2017.02.002
- Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., and Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *J. Neurosci.* 37, 9603–9613. doi: 10.1523/JNEUROSCI.3218-16.2017
- Boly, M., Moran, R., Murphy, M., Boveroux, P., Bruno, M.-A., Noirhomme, Q., et al. (2012). Connectivity changes underlying spectral EEG changes during propofol-induced loss of consciousness. *J. Neurosci.* 32, 7082–7090. doi: 10.1523/JNEUROSCI.3769-11.2012



- Bonhomme, V., Boveroux, P., Vanhaudenhuyse, A., Hans, P., Brichant, J. F., Jaquet, O., et al. (2011). Linking sleep and general anesthesia mechanisms: this is no walkover. *Acta Anaesthesiol. Belg.* 62, 161–171.
- Bonhomme, V., Fiset, P., Meuret, P., Backman, S., Plourde, G., Paus, T., et al. (2001). Propofol anesthesia and cerebral blood flow changes elicited by vibrotactile stimulation: a positron emission tomography study. *J. Neurophysiol.* 85, 1299–1308. doi: 10.1152/jn.2001.85.3.1299
- Bonhomme, V., Maquet, P., Phillips, C., Plenevaux, A., Hans, P., Luxen, A., et al. (2008). The effect of clonidine infusion on distribution of regional cerebral blood flow in volunteers. *Anesth. Analg.* 106, 899–909. doi: 10.1213/ane.0b013e3181619685
- Bonhomme, V., Vanhaudenhuyse, A., Demertzi, A., Bruno, M. A., Jaquet, O., Bahri, M. A., et al. (2016). Resting-state network-specific breakdown of functional connectivity during ketamine alteration of consciousness in volunteers. *Anesthesiology* 125, 873–888. doi: 10.1097/aln.0000000000001275
- Bouillon, T. W., Bruhn, J., Radulescu, L., Andresen, C., Shafer, T. J., Cohane, C., et al. (2004). Pharmacodynamic interaction between propofol and remifentanyl regarding hypnosis, tolerance of laryngoscopy, bispectral index, and electroencephalographic approximate entropy. *Anesthesiology* 100, 1353–1372. doi: 10.1097/0000542-200406000-00006
- Boveroux, P., Vanhaudenhuyse, A., Bruno, M.-A., Noirhomme, Q., Lauwick, S., Luxen, A., et al. (2010). Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. *Anesthesiology* 113, 1038–1053. doi: 10.1097/ALN.0b013e3181f697f5
- Brown, E. N., Purdon, P. L., and Van Dort, C. J. (2011). General anesthesia and altered states of arousal: a systems neuroscience analysis. *Annu. Rev. Neurosci.* 34, 601–628. doi: 10.1146/annurev-neuro-060909-153200
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* 5:198ra105. doi: 10.1126/scitranslmed.3006294
- Cavanna, F., Vilas, M. G., Palmucci, M., and Tagliazucchi, E. (2018). Dynamic functional connectivity and brain metastability during altered states of consciousness. *Neuroimage* 180, 383–395. doi: 10.1016/j.neuroimage.2017.09.065
- Colombo, M. A., Napolitani, M., Boly, M., Gosseries, O., Casarotto, S., Rosanova, M., et al. (2019). The spectral exponent of the resting EEG indexes the presence of consciousness during unresponsiveness induced by propofol, xenon, and ketamine. *Neuroimage* 189, 631–644. doi: 10.1016/j.neuroimage.2019.01.024
- Crone, J. S., Lutkenhoff, E. S., Bio, B. J., Laureys, S., and Monti, M. M. (2017). Testing proposed neuronal models of effective connectivity within the cortico-basal gangliathalamo-cortical loop during loss of consciousness. *Cereb. Cortex* 27, 2727–2738. doi: 10.1093/cercor/bhw112
- Darracq, M., Funk, C. M., Polyakov, D., Riedner, B., Gosseries, O., Nieminen, J. O., et al. (2018a). Evoked  $\alpha$  power is reduced in disconnected consciousness during sleep and anesthesia. *Sci. Rep.* 8:16664. doi: 10.1038/s41598-018-34957-9
- Darracq, M., Sleight, J., Banks, M. I., and Sanders, R. D. (2018b). Characterising the effect of propofol on complexity and stability in the EEG power spectrum. *Br. J. Anaesth.* 121, 1368–1369. doi: 10.1016/j.bja.2018.09.006
- Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Barttfeld, P., Raimondo, F., et al. (2019). Human consciousness is supported by dynamic patterns of brain signal coordination. *Sci. Adv.* 5:eaat7603. doi: 10.1126/sciadv.aat7603
- Driesen, N. R., McCarthy, G., Bhagwagar, Z., Bloch, M. H., Calhoun, V. D., D'Souza, D. C., et al. (2013a). The impact of NMDA receptor blockade on human working memory-related prefrontal function and connectivity. *Neuropsychopharmacology* 38, 2613–2622. doi: 10.1038/npp.2013.170
- Driesen, N. R., McCarthy, G., Bhagwagar, Z., Bloch, M., Calhoun, V., D'Souza, D. C., et al. (2013b). Relationship of resting brain hyperconnectivity and schizophrenia-like symptoms produced by the NMDA receptor antagonist ketamine in humans. *Mol. Psychiatry* 18, 1199–1204. doi: 10.1038/mp.2012.194
- Evered, L., Silbert, B., Knopman, D. S., Scott, D. A., DeKosky, S. T., Rasmussen, L. S., et al. (2018). Recommendations for the nomenclature of cognitive change associated with anaesthesia and surgery-2018. *Br. J. Anaesth.* 121, 1005–1012. doi: 10.1016/j.bja.2017.11.087
- Ferrarelli, F., Massimini, M., Sarasso, S., Casali, A., Riedner, B. A., Angelini, G., et al. (2010). Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proc. Natl. Acad. Sci. U S A* 107, 2681–2686. doi: 10.1073/pnas.0913008107
- Fiset, P., Paus, T., Daloze, T., Plourde, G., Meuret, P., Bonhomme, V., et al. (1999). Brain mechanisms of propofol-induced loss of consciousness in humans: a positron emission tomographic study. *J. Neurosci.* 19, 5506–5513. doi: 10.1523/JNEUROSCI.19-13-05506.1999
- Frolich, M. A., Banks, C., and Ness, T. J. (2017). The effect of sedation on cortical activation: a randomized study comparing the effects of sedation with midazolam, propofol, and dexmedetomidine on auditory processing. *Anesth. Analg.* 124, 1603–1610. doi: 10.1213/ane.0000000000002021
- Gómez, F., Phillips, C., Soddu, A., Boly, M., Boveroux, P., Vanhaudenhuyse, A., et al. (2013). Changes in effective connectivity by propofol sedation. *PLoS One* 8:e71370. doi: 10.1371/journal.pone.0071370
- Gao, W.-W., He, Y.-H., Liu, L., Yuan, Q., Wang, Y.-F., and Zhao, B. (2018). BIS monitoring on intraoperative awareness: a meta-analysis. *Curr. Med. Sci.* 38, 349–353. doi: 10.1007/s11596-018-1886-1
- Gaskell, A. L. L., Hight, D. F. F., Winders, J., Tran, G., Defresne, A., Bonhomme, V., et al. (2017). Frontal  $\alpha$ -delta EEG does not preclude volitional response during anaesthesia: prospective cohort study of the isolated forearm technique. *Br. J. Anaesth.* 119, 664–673. doi: 10.1093/bja/aex170
- Golkowski, D., Larroque, S. K., Vanhaudenhuyse, A., Plenevaux, A., Boly, M., Di Perri, C., et al. (2019). Changes in whole brain dynamics and connectivity patterns during sevoflurane- and propofol-induced unconsciousness identified by functional magnetic resonance imaging. *Anesthesiology* 130, 898–911. doi: 10.1097/aln.0000000000002704
- Greicius, M. D., Kiviniemi, V., Tervonen, O., Vainionpää, V., Alahuhta, S., Reiss, A. L., et al. (2008). Persistent default-mode network connectivity during light sedation. *Hum. Brain Mapp.* 29, 839–847. doi: 10.1002/hbm.20537
- Guldenmund, P., Demertzi, A., Boveroux, P., Boly, M., Vanhaudenhuyse, A., Bruno, M.-A., et al. (2013). Thalamus, brainstem and salience network connectivity changes during propofol-induced sedation and unconsciousness. *Brain Connect.* 3, 273–285. doi: 10.1089/brain.2012.0117
- Guldenmund, P., Gantner, I. S., Baquero, K., Das, T., Demertzi, A., Boveroux, P., et al. (2016). Propofol-induced frontal cortex disconnection: a study of resting-state networks, total brain connectivity and mean BOLD signal oscillation frequencies. *Brain Connect.* 6, 225–237. doi: 10.1089/brain.2015.0369
- Guldenmund, P., Vanhaudenhuyse, A., Sanders, R. D., Sleight, J., Bruno, M. A., Demertzi, A., et al. (2017). Brain functional connectivity differentiates dexmedetomidine from propofol and natural sleep. *Br. J. Anaesth.* 119, 674–684. doi: 10.1093/bja/aex257
- Hashmi, J. A., Loggia, M. L., Khan, S., Gao, L., Kim, J., Napadow, V., et al. (2017). Dexmedetomidine disrupts the local and global efficiencies of large-scale brain networks. *Anesthesiology* 126, 419–430. doi: 10.1097/ALN.0000000000001509
- Huang, Z., Liu, X., Mashour, G. A., and Hudetz, A. G. (2018). Timescales of intrinsic BOLD signal dynamics and functional connectivity in pharmacologic and neuropathologic states of unconsciousness. *J. Neurosci.* 38, 2304–2317. doi: 10.1523/JNEUROSCI.2545-17.2018
- Hutt, A., Lefebvre, J., Hight, D., and Sleight, J. (2018). Suppression of underlying neuronal fluctuations mediates EEG slowing during general anaesthesia. *Neuroimage* 179, 414–428. doi: 10.1016/j.neuroimage.2018.06.043
- Kafashan, M., Ching, S., and Palanca, B. J. A. (2016). Sevoflurane alters spatiotemporal functional connectivity motifs that link resting-state networks during wakefulness. *Front. Neural Circuits* 10:107. doi: 10.3389/fncir.2016.00107
- Kim, H., Moon, J.-Y., Mashour, G. A., and Lee, U. (2018). Mechanisms of hysteresis in human brain networks during transitions of consciousness and unconsciousness: theoretical principles and empirical evidence. *PLoS Comput. Biol.* 14:e1006424. doi: 10.1371/journal.pcbi.1006424
- Kim, M., Mashour, G. A., Moraes, S.-B., Vanini, G., Tarnal, V., Janke, E., et al. (2016). Functional and topological conditions for explosive synchronization develop in human brain networks with the onset of anesthetic-induced unconsciousness. *Front. Comput. Neurosci.* 10:1. doi: 10.3389/fncom.2016.00001



- Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22
- Kuizenga, M. H., Colin, P. J., Reyntjens, K. M. E. M., Touw, D. J., Nalbat, H., Knotnerus, F. H., et al. (2018). Test of neural inertia in humans during general anaesthesia. *Br. J. Anaesth.* 120, 525–536. doi: 10.1016/j.bja.2017.11.072
- Lee, H., Golkowski, D., Jordan, D., Berger, S., Ilg, R., Lee, J., et al. (2018). Relationship of critical dynamics, functional connectivity, and states of consciousness in large-scale human brain networks. *Neuroimage* 188, 228–238. doi: 10.1016/j.neuroimage.2018.12.011
- Lee, U., Kim, S., Noh, G. J., Choi, B. M., Hwang, E., and Mashour, G. A. (2009). The directionality and functional organization of frontoparietal connectivity during consciousness and anesthesia in humans. *Conscious. Cogn.* 18, 1069–1078. doi: 10.1016/j.concog.2009.04.004
- Lee, U., Ku, S., Noh, G., Baek, S., Choi, B., and Mashour, G. A. (2013). Disruption of frontal-parietal communication by ketamine, propofol, and sevoflurane. *Anesthesiology* 118, 1264–1275. doi: 10.1097/ALN.0b013e31829103f5
- Lee, H., Mashour, G. A., Noh, G.-J., Kim, S., and Lee, U. (2013). Reconfiguration of network hub structure after propofol-induced unconsciousness. *Anesthesiology* 119, 1347–1359. doi: 10.1097/aln.0b013e3182a8ec8c
- Lee, U., and Mashour, G. A. (2018a). Role of network science in the study of anesthetic state transitions. *Anesthesiology* 129, 1029–1044. doi: 10.1097/ALN.0000000000002228
- Lee, U., and Mashour, G. A. (2018b). Stochastic nature of neural inertia. *Br. J. Anaesth.* 121, 7–8. doi: 10.1016/j.bja.2018.04.018
- Lee, M., Sanders, R. D., Yeom, S. K., Won, D. O., Seo, K. S., Kim, H. J., et al. (2017). Network properties in transitions of consciousness during propofol-induced sedation. *Sci. Rep.* 7:16791. doi: 10.1038/s41598-017-15082-5
- Lewis, L. D., Piantoni, G., Peterfreund, R. A., Eskandar, E. N., Harrell, P. G., Akeju, O., et al. (2018). A transient cortical state with sleep-like sensory responses precedes emergence from general anesthesia in humans. *Elife* 7:e33250. doi: 10.7554/elife.33250
- Li, M., Woelfer, M., Colic, L., Safran, A., Chang, C., and Jochen, H. (2018). Default mode network connectivity change corresponds to ketamine's delayed glutamatergic effects. *Eur. Arch. Psychiatry Clin. Neurosci.* doi: 10.1007/s00406-018-0942-y [Epub ahead of print].
- Liang, Z., Huang, C., Li, Y., Hight, D. F., Voss, L. J., Sleight, J. W., et al. (2018). Emergence EEG pattern classification in sevoflurane anesthesia. *Physiol. Meas.* 39:045006. doi: 10.1088/1361-6579/aab4d0
- Liang, P., Zhang, H., Xu, Y., Jia, W., Zang, Y., and Li, K. (2015). Disruption of cortical integration during midazolam-induced light sedation. *Hum. Brain Mapp.* 36, 4247–4261. doi: 10.1002/hbm.22914
- Lichtner, G., Aukstulewicz, R., Kirilina, E., Velten, H., Mavrodis, D., Scheel, M., et al. (2018a). Effects of propofol anesthesia on the processing of noxious stimuli in the spinal cord and the brain. *Neuroimage* 172, 642–653. doi: 10.1016/j.neuroimage.2018.02.003
- Lichtner, G., Aukstulewicz, R., Velten, H., Mavrodis, D., Scheel, M., Blankenburg, F., et al. (2018b). Nociceptive activation in spinal cord and brain persists during deep general anaesthesia. *Br. J. Anaesth.* 121, 291–302. doi: 10.1016/j.bja.2018.03.031
- Linassi, F., Zanatta, P., Tellaroli, P., Ori, C., and Carron, M. (2018). Isolated forearm technique: a meta-analysis of connected consciousness during different general anaesthesia regimens. *Br. J. Anaesth.* 121, 198–209. doi: 10.1016/j.bja.2018.02.019
- Liu, X., Ward, B. D., Binder, J. R., Li, S. J., and Hudetz, A. G. (2014). Scale-free functional connectivity of the brain is maintained in anesthetized healthy participants but not in patients with unresponsive wakefulness syndrome. *PLoS One* 9:e92182. doi: 10.1371/journal.pone.0092182
- Marchant, N., Sanders, R., Sleight, J., Vanhaudenhuyse, A., Bruno, M. A., Brichant, J. F., et al. (2014). How electroencephalography serves the anesthesiologist. *Clin. EEG Neurosci.* 45, 22–32. doi: 10.1177/1550059413509801
- Mashour, G. A. (2018). Highways of the brain, traffic of the mind. *Anesthesiology* 129, 869–871. doi: 10.1097/aln.0000000000002385
- Mashour, G. A., and Hudetz, A. G. (2018). Neural correlates of unconsciousness in large-scale brain networks. *Trends Neurosci.* 41, 150–160. doi: 10.1016/j.tins.2018.01.003
- Mason, S. E., Noel-Storr, A., and Ritchie, C. W. (2010). The impact of general and regional anesthesia on the incidence of post-operative cognitive dysfunction and post-operative delirium: a systematic review with meta-analysis. *J. Alzheimers Dis.* 22, 67–79. doi: 10.3233/jad-2010-101086
- Muthukumaraswamy, S. D., and Liley, D. T. (2018). 1/F electrophysiological spectra in resting and drug-induced states can be explained by the dynamics of multiple oscillatory relaxation processes. *Neuroimage* 179, 582–595. doi: 10.1016/j.neuroimage.2018.06.068
- Nelson, L. E., Lu, J., Guo, T., Saper, C. B., Franks, N. P., and Maze, M. (2003). The  $\alpha$ 2-adrenoceptor agonist dexmedetomidine converges on an endogenous sleep-promoting pathway to exert its sedative effects. *Anesthesiology* 98, 428–436. doi: 10.1097/0000542-200302000-00024
- Nicolaou, N., and Georgiou, J. (2014). Neural network-based classification of anesthesia/awareness using granger causality features. *Clin. EEG Neurosci.* 45, 77–88. doi: 10.1177/1550059413486271
- Nourski, K. V., Steinschneider, M., Rhone, A. E., Kawasaki, H., and Howard, M. A. III, and Banks, M. I. (2018). Auditory predictive coding across awareness states under anesthesia: an intracranial electrophysiology study. *J. Neurosci.* 38, 8441–8452. doi: 10.1523/JNEUROSCI.0967-18.2018
- Numan, T., Slooter, A. J. C., van der Kooi, A. W., Hoekman, A. M. L., Suyker, W. J. L., Stam, C. J., et al. (2017). Functional connectivity and network analysis during hypoactive delirium and recovery from anesthesia. *Clin. Neurophysiol.* 128, 914–924. doi: 10.1016/j.clinph.2017.02.022
- Pal, D., Dean, J. G., Liu, T., Li, D., Watson, C. J., Hudetz, A. G., et al. (2018). Differential role of prefrontal and parietal cortices in controlling level of consciousness. *Curr. Biol.* 28, 2145.e5–2152.e5. doi: 10.1016/j.cub.2018.05.025
- Palanca, B. J. A., Mitra, A., Larson-Prior, L., Snyder, A. Z., Avidan, M. S., and Raichle, M. E. (2015). Resting-state functional magnetic resonance imaging correlates of sevoflurane-induced unconsciousness. *Anesthesiology* 123, 346–356. doi: 10.1097/aln.0000000000000731
- Pappas, I., Adapa, R. M., Menon, D. K., and Stamatakis, E. A. (2019). Brain network disintegration during sedation is mediated by the complexity of sparsely connected regions. *Neuroimage* 186, 221–233. doi: 10.1016/j.neuroimage.2018.10.078
- Pfanz, C. P., Pringle, A., Filippini, N., Warren, M., Gottwald, J., Cowen, P. J., et al. (2015). Effects of seven-day diazepam administration on resting-state functional connectivity in healthy volunteers: a randomized, double-blind study. *Psychopharmacology* 232, 2139–2147. doi: 10.1007/s00213-014-3844-3
- Plourde, G., and Boylan, J. F. (1991). The auditory steady state response during sufentanil anaesthesia. *Br. J. Anaesth.* 66, 683–691. doi: 10.1093/bja/66.6.683
- Purdon, P. L., Pierce, E. T., Mukamel, E. A., Prerau, M. J., Walsh, J. L., Wong, K. F. K., et al. (2013). Electroencephalogram signatures of loss and recovery of consciousness from propofol. *Proc. Natl. Acad. Sci. U S A* 110, E1142–E1151. doi: 10.1073/pnas.1221180110
- Radek, L., Kallionpää, R. E., Karvonen, M., Scheinin, A., Maksimov, A., Längsjö, J., et al. (2018). Dreaming and awareness during dexmedetomidine- and propofol-induced unresponsiveness. *Br. J. Anaesth.* 121, 260–269. doi: 10.1016/j.bja.2018.03.014
- Ranft, A., Golkowski, D., Kiel, T., Riedl, V., Kohl, P., Rohrer, G., et al. (2016). Neural correlates of sevoflurane-induced unconsciousness identified by simultaneous functional magnetic resonance imaging and electroencephalography. *Anesthesiology* 125, 861–872. doi: 10.1097/ALN.0000000000001322
- Rex, S., Meyer, P. T., Baumert, J.-H., Rossaint, R., Fries, M., Bull, U., et al. (2008). Positron emission tomography study of regional cerebral blood flow and flow-metabolism coupling during general anaesthesia with xenon in humans. *Br. J. Anaesth.* 100, 667–675. doi: 10.1093/bja/aen036
- Ribeiro de Paula, D., Ziegler, E., Abeyasinghe, P. M., Das, T. K., Cavaliere, C., Aiello, M., et al. (2017). A method for independent component graph analysis of resting-state fMRI. *Brain Behav.* 7:e00626. doi: 10.1002/brb3.626
- Rowley, P., Bonczyk, C., Gaskell, A., Absalom, A., Bonhomme, V., Coburn, M., et al. (2017). What do people expect of general anaesthesia? *Br. J. Anaesth.* 118, 486–488. doi: 10.1093/bja/aex040
- Sanders, R. D., Banks, M. I., Darracq, M., Moran, R., Sleight, J., Gosseries, O., et al. (2018). Propofol-induced unresponsiveness is associated with impaired feedforward connectivity in cortical hierarchy. *Br. J. Anaesth.* 121, 1084–1096. doi: 10.1016/j.bja.2018.07.006

- Sanders, R. D., Gaskell, A., Raz, A., Winders, J., Stevanovic, A., Rossaint, R., et al. (2017). Incidence of connected consciousness after tracheal intubation: a prospective, international, multicenter cohort study of the isolated forearm technique. *Anesthesiology* 126, 214–222. doi: 10.1097/ALN.0000000000001479
- Sanders, R. D., Tononi, G., Laureys, S., and Sleight, J. W. (2012). Unresponsiveness  $\neq$  unconsciousness. *Anesthesiology* 116, 946–959. doi: 10.1097/ALN.0b013e318249d0a7
- Sarasso, S., Boly, M., Napolitani, M., Gosseries, O., Charland-Verville, V., Casarotto, S., et al. (2015). Consciousness and complexity during unresponsiveness induced by propofol, xenon and ketamine. *Curr. Biol.* 25, 3099–3105. doi: 10.1016/j.cub.2015.10.014
- Scheidegger, M., Walter, M., Lehmann, M., Metzger, C., Grimm, S., Boeker, H., et al. (2012). Ketamine decreases resting state functional network connectivity in healthy subjects: implications for antidepressant drug action. *PLoS One* 7:e44799. doi: 10.1371/journal.pone.0044799
- Scheinin, H., Alkire, E. C., Scheinin, A., Alkire, M. T., Kantonen, O., and Langsjo, J. (2018). Using positron emission tomography in revealing the mystery of general anesthesia: study design challenges and opportunities. *Meth. Enzymol.* 603, 279–303. doi: 10.1016/bs.mie.2018.01.025
- Sleight, J., Warnaby, C., and Tracey, I. (2018). General anaesthesia as fragmentation of selfhood: insights from electroencephalography and neuroimaging. *Br. J. Anaesth.* 121, 233–240. doi: 10.1016/j.bja.2017.12.038
- Staquet, C., Vanhaudenhuyse, A., and Bonhomme, V. (2018). Aware beside an unconscious patient, not the inverse! On the necessity of knowing how anesthesia modulates consciousness. *Acta Anaesthesiol. Belg.* 69, 137–145.
- Tagliazucchi, E., Chialvo, D. R., Siniatchkin, M., Amico, E., Brichant, J.-F., Bonhomme, V., et al. (2016). Large-scale signatures of unconsciousness are consistent with a departure from critical dynamics. *J. R. Soc. Interface* 13:20151027. doi: 10.1098/rsif.2015.1027
- Thiery, T., Lajnef, T., Combrisson, E., Dehgan, A., Rainville, P., Mashour, G. A., et al. (2018). Long-range temporal correlations in the brain distinguish conscious wakefulness from induced unconsciousness. *Neuroimage* 179, 30–39. doi: 10.1016/j.neuroimage.2018.05.069
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.* 5:42. doi: 10.1186/1471-2202-5-42
- Uhl, R. R., Squires, K. C., Bruce, D. L., and Starr, A. (1980). Variations in visual evoked potentials under anesthesia. *Prog. Brain Res.* 54, 463–466. doi: 10.1016/s0079-6123(08)61662-3
- Uhrig, L., Sitt, J. D., Jacob, A., Tasserie, J., Barttfeld, P., Dupont, M., et al. (2018). Resting-state dynamics as a cortical signature of anesthesia in monkeys. *Anesthesiology* 129, 942–958. doi: 10.1097/ALN.0000000000002336
- Untergerhrer, G., Jordan, D., Kochs, E. F., Ilg, R., and Schneider, G. (2014). Fronto-parietal connectivity is a non-static phenomenon with characteristic changes during unconsciousness. *PLoS One* 9:e87498. doi: 10.1371/journal.pone.0087498
- van Dellen, E., van der Kooi, A. W., Numan, T., Koek, H. L., Klijn, F. A. M., Buijsrogge, M. P., et al. (2014). Decreased functional connectivity and disturbed directionality of information flow in the electroencephalography of intensive care unit patients with delirium after cardiac surgery. *Anesthesiology* 121, 328–335. doi: 10.1097/ALN.0000000000000329
- Vanhaudenhuyse, A., Demertzi, A., Schabus, M., Noirhomme, Q., Bredart, S., Boly, M., et al. (2011). Two distinct neuronal networks mediate the awareness of environment and of self. *J. Cogn. Neurosci.* 23, 570–578. doi: 10.1162/jocn.2010.21488
- Vlissides, P. E., Bel-Bahar, T., Lee, U. C., Li, D., Kim, H., Janke, E., et al. (2017). Neurophysiologic correlates of ketamine sedation and anesthesia: a high-density electroencephalography study in healthy volunteers. *Anesthesiology* 127, 58–69. doi: 10.1097/ALN.0000000000001671
- Vlissides, P. E., Bel-Bahar, T., Nelson, A., Chilton, K., Smith, E., Janke, E., et al. (2018). Subanaesthetic ketamine and altered states of consciousness in humans. *Br. J. Anaesth.* 121, 249–259. doi: 10.1016/j.bja.2018.03.011
- Vutsits, L. (2018). General anesthetics to treat major depressive disorder: clinical relevance and underlying mechanisms. *Anesth. Analg.* 126, 208–216. doi: 10.1213/ANE.0000000000002594
- Wang, J., Noh, G. J., Choi, B. M., Ku, S. W., Joo, P., Jung, W. S., et al. (2017). Suppressed neural complexity during ketamine- and propofol-induced unconsciousness. *Neurosci. Lett.* 653, 320–325. doi: 10.1016/j.neulet.2017.05.045
- Warnaby, C. E., Sleight, J. W., Hight, D., Jbabdi, S., and Tracey, I. (2017). Investigation of slow-wave activity saturation during surgical anesthesia reveals a signature of neural inertia in humans. *Anesthesiology* 127, 645–657. doi: 10.1097/ALN.0000000000001759
- Xie, G., Deschamps, A., Backman, S. B., Fiset, P., Chartrand, D., Dagher, A., et al. (2011). Critical involvement of the thalamus and precuneus during restoration of consciousness with physostigmine in humans during propofol anaesthesia: a positron emission tomography study. *Br. J. Anaesth.* 106, 548–557. doi: 10.1093/bja/aeq415

**Conflict of Interest Statement:** During the past 5 years, VB has received research grant from Orion Pharma and honoraria for consultancy from Medtronic.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Bonhomme, Staquet, Montupil, Defresne, Kirsch, Martial, Vanhaudenhuyse, Chatelle, Larroque, Raimondo, Demertzi, Bodart, Laureys and Gosseries. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Exploring the Phase-Locking Mechanisms Yielding Delayed and Anticipated Synchronization in Neuronal Circuits

Leonardo Dalla Porta<sup>1†</sup>, Fernanda S. Matias<sup>2†</sup>, Alfredo J. dos Santos<sup>3</sup>, Ana Alonso<sup>4</sup>, Pedro V. Carelli<sup>3</sup>, Mauro Copelli<sup>3</sup> and Claudio R. Mirasso<sup>4\*</sup>

<sup>1</sup> System Neuroscience Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain, <sup>2</sup> Instituto de Física, Universidade Federal de Alagoas, Maceió, Brazil, <sup>3</sup> Departamento de Física, Universidade Federal de Pernambuco, Recife, Brazil, <sup>4</sup> Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC, UIB-CSIC), Palma, Spain

## OPEN ACCESS

### Edited by:

Maurizio Mattia,  
Istituto Superiore di Sanità (ISS), Italy

### Reviewed by:

Simona Olmi,  
Italian National Research Council  
(CNR), Italy  
Gorka Zamora-López,  
Pompeu Fabra University, Spain

### \*Correspondence:

Claudio R. Mirasso  
claudio@ifisc.uib-csic.es

<sup>†</sup>These authors have contributed  
equally to this work

**Received:** 29 March 2019

**Accepted:** 05 August 2019

**Published:** 21 August 2019

### Citation:

Dalla Porta L, Matias FS, dos  
Santos AJ, Alonso A, Carelli PV,  
Copelli M and Mirasso CR (2019)  
Exploring the Phase-Locking  
Mechanisms Yielding Delayed and  
Anticipated Synchronization in  
Neuronal Circuits.  
Front. Syst. Neurosci. 13:41.  
doi: 10.3389/fnsys.2019.00041

Synchronization is one of the brain mechanisms allowing the coordination of neuronal activity required in many cognitive tasks. Anticipated Synchronization (AS) is a specific type of out-of-phase synchronization that occurs when two systems are unidirectionally coupled and, consequently, the information is transmitted from the sender to the receiver, but the receiver leads the sender in time. It has been shown that the primate cortex could operate in a regime of AS as part of normal neurocognitive function. However it is still unclear what is the mechanism that gives rise to anticipated synchronization in neuronal motifs. Here, we investigate the synchronization properties of cortical motifs on multiple scales and show that the internal dynamics of the receiver, which is related to its free running frequency in the uncoupled situation, is the main ingredient for AS to occur. For biologically plausible parameters, including excitation/inhibition balance, we found that the phase difference between the sender and the receiver decreases when the free running frequency of the receiver increases. As a consequence, the system switches from the usual delayed synchronization (DS) regime to an AS regime. We show that at three different scales, neuronal microcircuits, spiking neuronal populations and neural mass models, both the inhibitory loop and the external current acting on the receiver mediate the DS-AS transition for the sender-receiver configuration by changing the free running frequency of the receiver. Therefore, we propose that a faster internal dynamics of the receiver system is the main mechanism underlying anticipated synchronization in brain circuits.

**Keywords:** neuronal oscillations, neuronal dynamics, neuronal motifs, synchronization, anticipated synchronization

## 1. INTRODUCTION

Brain rhythms have been extensively studied and related to plenty of neurocognitive tasks in the last decades (Buzsáki, 2006). According to the communication through coherence hypothesis (Fries, 2005), neuronal oscillation locked at the appropriate phase may facilitate information transmission between brain regions. Despite the fact that the phase relations are associated to synaptic delays between distant regions, non-linear ingredients as inhibition and external noise acting locally

can also control the phase relation between coupled areas. When the oscillations of certain area A influence and lock those of another area B, it is expected that the phase between A and B (defined as  $\phi_B - \phi_A$ ) be positive [a regime we refer to as delayed synchronization (DS)]. However, a counterintuitive phase relation was observed between cortical regions in primates (Brovelli et al., 2004; Salazar et al., 2012). Under certain circumstances, a directional influence between two cortical areas is accompanied by a negative time delay (i.e., by a negative phase difference). This phenomenon has been explained by the concept of anticipated synchronization (AS) (Voss, 2000; Matias et al., 2014).

As proposed by Voss, two identical autonomous dynamical systems unidirectionally coupled in a sender-receiver configuration can exhibit anticipated synchronization if the receiver is subject to a delayed negative self-feedback:

$$\begin{aligned}\dot{\mathbf{S}} &= \mathbf{f}(\mathbf{S}(t)), \\ \dot{\mathbf{R}} &= \mathbf{f}(\mathbf{R}(t)) + K[\mathbf{S}(t) - \mathbf{R}(t - t_d)],\end{aligned}\quad (1)$$

where  $\mathbf{f}(\mathbf{S})$  is a vector function that describes the autonomous dynamical system,  $K$  is the coupling matrix and the delayed term  $\mathbf{R}(t - t_d)$  is the self-feedback (Voss, 2000). The solution  $\mathbf{R}(t) = \mathbf{S}(t + t_d)$  characterizes the regime of anticipated synchronization and has been verified in a variety of theoretical (Voss, 2000, 2001a,b, 2016, 2018; Masoller and Zanette, 2001; Hernández-García et al., 2002; Ciszak et al., 2003; Kostur et al., 2005; Sausedo-Solorio and Pisarchik, 2014) and experimental (Sivaprakasam et al., 2001; Tang and Liu, 2003; Ciszak et al., 2009; Stepp and Turvey, 2017) studies.

The AS regime has been reported in systems without the explicit delay term. For example, for a specific parameter mismatch between the sender and the receiver system that gives a first-order approximation to the delayed coupling (Corron et al., 2005). AS has also been reported in a chain consisting of a sender and two receivers with switching parameters (Pyragienė and Pyragas, 2015), between two Hodgkin-Huxley neurons with different depolarization parameters (Simonov et al., 2014) and in the presence of an inhibitory loop mediated by an interneuron with a free-running frequency greater than the others (Matias et al., 2017). It has also been shown that AS may appear between two neuron models directly coupled provided that the mean frequency of the free receiver is greater than the mean frequency of the sender with (Hayashi et al., 2016) and without the explicit time-delay (Pyragienė and Pyragas, 2013; Dima et al., 2018). AS has been verified in a system in which the delayed feedback has been replaced by a simple, low-order all-pass filter (Pyragienė and Pyragas, 2017). More recently, a novel theoretical viewpoint based on the mathematical object called canard, has been used to explain anticipation in excitable systems (Ersős et al., 2019).

In neuronal rhythms the relative phase between two coupled regions is an important characteristic of the dynamics since it can modulate the information flow of an unexpected stimuli (Barardi et al., 2014). Here we investigate the mechanisms underlying the transition from positive to negative phase locking (or equivalently the transition between the DS and AS regimes) on multiple scales. For synchronized systems, it is equivalent to

define the phase relation or the time delay between peaks of activity. We simulate three different motifs of unidirectionally coupled systems in which the negative delayed feedback of Equation (1) is replaced by a synaptic inhibitory loop. We extend previous results for three coupled neurons (Matias et al., 2011) and cortical-like populations (Matias et al., 2014), showing that the AS-DS transition can be mediated not only by the inhibitory synaptic conductance but also by the external stimulus at the receiver. We also show that a neural mass model, known to exhibit zero-lag synchronization (Gollob et al., 2014), can operate in the anticipated synchronization regime and the AS-DS transition can be mediated by the stimuli acting on the receiver as well by an inhibitory loop. Moreover, we show that when the sender and receiver are uncoupled the inhibitory loop and an external current acting at the receiver system change its internal dynamics which is reflected in its free running frequency. More important, we found that the phase difference between the sender and the receiver decreases when the free running frequency of the receiver increases. Therefore, we propose that for an excitation/inhibition balance and biologically plausible parameters a faster internal dynamics of the receiver as compared to the emitter is the mechanism underlying AS. We also suggest that the DS-AS transition studied here could be mediated by any parameter that turns the internal dynamics of the free-receiver faster (or equivalently the free-running frequency of the sender slower) and could also account for delay compensation in cortical systems.

## 2. MATERIALS AND METHODS

### 2.1. Microcircuit

The model for the 3-neurons motif is the one proposed by Matias et al. (2011). Neurons are described by the Hodgkin-Huxley model (Hodgkin and Huxley, 1952) composed by the currents  $I_{Na}$ ,  $I_K$  and  $I_L$ :

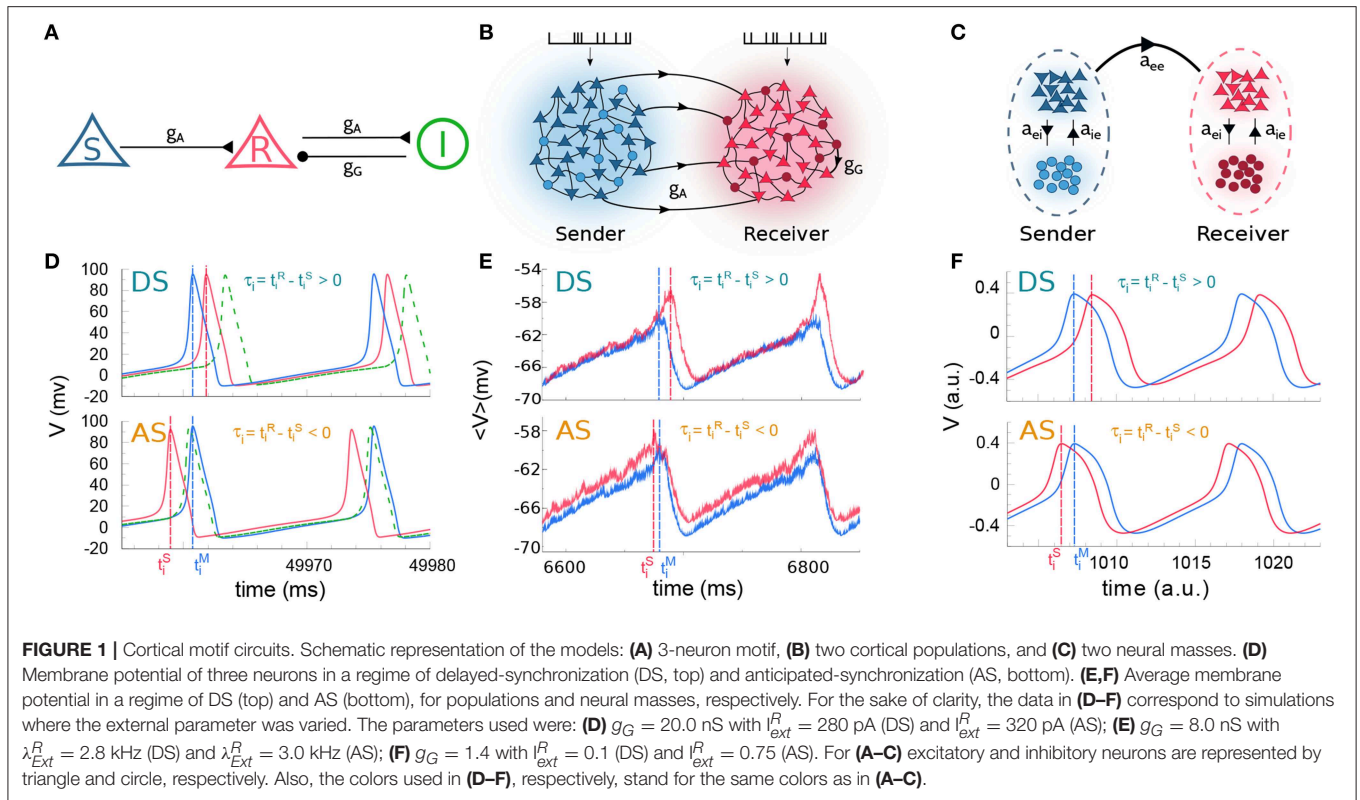
$$C_m \frac{dV}{dt} = I_{Na} + I_K + I_L + I_{ext} + I_{syn}, \quad (2)$$

where  $C_m = 9\pi\mu\text{F}$  is the membrane capacitance of a  $30 \times 30 \times \pi \mu\text{m}^2$  equipotential patch of membrane,  $I_{ext}$  is the external constant current that sets the neuron excitability and  $I_{syn}$  is the pre-synaptic current.

The ion channels follows the Hodgkin-Huxley formalism:  $dx/dt = \phi[\alpha_x(V)(1-x) - \beta_x(V)x]$ , being  $\phi = 1$  the temperature factor. The sodium current  $I_{Na} = g_{Na}m^3h(V_{Na} - V)$  has a maximal conductance  $g_{Na} = 1080\pi \text{ mS}$ , and rate constants  $\alpha_m(V) = 0.1(25 - V)/(e^{0.1(25-V)} - 1)$ ,  $\beta_m(V) = 4e^{-V/18}$ ,  $\alpha_h(V) = 0.07e^{-V/20}$  and  $\beta_h(V) = 1/(e^{0.1(30-V)} + 1)$ . The delayed-rectifier potassium current  $I_K = g_Kn^4(V_K - V)$  has a maximal conductance  $g_K = 324\pi \text{ mS}$  and rate constants  $\alpha_n(V) = 0.01(10 - V)/(e^{0.1(10-V)} - 1)$  and  $\beta_n(V) = 0.125e^{-V/80}$ . The leakage current  $I_L = g_L(V_L - V)$  has a maximal conductance  $g_L = 2.7\pi \text{ mS}$ . The reversal potentials are  $V_{Na} = 115 \text{ mV}$ ,  $V_K = -12 \text{ mV}$  and  $V_L = 10.6 \text{ mV}$ . In all the expressions above,  $V$  is measured in mV.

The synaptic current  $I_{syn} = g_{syn}r(V_{syn} - V)$  comprises the gating variable  $r$  following  $dr/dt = \alpha[T](1 - r) - \beta r$ , where  $\alpha$





and  $\beta$  are rate constants and  $[T](V_{pre}) = T_{max}/(1 + e^{(V_p - V_{pre})/K_p})$  is the neurotransmitter concentration in the synaptic cleft. In this model AMPA (A) and GABA<sub>A</sub> (G) are the excitatory and inhibitory synapses, respectively. The parameters concerning the synapses are:  $\alpha_A = 1.1$  (mM<sup>-1</sup>ms<sup>-1</sup>),  $\beta_A = 0.19$  (ms<sup>-1</sup>),  $\alpha_G = 5.0$  (mM<sup>-1</sup>ms<sup>-1</sup>),  $\beta_G = 0.30$  (ms<sup>-1</sup>),  $T_{max} = 1$  mM<sup>-1</sup>,  $K_p = 5$  mV,  $V_p = 62$  mV,  $V_{syn,AMPA} = 60$  mV and  $V_{syn,GABA} = -20$  mV. The three neurons are: sender, receiver, and interneuron. The sender projects excitatory synapses onto the receiver. The receiver projects excitatory and receives inhibitory synapses from the interneuron (see **Figure 1**). We kept  $g_A = 10.0$  nS,  $I_{ext}^S = 280.0$  pA and  $I_{ext}^{Inter} = 280.0$  pA fixed throughout all the simulations. The free parameters of this model are  $g_G$  and  $I_{ext}^R$ . When  $g_G^R$  was varied,  $I_{ext}^R = 280$  pA was kept fixed, whereas when  $I_{ext}^R$  was varied,  $g_G^R = 20.0$  nS was kept fixed.

## 2.2. Neuronal Populations

For two cortical populations we follow the ideas proposed by Matias et al. (2014). Each population, sender (S) and receiver (R), is composed of excitatory (80%) and inhibitory (20%) neurons, whose dynamics is described by the Izhikevich model (Izhikevich, 2003):

$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - u + \sum I_{syn}, \quad (3)$$

$$\frac{du}{dt} = a(bv - u), \quad (4)$$

where  $v$  and  $u$  stands for the membrane potential of the neuron and the membrane recovery variable (activation of K<sup>+</sup> and inactivation of Na<sup>+</sup> ionic currents), respectively.  $a, b, c$  and  $d$  are dimensionless parameters that account for the firing patterns heterogeneity which are randomly distributed accordingly to the neuron's nature. For excitatory neurons  $a = 0.02$ ,  $b = 0.20$ ,  $c = -65 + 15\sigma^2$  and  $d = 8 - 6\sigma^2$ , whereas for inhibitory neurons  $a = 0.02 + 0.08\sigma$ ,  $b = 0.25 - 0.05\sigma$ ,  $c = -65.0$  and  $d = 2.0$ .  $\sigma \in (0, 1)$  is a random variable. If a spike occurs, i.e.,  $v \geq -30$  mV,  $v$  is reset to  $c$  and  $u$  to  $u + d$ .

The synaptic transmissions are mediated by excitatory AMPA (A) and inhibitory GABA<sub>A</sub> (G). The pre-synaptic current is described as  $I_{syn} = -g_{syn}r(v - V_{syn})$ , where  $V_A = 0$  mV and  $V_G = -65$  mV.  $g_{syn}$  is the maximal conductance,  $g_A$  for excitatory and  $g_G$  for inhibitory synapses.  $r$  is the gating variable and follows a first-order kinetic dynamics:  $\tau_{syn}dr/dt = -r + D \sum_j \delta(t - t_j)$ , where  $\tau_A = 5.26$  ms,  $\tau_G = 5.60$  ms and the summation over  $j$  stands for the neighbor's pre-synaptic spikes at the previous time steps  $\{t_j\}$ .  $D$  is taken, without loss of generality, equal to 0.05.

The populations S and R are composed of 500 neurons each, among which 80% are pyramidal cells and 20% inhibitory interneurons. In the S population, each neuron receives 50 randomly chosen synapses from other neurons with excitatory conductances  $g_A^S = 0.5$  nS and inhibitory conductances  $g_G^S = 4$  nS, which remained fixed throughout the simulations. In the R population, each neuron receives 10 inhibitory synapses ( $g_G^{R,I} = 4$  nS for inhibitory neurons and  $g_G^R$  for excitatory

neurons) and 40 excitatory synapses ( $g_A^R = 0.5$  nS). For both populations, no autapses are allowed. The connectivity between S and R populations is such that all neurons within the R population receive 20 randomly chosen excitatory synapses from the S population ( $g_A^{SR} = 0.5$  nS, unless otherwise specified). Also, all neurons, within the S and R populations, are subject to an independent noisy spike train described by a Poisson distribution with rate  $\lambda$ . The input mimics excitatory synapses from neurons that are not included in the populations, with a maximal conductance  $g_A = 0.5$  nS. Without loss of generality, we assume for the S population  $\lambda^S = 2.4$  kHz. So, the free parameters for this model are  $g_G^R$  and  $\lambda_{ext}^R$ . When  $g_G^R$  was varied,  $\lambda_{ext}^R = 2.4$  kHz was kept fixed, whereas when  $\lambda_{ext}^R$  was varied,  $g_G^R = 8.0$  nS was kept fixed.

## 2.3. Neural Mass Models

The large-scale circuit model is the one used in Gollo et al. (2014). Briefly, the neural mass model (NMM) is composed by three state variables:  $V$  is the mean membrane potential of pyramidal neurons;  $Z$ , the mean membrane potential for interneurons; and  $W$  is the average number of open potassium ion channels. Here we made use of two ensembles  $i = S, R$ , namely Sender and Receiver. The equations for the dynamics are given by:

$$\begin{aligned} \frac{dV^i(t)}{dt} = & -\{g_{Ca} + (1.0 - C_{ji})r_{NMDA}a_{ee}Q_V^i(t) \\ & + C_{ji}r_{NMDA}a_{ee} < Q_V^j(t - \tau) >\}m_{Ca}(V^i(t) - V_{Ca}) \\ & -g_KW^i(t)(V^i(t) - V_K) - g_L(V^i(t) - V_L) \\ & -\{g_{Na}m_{Na} + a_{ee}(1.0 - C_{ji})Q_V^i(t) \\ & + C_{ji}a_{ee} < Q_V^j(t - \tau) >\}(V^i(t) - V_{Na}) \\ & -a_{ie}Z^i(t)Q_Z^i + a_{ne}I_{ext}^E, \end{aligned} \quad (5)$$

$$\frac{dZ^i(t)}{dt} = b(a_{ni}I_{ext}^I + a_{ei}V^i(t)Q_V^i(t)), \quad (6)$$

$$\frac{dW^i(t)}{dt} = \frac{\phi\{m_K - W^i(t)\}}{\tau_W}, \quad (7)$$

$$m_{ion} = 0.5\left[1 + \tanh\left(\frac{V^i(t) - T_{ion}}{\delta_{ion}}\right)\right], \quad (8)$$

$$Q_V^i(t) = 0.5Q_{Vmax}\left[1 + \tanh\left(\frac{V^i(t) - V_T}{\delta_V}\right)\right], \quad (9)$$

$$Q_Z^i(t) = 0.5Q_{Zmax}\left[1 + \tanh\left(\frac{Z^i(t) - Z_T}{\delta_Z}\right)\right], \quad (10)$$

where  $m_{ion}$  and  $Q_{V,Z}^i$  are the fraction of open channels and neuronal firing rates, respectively.

The parameters are:  $g_{Ca} = 1.1$ ,  $g_K = 2$ ,  $g_L = 0.5$ ,  $g_{Na} = 6.7$ ,  $r_{NMDA} = 0.25$ ,  $\phi = 0.7$ ,  $\tau_W = 1.0$ ,  $b = 0.1$ ,  $T_K = 0$ ,  $T_{Ca} = -0.01$ ,  $T_{Na} = 0.3$ ,  $\delta_K = 0.3$ ,  $\delta_{Na} = 0.15$ ,  $\delta_{Ca} = 0.15$ ,  $V_{Ca} = 1$ ,  $V_K = -0.7$ ,  $V_L = -0.5$ ,  $V_{Na} = 0.53$ ,  $V_T = Z_T = 0$ ,  $Q_{Vmax} = Z_{Vmax} = 0$ ,  $\delta_V = \delta_Z = 0.65$ ,  $a_{ei} = 2$ ,  $a_{ee} = 0.4$ ,  $a_{ne} = 1$ ,  $a_{ni} = 0.4$  and  $I_{ext}^I = 0.02$  (here the upper index stands for the inhibitory sub-population in both sender and receiver groups). All quantities are dimensionless. In the sender, the maximal conductance from interneurons to pyramidal neurons as well as

the external current in the excitatory neurons are kept constant,  $a_{ie}^S = 2.4$  and  $I_{ext}^S = 0.20$ , respectively. Thus,  $a_{ie}^R$  and  $I_{ext}^R$  are the free parameters of this model. The coupling ( $C_{ji}$ ) between the two neural masses are:  $C_{RS} = 0$  and  $C_{SR} = 0.4$ , thus guaranteeing an unidirectional *master-slave* configuration between S and R. When  $a_{ie}^R$  was varied  $I_{ext}^R = 0.3$  was kept fixed, whereas when  $I_{ext}^R$  was varied  $a_{ie}^R = 1.4$  was kept fixed.

## 2.4. Numerical Methods and Data Analysis

The model for 3-neurons-motif was implemented in a C code and simulated using a forth-order Runge-Kutta method with a time step of  $5 \times 10^{-3}$  ms. The equations for the neuronal populations and neural masses were implemented in a C++ code and simulated using the Euler method, with a time step of  $5 \times 10^{-2}$  ms for neuronal populations and of  $10^{-3}$  (arbitrary units) for neural masses. To compute the mean response of the membrane potential in NMMs we averaged over 10 realizations of the initial conditions while for the populations we averaged over 10 realizations of the external noise, network connectivity and neuron parameters.

The population membrane potential was estimated from the average value of the individual cell's membrane potential (a variable comparable to the local field potential (LFP) recorded in experiments). In order to smooth the noisy signal of the membrane potential we used a Butterworth low-pass filter of fourth order and cutoff frequency of 5 rad/s. From the filtered signal we extract the peak at times  $t_i$ . We then calculate the time delay in each cycle as  $\tau_i = t_i^R - t_i^S$  (see **Figure 1E**). Also, the main frequency ( $\omega$ ) of the neuronal population was obtained detecting the peak of the power spectrum computed via the Fast-Fourier Transform (FFT). For the 3-neurons-motif and also for NMMs, we compute the period in each cycle as the difference between consecutive peaks at times  $t_i$ , i.e.,  $T_i^* = t_{i+1}^* - t_i^*$  ( $*$  = R,S). The time delay is estimated as in the case of neuronal populations.

## 3. RESULTS

We investigated the phase locking characteristics, or equivalently in this case the synchronization characteristics, of a unidirectionally coupled system  $A \rightarrow B$ . Our main analysis assumes that the two dynamical nodes A and B are phase locked such that computing the phase difference is equivalent to compute the time difference between peaks. For this reason, from now on we talk about synchronization instead of phase locking. In a delayed synchronization condition, the time difference between the spikes in A and B is positive (A leads B), i.e., the pulse in B occurs after the pulse in A. In the less intuitive case of anticipated synchronization, the pulse in B precedes that in A, yielding a negative time difference.

### 3.1. The DS-AS Transition Can be Mediated by the External Input at the Receiver

In this section we analyze how an input in the receiver side can induce a transition from DS to AS (or vice versa) for the three systems under study.

### 3.1.1. 3-Neuron Motif

We start by studying the spiking dynamics of a circuit composed of 3 Hodgkin-Huxley cells coupled by chemical synaptic connections as in Matias et al. (2011) (for a schematic representation of the network architecture see **Figure 1A** and the Methods section for more details). The sender (S) neuron excites the receiver (R) neuron which also participates in an inhibitory loop mediated by an interneuron. For the simplest situation in which  $g_G = 20$  nS and all cells receive the same external current  $I_{ext}^S = I_{ext}^R = I_{ext}^{Inter} = 280$  pA, the neurons synchronize in the expected DS regime which exhibits the expected pre-post spike order. The neuronal time series show that the R neuron fires right after the S neuron (top panels of **Figure 1D**).

For an external constant current, after a transient time and within a synchronized state, the spike time difference converges to a constant value  $\tau = \tau_i$  that is independent of the initial conditions. By definition, DS is characterized by a positive  $\tau$  (or phase difference) and AS by a negative  $\tau$  (or phase difference). As we increase the inhibitory conductance  $g_G$ , the spike time difference  $\tau$  decreases, eventually changing sign and reaching negative values (see Matias et al., 2011). When this happens, the S and R neurons fire in a post-pre order (see bottom panel of **Figure 1D** for  $g_G = 20$  nS and  $I_{ext}^R = 320$  pA) characteristic of the AS regime. A similar effect can be obtained by increasing the external current at the receiver  $I_{ext}^R$ , without changing the value of the conductance  $g_G$ , as discussed below.

The dependence of  $\tau$  with  $g_G$  has been previously studied in Matias et al. (2011) for a large region of parameter space. The transition from DS to AS is continuous and smooth, and  $\tau$  is a function of  $g_G$  (see **Figure 2A**). Here we extend these findings, showing that a similar transition from DS to AS can be mediated by a different mechanism, namely increasing the external current of the R neuron. Starting from a DS regime in which  $I_{ext}^S = I_{ext}^R$  and  $g_G = 20$  nS, the spiking time difference  $\tau$  decreases as we increase  $I_{ext}^R$  (see **Figure 2D**).

### 3.1.2. Neuronal Populations

Similar patterns of out-of-phase synchronization have been reported for two unidirectionally coupled cortical-like populations composed of hundreds of neurons connected by chemical synapses (see Matias et al., 2014, **Figure 1B** and Methods for more details). Each population is composed by excitatory and inhibitory neurons, each of them receiving an independent Poisson input with rate  $\lambda$ , which accounts for excitatory synapses from neurons that are not included in the population. By construction, both Sender (S) and receiver (R) populations have inhibitory loops within the populations. Depending on the synaptic conductances and the external Poisson current, the mean activity of all neurons in each population may exhibit an oscillatory component. Moreover, the activity of the S and R populations can synchronize with a specific phase difference or equivalently time difference. As an example, it can be seen in the top panel of **Figure 1E** that if the neurons from both populations receive a noisy spike train with distribution rate  $\lambda_{ext}^S = 2.4$  kHz and  $\lambda_{ext}^R = 2.8$  kHz and the inhibitory synaptic conductance are  $g_G^S = 4.0$  nS and  $g_G^R = 8.0$  nS, the system operates in a DS regime. The peak of the

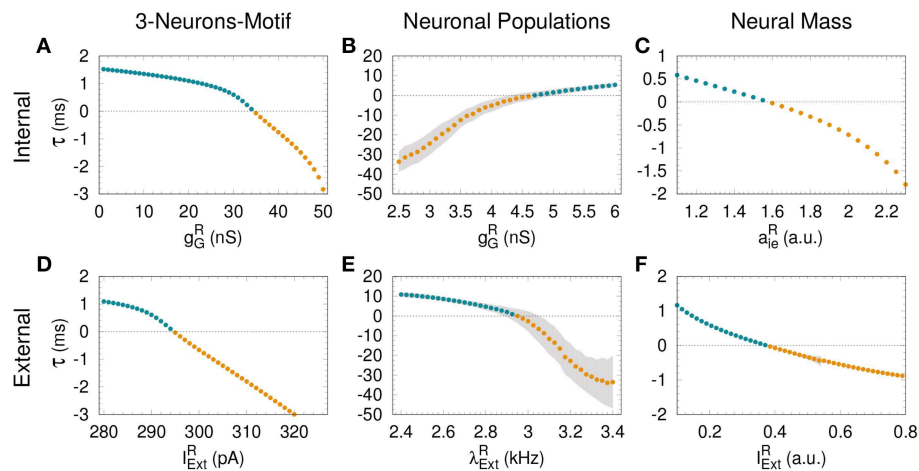
mean activity  $\langle V \rangle$  of the S population occurs before the peak of the R population. For the populations the spike time difference in each cycle  $i$  defined as  $\tau_i = t_i^R - t_i^S$ , where  $t_i^S$  is the peak of the mean activity of all neurons in S at the  $i$ -th cycle. Due to the noise, we can also define a spike time difference  $\tau$  as the mean of  $\tau_i$  averaged over many cycles. If we increase the external Poisson input at the R population,  $\tau$  decreases and the system reaches an AS regime (see the bottom panel of **Figure 1E** for  $\lambda_{ext}^R = 3.0$  kHz). The transition from DS to AS is continuous and smooth (see **Figure 2E**). To our knowledge, this is the first time that the DS-AS transition mediated by the level of noise in the receiver system is reported. Conversely, the dependence of  $\tau$  with the inhibitory conductance has been previously reported in Matias et al. (2014) (see **Figure 2B**).

### 3.1.3. Neural Mass Models

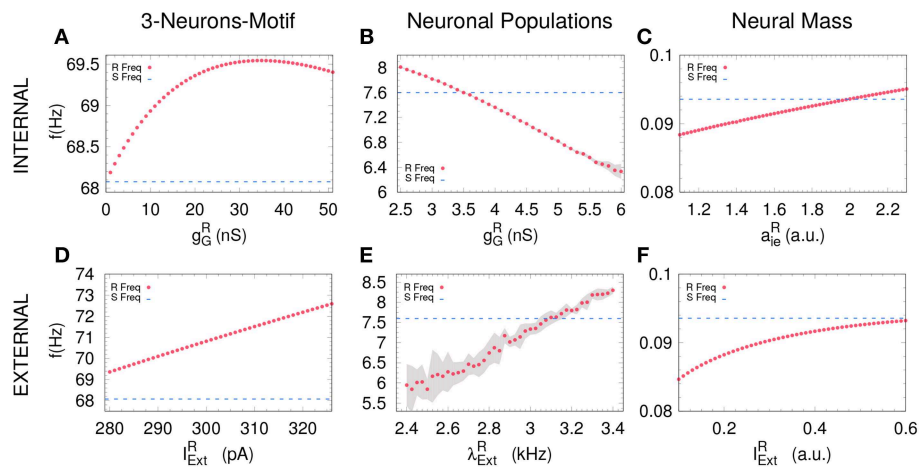
To further investigate the robustness of the relationship between the local parameters of the receiver system and the existence of anticipated synchronization we also studied the case of two unidirectionally coupled neural mass models, which represents a reduced model of spontaneous cortical dynamics. The neural mass model used here accounts for the neuronal population dynamics and uses three non-linear differential equations per node: one equation for the excitatory subpopulation, one for the inhibitory subpopulation and one for the number of open potassium channels. By its own definition the model has an inhibitory loop mediated by an effective conductance  $a_{ei}$  from excitatory to inhibitory neurons and  $a_{ie}$  from inhibitory to excitatory neurons (see **Figure 1C**). We find, as in previous cases, that two neural masses unidirectionally connected may synchronize and the inhibitory conductance  $a_{ie}$  as well as the external current  $I_{ext}^R$  at the receiver node can control the phase-locking difference between them (see **Figure 1F**). The transition from DS to AS via zero-lag can be obtained by increasing  $a_{ie}$  or  $I_{ext}^R$  (see **Figures 2C,F**).

## 3.2. The Frequency of the Free-Running Receiver Serves as a Mechanism Underlying the Synchronization Transition

Based on previous work on anticipated synchronization in the framework of Equation (1) (Pyragienė and Pyragas, 2013; Hayashi et al., 2016; Dima et al., 2018), we studied the effect that an inhibitory connection and the external current plays in determining the frequency of the receiver system when the sender and the receiver are uncoupled (the receiver free-running frequency  $\omega_R$ ). In fact, we find that both the inhibitory conductance and the external stimuli modify the receiver internal dynamics. More important, we find a correlation between the transition from DS to AS regime and the increase of the receiver free running frequency  $\omega_R$ , and consequently in  $\omega_R - \omega_S$  since  $\omega_S$  is fixed. We find that the DS-AS transition can be mediated by a change in the receiver free running frequency in multiple scales and by two different parameters (see **Figure 4**). Therefore, we propose that a faster free-running frequency of the receiver is the mechanism yielding anticipation. Similarly, it is also possible to obtain AS by keeping the Receiver intact and slowing-down the Sender free-running frequency.



**FIGURE 2 |** Assessing anticipated and delayed synchronization in cortical motifs. For all the models—3-neurons-motif (left), neuronal populations (middle) and neural masses (right)—a transition from delayed synchronization (DS,  $\tau > 0$ , cyan dots) to anticipated synchronization (AS,  $\tau < 0$ , yellow dots) is possible increasing the inhibition in the receiver (**A–C**); or increasing the external stimulus (**D–F**). Gray shadow represents the standard deviation over 10 runs (see Methods). For the parameters used here see Methods.



**FIGURE 3 |** Receiver free-running frequency. Assessing the free-running frequency of the receiver when uncoupled from the sender for all the models. (**A–C**) show how the receiver's frequency changes when varying the internal parameter while (**D–F**) when changing an external parameter. Blue dashed line represents the sender's natural frequency and magenta dots represents the receiver's frequency. Gray shadow represents the standard deviation over 10 runs (see Methods). For the parameters used here see Methods.

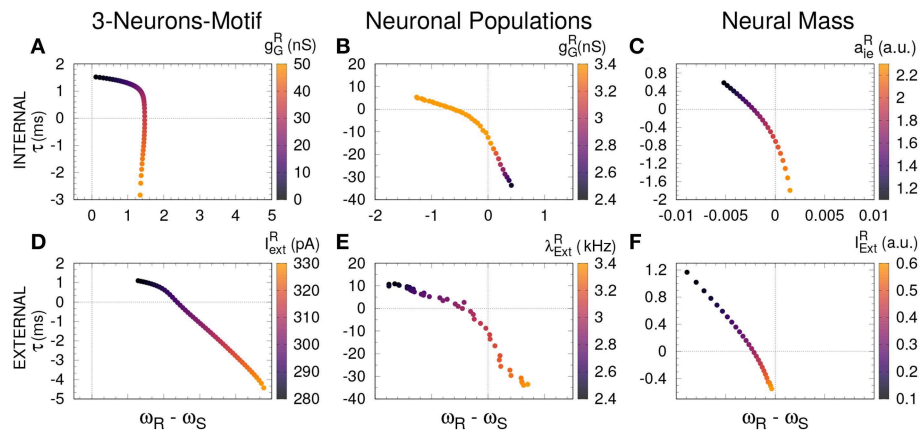
An increasing external stimuli increases the frequency of the uncoupled receiver for the three systems (see **Figures 3D–F**). On the contrary, an increasing inhibitory conductance increases the receiver frequency for the 3-neuron motif and the neural mass model but decreases the free-running frequency in the case of neuronal populations (compare **Figures 3A,C** with **Figure 3B**). Despite the fact that more inhibition is typically associated to less activity, it is well-established that resonant neurons, as Hodgkin Huxley model, can exhibit inhibition-induced spiking (Izhikevich, 2003). For neuronal populations and the neural mass model, the transition from DS to AS occurs roughly when the receiver pulses faster than the sender, whether it is due to the internal (**Figures 4B,C**) or the external factor (**Figures 4E,F**).

Nevertheless, this effect cannot be observed for the 3-neuron motif (**Figures 4A,D**), when the system is uncoupled; the Receiver always pulses at the same frequency or faster than the Sender (**Figures 3A–D**). This is due to the fact that, unlike other motifs, the Sender neuron is not subjected to any kind of inhibition. This means that in the uncoupled configuration, in the absence of inhibition at the receiver and same external current both neurons are identical and consequently  $\omega_R = \omega_S$ .

## 4. DISCUSSION

In this paper we have studied the effects that a change in the inhibitory conductance or an increase in the external forcing play





**FIGURE 4 |** Delayed and anticipated synchronization as a function of the free-running frequency. For all the cases,  $\tau$  (y-axis) was computed in a coupled system, while the frequency difference (x-axis) was computed for an uncoupled system; the parameters used in both cases are specified by the color code. **(A–C)** Internal parameter vs. frequency differences and **(D–F)** external parameter vs. frequency differences, for 3-neurons-motif, neuronal populations, and neural masses, respectively. Horizontal and vertical dashed lines represent zero-lag synchronization (coupled system) and a perfect match between sender and receiver free-running frequencies, respectively. For the parameters used here see Methods.

in the transition from delayed to anticipated synchronization in neuronal circuits. Our study covers three cases of unidirectionally coupled systems: two Hodgkin-Huxley (HH) neurons where the receiver neuron is coupled to an inhibitory interneuron, two populations and two neural mass (NM) models. The results obtained for the HH neurons and the neuron populations confirmed that, when changing the inhibitory conductance in the receiver side, the systems can undergo a delayed to anticipated synchronization transition (Matias et al., 2011, 2014). Similar results were obtained when analyzing two coupled neural mass (NM) models. Interestingly, we found a second mechanism that yields similar results. If we fix the inhibitory conductance in all cases but change the external input (external current for the HH and NM models or the Poisson rate in the populations) the system can also undergo a DS to AS transition.

To unveil if the two mechanisms are independent or not, we studied how the pulsating frequency of an isolated system (a pair of excitatory-inhibitory neurons, a single population containing excitatory and inhibitory neurons or a NM model) changes when changing the inhibitory conductance or the external forcing. We found for the excitatory-inhibitory pair and the NM model that an increase in the inhibitory conductance indeed increases the pulsating frequency (see **Figures 3A–C**), as happens when we increase the external current. However, for the parameters we have used for the populations, an opposite behavior is observed: an increase in the inhibitory conductance decreases the oscillating frequency of the population. These results suggest that the two mechanisms might be independent although a more exhaustive analysis is necessary. Nevertheless, we propose that the unifying mechanism that promotes the transition between DS and AS in the sender-receiver motif is indeed an increase of the internal dynamics of the receiver system. This was observed at the three scales that we examined. This result is in agreement with previous studies of AS in a theoretical framework with the explicit time-delay in Equation (1) (Hayashi et al., 2016) and in simplified

neural models without the explicit time-delay (Pyragienė and Pyragas, 2013, 2015; Dima et al., 2018).

Previous studies have shown that a biologically plausible mechanism for anticipation of pre-synaptic inputs is a combination of short-term synaptic depression (STD) and intrinsic spike-frequency adaptation (SFA) (Puccini et al., 2006, 2007). In the presence of both STD and SFA the post-synaptic system approximately computes the derivative a pre-synaptic stimuli which allows it to anticipate temporally incoming synaptic inputs. Their findings quantitatively agree with experimental results on anticipatory responses to moving stimuli in the primary visual cortex (Jancke et al., 2004; Puccini et al., 2007). We propose that the DS-AS transition studied here could be mediated by any parameter that turns the internal dynamics of the free-receiver faster and could also account for delay compensation in cortical systems. Specially the DS-AS transition could explain commonly reported short latency in visual systems (Orban et al., 1985; Nowak et al., 1995; Kerzel and Gegenfurtner, 2003; Jancke et al., 2004; Puccini et al., 2007; Stepp and Turvey, 2010, 2017; Martinez et al., 2014). Therefore, our results open new possibilities to further experimental investigation of anticipatory dynamics in neuronal systems.

## DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

## AUTHOR CONTRIBUTIONS

LD, FM, PC, MC, and CM designed the research and wrote the article. LD, AdS, and AA performed the simulations. LD performed the data analysis.

## FUNDING

We thank CAPES (grant PVE 88881.068077/2014-01) for financial support. MC received financial support from CNPq (grants 301744/2018-1 and 425329/2018-6) and FAPESP Center for Neuromathematics (grant 2013/07699-0, Sao Paulo Research Foundation). PC received financial support from FACEPE (APQ 0826-1.05/15). MC, PC, and AdS received

support from CAPES PROEX program 534/2018 (grant 23038.003382/2018-39). FM was supported by CNPq (grant 432429/2016-6) and CAPES (grant 88881.120309/2016-01). CM acknowledge the Spanish State Research Agency, through the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence in R&D (MDM-2017-0711). MC and PC acknowledge support from FACEPE Grant No. APQ-0642-1.05/18.

## REFERENCES

- Barardi, A., Sancristóbal, B., and Garcia-Ojalvo, J. (2014). Phase-coherence transitions and communication in the gamma range between delay-coupled neuronal populations. *PLoS Comput. Biol.* 10:e1003723. doi: 10.1371/journal.pcbi.1003723
- Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., and Bressler, S. L. (2004). Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by granger causality. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9849–9854. doi: 10.1073/pnas.0308538101
- Buzsáki, G. (2006). *Rhythms of the Brain* (New York, NY: Oxford University Press).
- Ciszak, M., Calvo, O., Masoller, C., Mirasso, C. R., and Toral, R. (2003). Anticipating the response of excitable systems driven by random forcing. *Phys. Rev. Lett.* 90:204102. doi: 10.1103/PhysRevLett.90.204102
- Ciszak, M., Mirasso, C. R., Toral, R., and Calvo, O. (2009). Predict-prevent control method for perturbed excitable systems. *Phys. Rev. E* 79:046203. doi: 10.1103/PhysRevE.79.046203
- Corron, N. J., Blakely, J. N., and Pethel, S. D. (2005). Lag and anticipating synchronization without time-delay coupling. *Chaos* 15:023110. doi: 10.1063/1.1898597
- Dima, G. C., Copelli, M., and Mindlin, G. B. (2018). Anticipated synchronization and zero-lag phases in population neural models. *Int. J. Bifur. Chaos* 28:1830025. doi: 10.1142/S0218127418300252
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cognit. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011
- Gollo, L. L., Mirasso, C., Sporns, O., and Breakspear, M. (2014). Mechanisms of zero-lag synchronization in cortical motifs. *PLoS Comput. Biol.* 10, 1–17. doi: 10.1371/journal.pcbi.1003548
- Hayashi, Y., Nasuto, S. J., and Eberle, H. (2016). Renormalized time scale for anticipating and lagging synchronization. *Phys. Rev. E* 93:052229. doi: 10.1103/PhysRevE.93.052229
- Hernández-García, E., Masoller, C., and Mirasso, C. (2002). Anticipating the dynamics of chaotic maps. *Phys. Lett. A* 295, 39–43. doi: 10.1016/S0375-9601(02)00147-0
- Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Neurophysiol.* 17, 500–544. doi: 10.1113/jphysiol.1952.sp004764
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Trans. Neural Netw.* 14, 1569–1572. doi: 10.1109/TNN.2003.820440
- Jancke, D., Erlhagen, W., Schöner, G., and Dinse, H. R. (2004). Shorter latencies for motion trajectories than for flashes in population responses of cat primary visual cortex. *J. Physiol.* 556, 971–982. doi: 10.1113/jphysiol.2003.058941
- Kerzel, D., and Gegenfurtner, K. R. (2003). Neuronal processing delays are compensated in the sensorimotor branch of the visual system. *Curr. Biol.* 13, 1975–1978. doi: 10.1016/j.cub.2003.10.054
- Köksal Ersős, E., Desroches, M., Mirasso, C. R., and S. (2019). Anticipation via canards in excitable systems. *Chaos* 29, 01311–01311-18. doi: 10.1063/1.5050018
- Kostur, M., Hänggi, P., Talkner, P., and Mateos, J. L. (2005). Anticipated synchronization in coupled inertial ratchets with time-delayed feedback: a numerical study. *Phys. Rev. E* 72:036210. doi: 10.1103/PhysRevE.72.036210
- Martínez, L. M., Molano-Mazón, M., Wang, X., Sommer, F. T., and Hirsch, J. A. (2014). Statistical wiring of thalamic receptive fields optimizes spatial sampling of the retinal image. *Neuron* 81, 943–956. doi: 10.1016/j.neuron.2013.12.014
- Masoller, C., and Zanette, D. H. (2001). Anticipated synchronization in coupled chaotic maps with delays. *Phys. A* 300, 359–366. doi: 10.1016/S0378-4371(01)00362-4
- Matias, F. S., Carelli, P. V., Mirasso, C. R., and Copelli, M. (2011). Anticipated synchronization in a biologically plausible model of neuronal motifs. *Phys. Rev. E* 84:021922. doi: 10.1103/PhysRevE.84.021922
- Matias, F. S., Carelli, P. V., Mirasso, C. R., and Copelli, M. (2017). Anticipated synchronization in neuronal circuits unveiled by a phase-response-curve analysis. *Phys. Rev. E* 95:052410. doi: 10.1103/PhysRevE.95.052410
- Matias, F. S., Gollo, L. L., Carelli, P. V., Copelli, M., and Mirasso, C. R. (2014). Modeling positive Granger causality and negative phase lag between cortical areas. *Neuroimage* 99:411. doi: 10.1016/j.neuroimage.2014.05.063
- Nowak, L. G., Munk, M. H., Girard, P., and Bullier, J. (1995). Visual latencies in areas v1 and v2 of the macaque monkey. *Vis. Neurosci.* 12, 371–384. doi: 10.1017/S095252380000804X
- Orban, G. A., Hoffmann, K. P., and Duysens, J. (1985). Velocity selectivity in the cat visual system. I. Responses of lgn cells to moving bar stimuli: a comparison with cortical areas 17 and 18. *J. Neurophysiol.* 54, 1026–1049. doi: 10.1152/jn.1985.54.4.1026
- Puccini, G. D., Sanchez-Vives, M. V., and Compte, A. (2006). Selective detection of abrupt input changes by integration of spike-frequency adaptation and synaptic depression in a computational network model. *J. Physiol.* 100, 1–15. doi: 10.1016/j.jphysparis.2006.09.005
- Puccini, G. D., Sanchez-Vives, M. V., and Compte, A. (2007). Integrated mechanisms of anticipation and rate-of-change computations in cortical circuits. *PLoS Comput. Biol.* 3:e82. doi: 10.1371/journal.pcbi.0030082
- Pyragienė, T., and Pyragas, K. (2013). Anticipating spike synchronization in nonidentical chaotic neurons. *Nonlinear Dyn.* 74, 297–306. doi: 10.1007/s11071-013-0968-7
- Pyragienė, T., and Pyragas, K. (2015). Anticipating chaotic synchronization via act-and-wait coupling. *Nonlinear Dyn.* 79, 1901–1910. doi: 10.1007/s11071-014-1782-6
- Pyragiene, T., and Pyragas, K. (2017). Anticipatory synchronization via low-dimensional filters. *Phys. Lett. A* 381, 1893–1898. doi: 10.1016/j.physleta.2017.04.005
- Salazar, R. F., Dotson, N. M., Bressler, S. L., and Gray, C. M. (2012). Content-specific fronto-parietal synchronization during visual working memory. *Science* 338, 1097–1100. doi: 10.1126/science.1224000
- Sausedo-Solorio, J., and Pisarchik, A. (2014). Synchronization of map-based neurons with memory and synaptic delay. *Phys. Lett. A* 378, 2108–2112. doi: 10.1016/j.physleta.2014.05.003
- Simonov, A. Y., Gordleeva, S. Y., Pisarchik, A., and Kazantsev, V. (2014). Synchronization with an arbitrary phase shift in a pair of synaptically coupled neural oscillators. *JETP Lett.* 98, 632–637. doi: 10.1134/S0021364013230136
- Sivaprakasam, S., Shahverdiev, E. M., Spencer, P. S., and Shore, K. A. (2001). Experimental demonstration of anticipating synchronization in chaotic semiconductor lasers with optical feedback. *Phys. Rev. Lett.* 87:154101. doi: 10.1103/PhysRevLett.87.154101
- Stepp, N., and Turvey, M. T. (2010). On strong anticipation. *Cognit. Syst. Res.* 11, 148–164. doi: 10.1016/j.cogsys.2009.03.003
- Stepp, N., and Turvey, M. T. (2017). Anticipation in manual tracking with multiple delays. *J. Exp. Psychol. Hum. Percept. Perform.* 43:914. doi: 10.1037/xhp0000393
- Tang, S., and Liu, J. M. (2003). Experimental verification of anticipated and retarded synchronization in chaotic semiconductor lasers. *Phys. Rev. Lett.* 90:194101. doi: 10.1103/PhysRevLett.90.194101

- Voss, H. U. (2000). Anticipating chaotic synchronization. *Phys. Rev. E* 61:5115. doi: 10.1103/PhysRevE.61.5115
- Voss, H. U. (2001a). Dynamic long-term anticipation of chaotic states. *Phys. Rev. Lett.* 87:014102. doi: 10.1103/PhysRevLett.87.014102
- Voss, H. U. (2001b). Erratum: anticipating chaotic synchronization [phys. rev. e 61, 5115 (2000)]. *Phys. Rev. E* 64:039904. doi: 10.1103/PhysRevE.64.039904
- Voss, H. U. (2016). Signal prediction by anticipatory relaxation dynamics. *Phys. Rev. E* 93:030201. doi: 10.1103/PhysRevE.93.030201
- Voss, H. U. (2018). A delayed-feedback filter with negative group delay. *Chaos Interdiscip. J. Nonlinear Sci.* 28:113113. doi: 10.1063/1.5052497

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Dalla Porta, Matias, dos Santos, Alonso, Carelli, Copelli and Mirasso. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Coupling the State and Contents of Consciousness

Jaan Aru<sup>1,2,3\*</sup>, Mototaka Suzuki<sup>4</sup>, Renate Rutiku<sup>3</sup>, Matthew E. Larkum<sup>1,4</sup>  
and Talis Bachmann<sup>3</sup>

<sup>1</sup>Institute of Biology, Humboldt University of Berlin, Berlin, Germany, <sup>2</sup>Institute of Computer Science, University of Tartu, Tartu, Estonia, <sup>3</sup>School of Law, University of Tartu, Tartu, Estonia, <sup>4</sup>Neurocure Center for Excellence, Charité Universitätsmedizin Berlin, Berlin, Germany

One fundamental feature of consciousness is that the contents of consciousness depend on the state of consciousness. Here, we propose an answer to why this is so: both the state and the contents of consciousness depend on the activity of cortical layer 5 pyramidal (L5p) neurons. These neurons affect both cortical and thalamic processing, hence coupling the cortico-cortical and thalamo-cortical loops with each other. Functionally this coupling corresponds to the coupling between the state and the contents of consciousness. Together the cortico-cortical and thalamo-cortical loops form a thalamo-cortical broadcasting system, where the L5p cells are the central elements. This perspective makes one quite specific prediction: cortical processing that does not include L5p neurons will be unconscious. More generally, the present perspective suggests that L5p neurons have a central role in the mechanisms underlying consciousness.

**Keywords:** consciousness, thalamus, pyramidal neurons, dendrites, unconscious processing, state of consciousness

## OPEN ACCESS

### Edited by:

Igor Timofeev,  
Laval University, Canada

### Reviewed by:

Francisco Clasca,  
Autonomous University of Madrid,  
Spain  
Marco Atzori,  
Universidad Autónoma de San Luis  
Potosí, Mexico

### \*Correspondence:

Jaan Aru  
jaan.aru@gmail.com

**Received:** 29 March 2019

**Accepted:** 13 August 2019

**Published:** 30 August 2019

### Citation:

Aru J, Suzuki M, Rutiku R,  
Larkum ME and Bachmann T  
(2019) Coupling the State and  
Contents of Consciousness.  
*Front. Syst. Neurosci.* 13:43.  
doi: 10.3389/fnsys.2019.00043

## INTRODUCTION

Each of us can be fully conscious, have a dream, be in deep sleep or be anesthetized. These are typical *states of consciousness*. On the other hand, we can be conscious of a dog, a paper, coconut taste, itch etc. These are examples of the *contents of consciousness* (from among a world of different possibilities). Unfortunately for consciousness research, the state of consciousness is mostly studied separately from the contents (Bachmann and Hudetz, 2014). The research done on the state of consciousness mainly revolves around the thalamus and thalamocortical interactions (reviews: Laureys, 2005; Alkire et al., 2008; Schiff, 2010). On the other hand, the search for the correlates of the contents of consciousness is mostly focused on cortical processing (reviews: Rees et al., 2002; Dehaene and Changeux, 2011; Koch et al., 2016).

However, one basic fact about consciousness is that the state of consciousness can never be dissociated from the contents of consciousness. One cannot be conscious of the coconut taste while being in an unconscious state. And the other way around: in typical healthy subjects, one cannot be in a conscious state while not being conscious of anything at all<sup>1</sup>. In other words, contents of consciousness and states of consciousness make up an integrated whole. Studying one while disregarding the other can only provide half of an answer.

<sup>1</sup>Here one might want to suggest that during some forms of meditation people can be conscious without being conscious of anything, but even if one agrees with this claim, the length of the necessary training demonstrates how hard it is to separate the state and the contents of consciousness. Furthermore, it is possible to say that the “emptiness” of conscious experience itself is a content (according to set theory, an empty set is a subset of all sets).



The intertwinement of state and contents of consciousness is a well-known issue in consciousness research (e.g., Hohwy, 2009; Bachmann, 2012; Bachmann and Hudetz, 2014; Mashour and Hudetz, 2017). However, it is currently unclear why contents and state of consciousness have to be so tightly coupled in the first place. In this work, we highlight a clear neurobiological correlate for this intertwinement. Namely, cortical layer 5 pyramidal (L5p) neurons participate in both the thalamo-cortical and cortico-cortical loops and couple these two loops in a unique way, hence functionally coupling the state and contents of consciousness (**Figure 1**). This view reconciles the two traditions of consciousness research: the tradition of studying the state of consciousness with the focus on the thalamo-cortical system and the tradition of investigating the contents of consciousness with a particular emphasis on the cortico-cortical processing. We will first very briefly review these two traditions and then explain how L5p neurons couple the activity patterns of the thalamo-cortical and cortico-cortical loops.

## THALAMO-CORTICAL LOOP AND THE STATE OF CONSCIOUSNESS

Conscious experience alternates between being present and not being present, with intermediate states between full absence of conscious mentation and full presence of clear, vivid and temporally seamless subjective experience (e.g., sleep, vegetative state, minimally conscious state, hypnagogic state, fully conscious state—with varying levels of clarity or degree of experience across different states of alertness).

The idea that thalamus plays a key role in controlling the states of consciousness is not new. According to the prevailing views of the mid-20th century thalamus is part of the reticulo-thalamic system with two general types of thalamo-cortical pathways: (1) the so-called specific pathways (SP) which function as the carriers of afferent information to the cortex [the prime example is the lateral geniculate nucleus (LGN) for vision]; and (2) the so-called non-specific pathways (NSP) that have no direct role in transmitting information to cortex, but are capable of modulating the state of the cortex (**Figure 1**; Moruzzi and Magoun, 1949; Jung, 1958; Magoun, 1958; Riklan and Levita, 1969; Doty, 1970; Purpura, 1970; Somjen, 1972; Brooks and Jung, 1973; Brazier, 1977; Hassler, 1978; Smirnov et al., 1978; Singer, 1979; Livingstone and Hubel, 1980; Steriade, 1981a; Mesulam, 1985; Newman, 1995a). Historically, these two types of pathways were first distinguished based on the identity of the thalamic nucleus they originated from (LGN would be an example of SP nuclei, the intralaminar thalamic nuclei are classic examples of NSP nuclei). However, it later turned out that the projections are more intermingled (e.g., Jones, 2012; Clascá et al., 2016), hence it is too simplified to refer to SP and NSP nuclei. In this manuscript when we refer to NSP thalamus we have in mind these thalamic cells that project to the superficial and deep layers of the cortex (see also **Figure 1**). Such thalamocortical cells mainly reside in the classic NSP nuclei, but there is much diversity among them<sup>2</sup>.

<sup>2</sup>We note that even this division is a simplification, as there also exist thalamo-cortical projections that have the specific projection pattern (i.e., To layer 4) in one

There are several types of evidence demonstrating that especially the NSP thalamus is directly involved in controlling the state of consciousness.

First, alternation of sleep and wakefulness depends on NSP; for example, intralaminar thalamocortical neurons increase their firing rate about 10 s before EEG desynchronization in natural transitions from slow-wave sleep to waking or active sleep (Steriade, 1981b). At sleep onset thalamic deactivation is observed first, followed by cortical changes (Magnin et al., 2010). Recently, Honjoh et al. (2018) showed that stimulation of the NSP ventromedial (VM) thalamic nucleus awoke mice from NREM sleep and anesthesia and caused EEG activation in the high-frequency band.

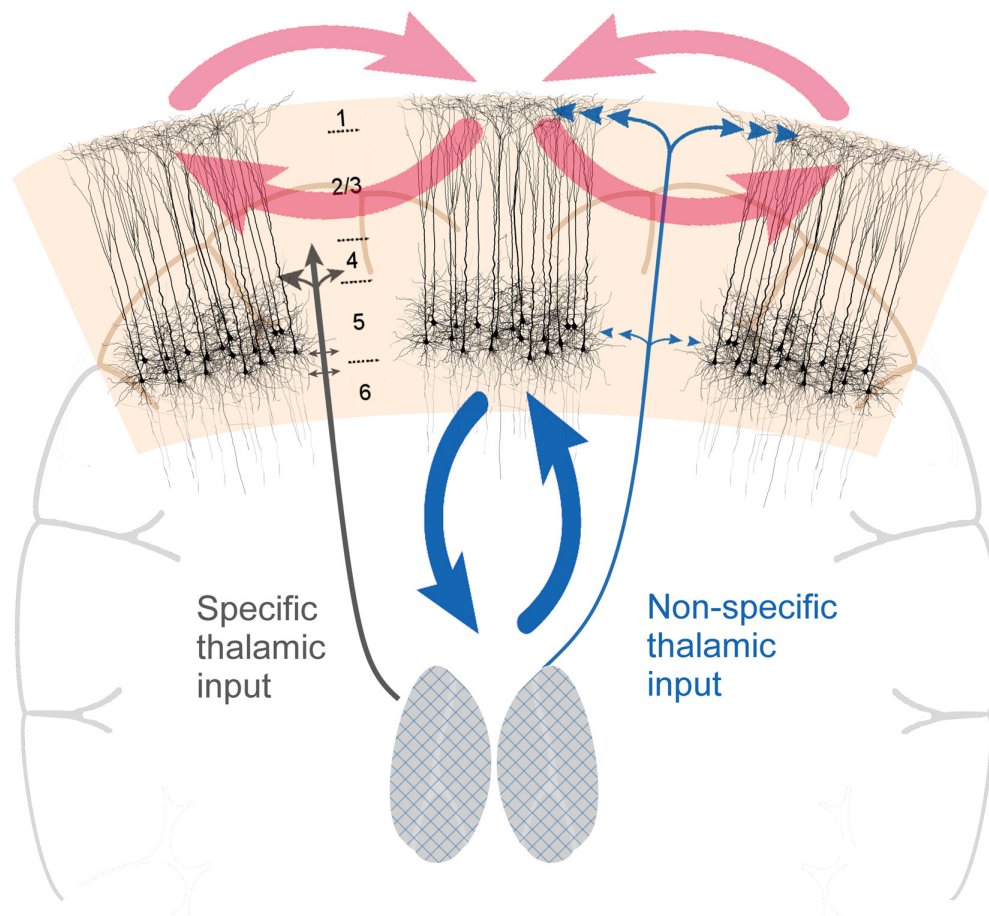
In addition, virtually all general anesthetics, despite their differences in neurobiological effects elsewhere, have common target in the NSP thalamus (Alkire et al., 2000, 2008). If the NSP thalamus of the anesthetized experimental animals is electrically stimulated, desynchronization of EEG occurs with activity patterns typical to the awake brain (Bremer, 1935; Moruzzi and Magoun, 1949; Brazier, 1977; Munk et al., 1996). These results are corroborated by the fact that injuries and tumors localized in the NSP thalamus often cause absence of consciousness in patients despite the fact that the SP system has remained relatively intact (e.g., Riklan and Levita, 1969; Penfield, 1975; Kinney et al., 1994; Bogen, 1995a,b; Newman, 1995b). Taken together, these findings show that thalamus is in an optimal position to modulate and control the state and level of consciousness (but see Alkire et al., 2008; Hudetz, 2012; Mashour, 2014; for counter-arguments suggesting that the effects in thalamus are only secondary to those in cortex).

Based on this evidence, thalamo-cortical theories of consciousness were introduced (e.g., Purpura, 1970; Hassler, 1978; Bachmann, 1984, 1989, 1999; Bogen, 1995a,b; Newman, 1995a; Llinás, 1996; LaBerge, 1997; Purpura and Schiff, 1997; Tononi and Edelman, 1998; Ward, 2011). As the name suggests, a thalamocortical theory of consciousness proposes that conscious experience is based on the interactions between thalamic nuclei and areas of cortex. According to these theories, there is no specific area of the cortex that is related to consciousness: it is the interplay with the thalamus that matters.

## CORTICO-CORTICAL LOOP AND THE CONTENTS OF CONSCIOUSNESS

Since the turn of the century, many leading theories of consciousness have emphasized the relationship between consciousness and cortical processing. For example, although the neural global workspace theory (Dehaene and Naccache, 2001; Dehaene and Changeux, 2011) includes thalamocortical processing, it has a minor role compared to the cortical frontoparietal broadcasting system (Dehaene and Changeux, 2011). In addition, the theory that consciousness is related

cortical target area, while having a non-specific projection pattern (i.e., To layers 1 and 5/6) in another (the so-called multi-specific thalamocortical projection cells, see Clascá et al., 2016).



**FIGURE 1 |** Cortical layer 5 pyramidal (L5p) neurons (the black-colored neurons on the image) play a central role in both cortico-cortical and thalamo-cortical loops. By being central to both loops, they effectively couple them, functionally coupling the state and contents of consciousness. Two types of thalamic projections are highlighted: specific (SP) and nonspecific (NSP) types of projections that have different cortical projection patterns. The grids on the thalami are illustrative of the fact that thalamocortical neurons with NSP types of projections can be found in different parts of the thalamus.

to cortical recurrent processing is purely cortico-centric (Lamme, 2003, 2010). As a final example, the higher order thought theory of consciousness claims that consciousness is tightly linked to computations that crucially depend on the prefrontal cortex (PFC; Lau and Rosenthal, 2011).

Perhaps this change of perspective from thalamus to cortex was driven by methodological advances: by combining appropriate experimental approaches with microelectrode recordings and functional magnetic resonance imaging (fMRI) much evidence could be gathered about the role of cortical areas in consciousness. For example, it was shown that during binocular rivalry the firing of cells in primary visual cortex was not modulated by conscious experience of the monkey (Leopold and Logothetis, 1996). In contrast, the firing of neurons in inferior temporal cortex was stronger when the preferred stimulus was consciously perceived (Sheinberg and Logothetis, 1997). This result was celebrated as paving the way for understanding the neural correlates of the contents of consciousness. For instance, in the enthusiastic commentary accompanying the first of the above-mentioned articles, Francis

Crick wrote about the problem of consciousness that “with a little luck, we may glimpse the outline of the solution before the end of the century” (Crick, 1996).

In 1998, as one of the first seminal studies about the neural markers of the contents of consciousness with human neuroimaging, Lumer et al. (1998) used binocular rivalry and fMRI to demonstrate that the activity of the frontoparietal networks is involved in the changes of the content of consciousness (i.e., the dominating image in rivalry). In a later study using visual masking (Dehaene et al., 2001) it was revealed with fMRI that visible words elicited a strong activation of the frontoparietal network that was not observed when the same words were made invisible through masking. Based on these findings the neural global workspace theory (Dehaene and Naccache, 2001; Dehaene and Changeux, 2011) was established and became one of the most influential theories of consciousness.

Hence, the experimental techniques combined with psychophysical paradigms had opened up new venues for studying the contents of consciousness. Consciousness research became dominated by studies done on healthy humans and the

contents of consciousness. This research has generated many interesting findings (reviewed in Rees et al., 2002; Dehaene and Changeux, 2011; Koch et al., 2016) and many controversies (Aru et al., 2012; De Graaf et al., 2012, for a recent debate see for example Boly et al., 2017; Odegaard et al., 2017). However, this focus on the contents of consciousness in humans has left unanswered the question why these contents of consciousness are dependent on the state of consciousness. We would like to demonstrate next how thinking about this issue might also lead us closer to understanding the neurobiological mechanisms of consciousness in general.

## STATE AND CONTENT OF CONSCIOUSNESS INTERACT IN LAYER 5 PYRAMIDAL NEURONS

Why are the contents of consciousness dependent on the state of consciousness? In a nutshell, we claim that thalamo-cortical and cortico-cortical processes mechanistically interact at the level of cortical L5p cells, functionally corresponding to the intertwining of the state and contents of consciousness. Our view, thus, has the potential to reconcile the thalamo-cortical and cortico-cortical theories of consciousness.

From the inspection of the thalamocortical system, it becomes evident that thalamo-cortical and cortico-cortical loops uniquely intersect at the level of L5p neurons (**Figure 1**). L5p cells participate in both loops by receiving input from and sending output to both thalamus and cortex (Sherman and Guillery, 2001; Jones, 2012; Harris and Shepherd, 2015). There are at least two subpopulations of L5p cells and it might be that some of them (L5A neurons, in upper layer 5) are mainly involved in cortico-cortical loops whereas others (L5B neurons, in deeper layer 5) mainly participate in the loop with the NSP thalamic nuclei (Larsen et al., 2008; Kawaguchi, 2017; Takahashi et al., under revision). Thus, as a population, L5p neurons affect both thalamo-cortical and cortico-cortical processing. It has long been suspected that these cells also have a central role in the neural mechanisms underlying consciousness (Crick and Koch, 1992; Angel, 1993; Llinás and Ribary, 2001; LaBerge, 2006; Meyer, 2015).

Cortical L5p neurons have two functionally distinct sites of integration, one in the soma (somatic compartment) and one near the top of their apical trunk (apical compartment; e.g., Larkum, 2013). This segregation is interesting as the input to the basal dendrites and the apical tuft are known to be both functionally and anatomically quite distinct (Larkum, 2013). In particular, the apical tuft receives diverse input from higher cortical areas and NSP thalamic nuclei whereas the basal dendrites receive more specific feedforward information from lower cortical areas (and, in sensory cortex, from thalamic relay nuclei; Llinás and Ribary, 2001; Jones, 2012; Larkum, 2013).

If basal and apical compartments are depolarized at around the same time, the probability that the neuron fires a burst of action potentials is greatly increased (e.g., Larkum et al., 1999; Larkum, 2013). These bursts have a profound effect on both cortical and thalamic processing, as a burst makes it more likely

that the spiking of a neuron contributes to learning (Lisman, 1997) and behavior (Takahashi et al., 2016). The NSP thalamic nuclei have diffuse projection patterns across the entire cortex, hence a burst firing of a few cortical columns could reach the whole thalamo-cortical system. For example, it is known that the secondary somatosensory cortex (S2) can be activated by L5p neurons in S1 through the NSP thalamus (Theyel et al., 2010). After chemically deactivating the respective thalamic nucleus, the activation of S2 was eliminated, suggesting that this trans-thalamic pathway is a key component for cortical activity propagation (Theyel et al., 2010). Similarly, there is a circuit from L5p cells of S1 through the NSP thalamus to the primary motor cortex (M1; Mo and Sherman, 2019). Thus, the NSP thalamus is the basis for a thalamo-cortical broadcasting system and L5p cells are the central elements of it.

The direct relationship between L5p cell activity and the state of consciousness has been described in Murayama and Larkum (2009). These authors performed fiberoptic  $\text{Ca}^{2+}$  imaging of the apical dendrites in L5p cells. They compared the effect of different states of consciousness on the activity of the apical compartment while delivering brief air-puffs to the hindlimb of the animal. The apical dendritic response to hindlimb stimulation was 4-fold stronger in the quiet awake state than in the anesthetized state. However, when the animal moved its hindlimb in response to the air-puff, the apical dendritic response was an astonishing 14-fold stronger than under anesthesia. This increase in the  $\text{Ca}^{2+}$  activity mostly stemmed from a prolonged duration of the response: when the animal moved its hindlimb, the apical dendritic response lasted for several seconds. This result shows that the awake state has an enormous impact on the activity of L5p apical dendrites (Phillips et al., 2018). Moreover, while the authors did not interpret their findings in this fashion, it is also possible to see the interaction between the state and contents of consciousness in these results. Namely, one way to interpret the behavior of the animals is that when they reacted to the air-puff with a movement, they perceived it, whereas when they did not move in reaction to the air-puff then they did not. Under this scenario, the massive increase in the duration of the activity of L5p apical dendrites is reflecting the interaction between the state and the contents of consciousness.

There is more evidence showing that L5p neurons are tightly linked to the contents of perception. Takahashi et al. (2016) found that manipulation of apical dendritic activity of L5p cells affects the behavioral report of the animal. In this experiment awake behaving mice learned to detect weak whisker stimuli of different magnitudes i.e., sometimes at, sometimes below, sometimes above the detection threshold. This allowed the researchers to delineate psychometric curves for whisker stimulation detection and to correlate these curves with neurometric curves. To monitor the activity of the apical tuft dendrites the authors performed fast-scanning two-photon  $\text{Ca}^{2+}$  imaging. They observed that the  $\text{Ca}^{2+}$  signal of apical dendrites of L5p cells was well correlated with the behavior of the animal. In addition to aligning with the psychometric curves, the apical dendritic signals could predict the behavioral hits and misses of threshold stimuli. Most importantly, directly modulating the dendritic activity through pharmacological intervention



or optogenetics had a measurable influence on the detection behavior of the animal by shifting the psychometric curve. Optogenetic enhancement of apical dendritic activity also led to false alarms. In other words, activating the apical compartment caused the animal to lick as if a whisker stimulus would have been present. A bold but highly plausible interpretation would be that the animal experienced an illusory stimulus, just like humans have been shown to do when expecting a stimulus given some context (Aru and Bachmann, 2017; Aru et al., 2018).

Interestingly, in the follow-up experiments, it was learned that the positive relationship between dendritic activity and animal behavior is largely constrained to the pyramidal cells in layer 5B that project to subcortical regions including non-specific thalamic nuclei (Takahashi et al., under revision). On the other hand, layer 5A cells with denser corticocortical projections were not so tightly correlated with the perceptual report of the animal (Takahashi et al., under revision). Hence, these findings support the conclusion that L5p cells, especially those projecting to the thalamus, are vital for conscious perception.

## NON-CONSCIOUS CONTENTS IN A CONSCIOUS STATE

A framework that proposes how the state and content of consciousness interact in the brain has to give answers to two questions. First, why are state and content so tightly bound in the brain? This question was answered in the first part of this work: we claim that the intertwining of state and contents of consciousness arises at the level of single L5p neurons because they hold a central position in both thalamo-cortical and cortico-cortical loops. Second, the theory has to explain the existence of unconscious processing of contents: if state and content are bound in the brain, then how can some processing of the contents be unconscious? This is the question we turn to next.

Mental information processing can proceed also in its subliminal mode, meaning that not all sensory signals that are veridically encoded and adequately responded to are consciously experienced (for reviews, see Tulving et al., 1982; Kim and Blake, 2005; Mitchell, 2006; Kouider and Dehaene, 2007; Riener, 2017). Only part of the signals carrying contents become conscious. In fact, consciousness only has access to very specific levels of representation. Other levels are firmly locked away from it. For example, we do not have access to the processes that segregate our surrounding environment into discrete objects. How border ownership is calculated remains elusive to consciousness—we only experience the result of it. Similarly, as we speak or write in sentences we have no conscious experience of how our brain constructs them, i.e., chooses the words and puts them in the right order with correct grammar. Hence, conscious experience is restricted to certain contents and computations.

The present theory offers a natural way to understand this: all subcortical and cortical processing that does not involve L5p neurons will remain non-conscious. For example, motor control is based on the computations of basal ganglia and cerebellum, which are detached from the loop of consciousness. Both basal ganglia and the cerebellum have connections to the thalamus and cortex, which means that the outputs

of the processes happening there can be incorporated into consciousness, while the computations themselves taking place within the cerebellum and basal ganglia remain unconscious (Tononi, 2004).

However, non-conscious processing also happens in cortex. According to the present view, if the processing does not involve L5p cells or the apical-dendritic modulation of these cells remains insufficient, this processing will not reach the NSP thalamus and will not be conscious. In other words, we make the strong prediction that cortical processing in itself, when not integrated with the NSP thalamic nuclei *via* L5p neurons, is not conscious. In particular, feedforward cortical processing, where information is mainly flowing within the cortical superficial layers bypassing thalamocortical neurons, is non-conscious.

## THALAMO-CORTICAL BROADCASTING IN ACTION

As a closing argument, we wish to demonstrate how the presently proposed view can explain one central phenomenon often used in consciousness research. It is well known that conscious experience has a limited temporal resolution, meaning that if the presentation of different stimuli occurs too rapidly then some fail to be consciously experienced. This is evidenced in empirical phenomena like backward masking (for review, see Bachmann and Francis, 2013) and attentional blink (for review, see Martens and Wyble, 2010). Even when an identical stimulus is switched on and off intermittently at very high frequency, the multiplicity of stimulus replications will not be consciously perceived and the stimuli will fuse into a continuous experience (Watson, 1986). This characteristic is tightly related to the previous one: if conscious perception is based on the processing within the thalamocortical loop, then it is hard for conscious perception to resolve anything that happens *faster* than the processing time of this loop. In other words, we claim that the temporal resolution of conscious experience stems from the propagation time between the L5p neurons, NSP thalamus and higher cortical areas.

In *backward masking*, two stimuli are presented in quick succession. Under certain conditions, the second stimulus (“the mask”) can completely abolish the first (“the target”) from consciousness (Bachmann, 1984, 1994, 2000). Thus, although the target is presented 50–100 ms before the mask, it is not consciously perceived. The present theory explains this phenomenon in the following way. When the first stimulus, the target, activates L5p neurons, it initiates the thalamocortical broadcasting loop. However, by the time this activity propagates from the L5p neurons to the NSP thalamus and back to the apical dendrites of L5p neurons, the second stimulus, the mask, has taken over early cortical representation and now “steals the fire” which was started by the target. The target starts the loop, but the mask benefits from it.

## THE FUTURE OF THE PRESENT HYPOTHESIS

The present conjecture can be directly tested in the rodent model, where it is possible to directly manipulate the different



components of the loops depicted in **Figure 1**. For example, it is possible to specifically affect the different compartments of L5p neurons (e.g., Takahashi et al., 2016; Suzuki and Larkum, 2017). Which elements of the L5p neurons are most affected by the changes in the state of consciousness? It has been suggested that anesthesia especially affects the apical dendritic compartment (Phillips et al., 2018), but there are several other possibilities too (Meyer, 2015). It is also possible to specifically target the NSP thalamus (Honjoh et al., 2018) and to combine this with sensory stimulation. Future studies using the modern technological toolkit will directly test the merit of the present proposal. In this sense, the present ideas are clearly preliminary and will need to be refined, but the goal here was to propose that future work on the mechanisms of consciousness should specifically target the L5p cells.

This work will also inform the measures and theories of consciousness. In healthy subjects, it has been possible to show that in deep sleep (Massimini et al., 2005) and anesthesia (Ferrarelli et al., 2010) there is a breakdown of effective connectivity. These studies used a combination of high-density EEG with transcranial magnetic stimulation (TMS). By perturbing the cortex focally with TMS during conscious wakefulness the global deterministic brain response reflects high effective connectivity, i.e., the interactions between brain regions are invariably complex and extend over space and time. When the same stimulation is performed during NREM sleep, however, a qualitatively different pattern emerges. Now the brain's response is dominated by bistability and even if the local bistable response to perturbation is very strong, it fails to engage the rest of the brain in complex interaction (Massimini et al., 2005; Ferrarelli et al., 2010).

These two reactivity patterns can be formally quantified by the perturbational complexity index (PCI; Casali et al., 2013)—arguably the most successful neuronal measure of conscious state so far. PCI is inspired by integrated information theory of consciousness (IIT; Tononi, 2004; Tononi et al., 2016) and aims to capture both segregation and integration of neural processing in one measure. It can differentiate between the conscious state and unconscious state, generalizing to different types of anesthesia, deep sleep and disorders of consciousness (Casali et al., 2013). Based on the present hypothesis we predict that the central reason for why information integration breaks down in the unconscious state is to be found on the level of L5p neurons and their interactions with the thalamo-cortical system. Hence, we hope that the present work can lead to a better understanding of the neurobiological mechanisms underlying the measures of consciousness.

## LIMITATIONS

The hypothesis presented in this work has several limitations mainly related to the anatomy of the cortico-thalamo-cortical circuits. First, not much is known about human L5p cells. Recent work (Beaulieu-Laroche et al., 2018) has demonstrated that dendritic integration is functionally more segregated in

human as compared to rat L5p neurons. Most importantly for the current purposes, virtually nothing is known about the projection patterns of human L5p neurons. We do not know whether similarly to the rodent there are two classes of L5p neurons that project either to other cortical areas or NSP thalamus (for rodent data see Larsen et al., 2008; Kawaguchi, 2017; Takahashi et al., under revision). Most likely there are even more than simply these two classes. This brings us to the second problem: if there are distinct classes of L5p neurons that participate in cortico-cortical and cortico-thalamo-cortical loops, then is it even meaningful to suggest that there is an intersection of these loops? Which of these cell types (e.g., L5A or L5B pyramidal cells) are then crucially related to consciousness? If we would have to pick one specific type of neuron, then it would be the thick-tufted L5B neurons, which are the major output neurons of the cortex, project heavily to NSP thalamus and have also been historically related to consciousness. Even if it would turn out that these cells do not send long-range cortico-cortical projections, it seems that they do *receive* such long-range cortico-cortical projections, especially on their apical dendrites (Cauller et al., 1998; Feldmeyer, 2012; Lee et al., 2013). We note that the direct evidence for this last claim is scarce, but we hope that the next years will bring clarity about this issue.

Third, the thalamic projection patterns as presented here are necessarily simplified. For clarity and simplicity, we mainly adhered to the classic distinction between SP and NSP thalamus (as in functional terms this distinction has not been invalidated). However, this distinction is only part of the story, with no clear consensus on how to classify thalamo-cortical projections (Jones, 2012 advocates the distinction of thalamic core vs. matrix cells; however, this distinction also seems too simplistic, see Clascá et al., 2016).

## CONCLUSION

The contents of consciousness depend on the state of consciousness. Here, we argued that there is a clear neurobiological reason for this intertwining: L5p cells are in the center of both cortico-cortical and thalamo-cortical loops, hence functionally coupling the state and contents of consciousness. The main message going forward is that given the advances in understanding the neural computations, more attention should be given to the relationship between consciousness and the L5p cells. After all, Santiago Ramón y Cajal called them the “psychic cells.”

## DATA AVAILABILITY

No datasets were generated or analyzed for this study.

## AUTHOR CONTRIBUTIONS

JA and TB wrote the initial version of the manuscript. All authors read and commented on the manuscript.

## FUNDING

This work was supported by Grants IUT20-40 and PUT1476 from the Estonian Ministry of Education and Research (Eesti Teadusagentuur). JA was also supported by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 799411. MS was supported by German Research Foundation EXC 257. ML was supported by the European Union's Horizon 2020 research and innovation program and Euratom research and training program 2014-2018 under Grant Agreement No. 670118. This project has received funding from the European

Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 720270 (Human Brain Project SGA1 and SGA2) and from Deutsche Forschungsgemeinschaft (Grant No. LA 3442/3-1 and Grant No. LA 3442/5-1).

## ACKNOWLEDGMENTS

We are indebted to Timothy Zolnik for reading and commenting a draft of this article. We thank Prof. Igor Timofeev for kind assistance with the publication process.

## REFERENCES

- Alkire, M. T., Haier, R. J., and Fallon, J. H. (2000). Toward a unified theory of narcosis: brain imaging evidence for a thalamocortical switch as the neurophysiologic basis of anesthetic-induced unconsciousness. *Conscious. Cogn.* 9, 370–386. doi: 10.1006/ccog.1999.0423
- Alkire, M. T., Hudetz, A. G., and Tononi, G. (2008). Consciousness and anesthesia. *Science* 322, 876–880. doi: 10.1126/science.1149213
- Angel, A. (1993). Central neuronal pathways and the process of anaesthesia. *Br. J. Anaesth.* 71, 148–163. doi: 10.1093/bja/71.1.148
- Aru, J., and Bachmann, T. (2017). Expectation creates something out of nothing: the role of attention in iconic memory reconsidered. *Conscious. Cogn.* 53, 203–210. doi: 10.1016/j.concog.2017.06.017
- Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746. doi: 10.1016/j.neubiorev.2011.12.003
- Aru, J., Tulver, K., and Bachmann, T. (2018). It's all in your head: expectations create illusory perception in a dual-task setup. *Conscious. Cogn.* 65, 197–208. doi: 10.1016/j.concog.2018.09.001
- Bachmann, T. (1984). The process of perceptual retouch: nonspecific afferent activation dynamics in explaining visual masking. *Percept. Psychophys.* 35, 69–84. doi: 10.3758/bf03205926
- Bachmann, T. (1989). *Psychophysiological Bases of Visual Masking (in Russian)*. Tartu: Tartu University Press.
- Bachmann, T. (1994). *Psychophysiology of Visual Masking. The Fine Structure of Conscious Experience*. Commack, NY: Nova Science Publishers, Inc.
- Bachmann, T. (1999). "Twelve spatiotemporal phenomena and one explanation," in *Cognitive Contributions to the Perception of Spatial and Temporal Events*, eds G. Aschersleben, T. Bachmann and J. Müsseler (Amsterdam: Elsevier/North-Holland), 173–206.
- Bachmann, T. (2000). *Microgenetic Approach to the Conscious Mind*. Amsterdam/Philadelphia: John Benjamins.
- Bachmann, T. (2012). How to begin to overcome the ambiguity present in differentiation between contents and levels of consciousness? *Front. Psychol.* 3:82. doi: 10.3389/fpsyg.2012.00082
- Bachmann, T., and Francis, G. (2013). *Visual Masking: Studying Perception, Attention, and Consciousness*. San Diego, CA: Academic Press.
- Bachmann, T., and Hudetz, A. G. (2014). It is time to combine the two main traditions in the research on the neural correlates of consciousness:  $C = L \times D$ . *Front. Psychol.* 5:940. doi: 10.3389/fpsyg.2014.00940
- Beaulieu-Laroche, L., Toloza, E. H., van der Goes, M. S., Lafourcade, M., Barnagian, D., Williams, Z. M., et al. (2018). Enhanced dendritic compartmentalization in human cortical neurons. *Cell* 175, 643–651. doi: 10.1016/j.cell.2018.08.045
- Bogen, J. E. (1995a). On the neurophysiology of consciousness: I. An overview. *Conscious. Cogn.* 4, 52–62. doi: 10.1006/ccog.1995.1003
- Bogen, J. E. (1995b). On the neurophysiology of consciousness: Part II. Constraining the semantic problem. *Conscious. Cogn.* 4, 137–158. doi: 10.1006/ccog.1995.1020
- Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., and Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *Neuron* 37, 9603–9613. doi: 10.1523/jneurosci.3218-16.2017
- Brazier, M. A. B. (1977). *Electrical Activity of the Nervous System*. London: Pitman.
- Bremer, F. (1935). Cerveau "isole" et physiologie du sommeil. *C. R. Soc. Bio.* 118, 1235–1241.
- Brooks, B., and Jung, R. (1973). "Neuronal physiology of the visual cortex," in *Handbook of Sensory Physiology. Central Processing of Visual Information. Part B. (Vol. VII/3)*, ed. R. Jung (New York, NY: Springer-Verlag), 325–440.
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* 5, 198ra105–198ra105. doi: 10.1126/scitranslmed.3006294
- Cauler, L. J., Clancy, B., and Connors, B. W. (1998). Backward cortical projections to primary somatosensory cortex in rats extend long horizontal axons in layer I. *J. Comp. Neurol.* 390, 297–310. doi: 10.1002/(sici)1096-9861(19980112)390:2<297::aid-cne11>3.3.co;2-0
- Clascá, F., Porrero, C., Galazo, M. J., Rubio-Garrido, P., and Evangelio, M. (2016). "Anatomy and development of multispecific thalamocortical axons: implications for cortical dynamics and evolution," in *Axons and Brain Architecture*, ed. S. R. Kathleen (Cambridge, MA: Academic Press), 69–92.
- Crick, F. (1996). Visual perception: rivalry and consciousness. *Nature* 379, 485–486. doi: 10.1038/379485a0
- Crick, F., and Koch, C. (1992). The problem of consciousness. *Sci. Am.* 267, 152–159. doi: 10.1038/scientificamerican0992-152
- De Graaf, T. A., Hsieh, P. J., and Sack, A. T. (2012). The 'correlates' in neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 191–197. doi: 10.1016/j.neubiorev.2011.05.012
- Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018
- Dehaene, S., and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79, 1–37. doi: 10.1016/s0010-0277(00)00123-2
- Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.* 4, 752–758. doi: 10.1038/89551
- Doty, R. W. (1970). "Modulation of visual input by brain-stem system," in *Early Experience and Visual Information Processing in Perceptual and Reading Disorders*, eds F. A. Young and D. B. Lindsay (Washington, DC: National Academy of Sciences), 143–150.
- Feldmeyer, D. (2012). Excitatory neuronal connectivity in the barrel cortex. *Front. Neuroanat.* 6:24. doi: 10.3389/fnana.2012.00024
- Ferrarelli, F., Massimini, M., Sarasso, S., Casali, A., Riedner, B. A., Angelini, G., et al. (2010). Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proc. Natl. Acad. Sci. U S A* 107, 2681–2686. doi: 10.1073/pnas.0913008107
- Harris, K. D., and Shepherd, G. M. (2015). The neocortical circuit: themes and variations. *Nat. Neurosci.* 18, 170–181. doi: 10.1038/nn.3917
- Hassler, R. (1978). "Interaction of reticular activating system for vigilance and the thalamocortical and pallidal systems for directing awareness and attention

- under striatal control,” in *Cerebral Correlates of Conscious Experience*, eds P. A. Buser and A. Rougeul-Buser (Amsterdam: North-Holland), 111–129.
- Hohwy, J. (2009). The neural correlates of consciousness: new experimental approaches needed? *Conscious. Cogn.* 18, 428–438. doi: 10.1016/j.concog.2009.02.006
- Honjoh, S., Sasai, S., Schiereck, S. S., Nagai, H., Tononi, G., and Cirelli, C. (2018). Regulation of cortical activity and arousal by the matrix cells of the ventromedial thalamic nucleus. *Nat. Commun.* 9:2100. doi: 10.1038/s41467-018-04497-x
- Hudetz, A. G. (2012). General anesthesia and human brain connectivity. *Brain Connect.* 2, 291–302. doi: 10.1089/brain.2012.0107
- Jones, E. G. (2012). *The Thalamus*. New York, NY: Springer Science and Business Media.
- Jung, R. (1958). “Coordination of specific and nonspecific afferent impulses at single neurons of the visual cortex,” in *Reticular Formation of the Brain*, ed. H. H. Jasper (Boston, MA: Little, Brown and Co), 423–434.
- Kawaguchi, Y. (2017). Pyramidal cell subtypes and their synaptic connections in layer 5 of rat frontal cortex. *Cereb. Cortex* 27, 5755–5771. doi: 10.1093/cercor/bhx252
- Kim, C. Y., and Blake, R. (2005). Psychophysical magic: rendering the visible ‘invisible’. *Trends Cogn. Sci.* 9, 381–388. doi: 10.1016/j.tics.2005.06.012
- Kinney, H. C., Korein, J., Panigrahy, A., Dikkes, P., and Goode, R. (1994). Neuropathological findings in the brain of Karen Ann Quinlan—the role of the thalamus in the persistent vegetative state. *N. Engl. J. Med.* 330, 1469–1475. doi: 10.1056/nejm199405263302101
- Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.61
- Kouider, S., and Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 857–875. doi: 10.1098/rstb.2007.2093
- LaBerge, D. (1997). Attention, awareness and the triangular circuit. *Conscious. Cogn.* 6, 149–181. doi: 10.1006/ccog.1997.0305
- LaBerge, D. (2006). Apical dendrite activity in cognition and consciousness. *Conscious. Cogn.* 15, 235–257. doi: 10.1016/j.concog.2005.09.007
- Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci.* 7, 12–18. doi: 10.1016/s1364-6613(02)00013-x
- Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cogn. Neurosci.* 1, 204–220. doi: 10.1080/17588921003731586
- Larkum, M. E. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* 36, 141–151. doi: 10.1016/j.tins.2012.11.006
- Larkum, M. E., Zhu, J. J., and Sakmann, B. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature* 398, 338–341. doi: 10.1038/18686
- Larsen, D. D., Wickersham, I. R., and Callaway, E. M. (2008). Retrograde tracing with recombinant rabies virus reveals correlations between projection targets and dendritic architecture in layer 5 of mouse barrel cortex. *Front. Neural Circuits* 2:5. doi: 10.3389/neuro.04.005.2007
- Lau, H., and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci.* 15, 365–373. doi: 10.1016/j.tics.2011.05.009
- Laureys, S. (2005). The neural correlate of (un) awareness: lessons from the vegetative state. *Trends Cogn. Sci.* 9, 556–559. doi: 10.1016/j.tics.2005.10.010
- Lee, S., Kruglikov, I., Huang, Z. J., Fishell, G., and Rudy, B. (2013). A disinhibitory circuit mediates motor integration in the somatosensory cortex. *Nat. Neurosci.* 16, 1662–1670. doi: 10.1038/nn.3544
- Leopold, D. A., and Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys’ percepts during binocular rivalry. *Nature* 379, 549–553. doi: 10.1038/379549a0
- Lisman, J. E. (1997). Bursts as a unit of neural information: making unreliable synapses reliable. *Trends Neurosci.* 20, 38–43. doi: 10.1016/s0166-2236(96)10070-9
- Livingstone, M. S., and Hubel, D. H. (1980). Evoked responses and spontaneous activity of cells in the visual cortex during waking and slow-wave sleep. *ARVO 1980. Supplement to Investigative Ophthalmology and Visual Science*, 223.
- Llinás, R. (1996). “Content and context in the thalamocortical system: the basis for cognition,” in *Paper Presented at Toward a Science of Consciousness 1996. “Tucson II,” International Conference, April 8–13, Tucson, Arizona*, 8–13.
- Llinás, R., and Ribary, U. (2001). Consciousness and the brain. The thalamocortical dialogue in health and disease. *Ann. N. Y. Acad. Sci.* 929, 166–175. doi: 10.1111/j.1749-6632.2001.tb05715.x
- Lumer, E. D., Friston, K. J., and Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science* 280, 1930–1934. doi: 10.1126/science.280.5371.1930
- Magnin, M., Rey, M., Bastuji, H., Guillemant, P., Mauguier, F., and Garcia-Larrea, L. (2010). Thalamic deactivation at sleep onset precedes that of the cerebral cortex in humans. *Proc. Natl. Acad. Sci. U S A* 107, 3829–3833. doi: 10.1073/pnas.0909710107
- Magoun, H. W. (1958). *The Waking Brain*. Springfield: C. C. Thomas.
- Martens, S., and Wyble, B. (2010). The attentional blink: past, present, and future of a blind spot in perceptual awareness. *Neurosci. Biobehav. Rev.* 34, 947–957. doi: 10.1016/j.neubiorev.2009.12.005
- Mashour, G. A. (2014). Top-down mechanisms of anesthetic-induced unconsciousness. *Front. Syst. Neurosci.* 8:115. doi: 10.3389/fnsys.2014.00115
- Mashour, G. A., and Hudetz, A. G. (2017). Bottom-up and top-down mechanisms of general anesthetics modulate different dimensions of consciousness. *Front. Neural Circuits* 11:44. doi: 10.3389/fncir.2017.00044
- Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. doi: 10.1126/science.1117256
- Mesulam, M.-M. (1985). *Principles of Behavioral Neurology*. Philadelphia, PA: F. A. Davis.
- Meyer, K. (2015). The role of dendritic signaling in the anesthetic suppression of consciousness. *Anesthesiology* 122, 1415–1431. doi: 10.1097/ALN.0000000000000673
- Mitchell, D. B. (2006). Nonconscious priming after 17 years: invulnerable implicit memory? *Psychol. Sci.* 17, 925–929. doi: 10.1111/j.1467-9280.2006.01805.x
- Mo, C., and Sherman, S. M. (2019). A sensorimotor pathway via higher-order thalamus. *J. Neurosci.* 39, 692–704. doi: 10.1523/JNEUROSCI.1467-18.2018
- Moruzzi, G., and Magoun, H. W. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalogr. Clin. Neurophysiol.* 1, 455–473. doi: 10.1016/0013-4694(49)90066-8
- Munk, M. H. J., Roelfsema, P. R., König, P., Engel, A. K., and Singer, W. (1996). Role of reticular activation in the modulation of intracortical synchronization. *Science* 272, 271–274. doi: 10.1126/science.272.5259.271
- Murayama, M., and Larkum, M. E. (2009). Enhanced dendritic activity in awake rats. *Proc. Natl. Acad. Sci. U S A* 106, 20482–20486. doi: 10.1073/pnas.0910379106
- Newman, J. (1995a). Thalamic contributions to attention and consciousness. *Conscious. Cogn.* 4, 172–193. doi: 10.1006/ccog.1995.1024
- Newman, J. (1995b). Reticular-thalamic activation of the cortex generates conscious contents. *Behav. Brain Sci.* 18, 691–692. doi: 10.1017/s0140525x00040577
- Odegaard, B., Knight, R. T., and Lau, H. (2017). Should a few null findings falsify prefrontal theories of conscious perception? *J. Neurosci.* 37, 9593–9602. doi: 10.1523/JNEUROSCI.3217-16.2017
- Penfield, W. (1975). *The Mystery of the Mind: A Critical Study of Consciousness and the Human Brain*. Toledo, OH: Princeton University Press.
- Phillips, W. A., Bachmann, T., and Storm, J. F. (2018). Apical function in neocortical pyramidal cells: a common pathway by which general anesthetics can affect mental state. *Front. Neural Circuits* 12:50. doi: 10.3389/fncir.2018.00050
- Purpura, D. P. (1970). “Operations and processes in thalamic and synaptically related neural subsystems,” in *The Neurosciences. Second Study Program*, ed. F. O. Schmitt (New York, NY: Rockefeller University Press), 458–470.
- Purpura, K. P., and Schiff, N. D. (1997). The thalamic intralaminar nuclei: a role in visual awareness. *Neuroscientist* 3, 8–15. doi: 10.1177/107385849700300110
- Rees, G., Kreiman, G., and Koch, C. (2002). Neural correlates of consciousness in humans. *Nat. Rev. Neurosci.* 3, 261–270. doi: 10.1038/nrn783

- Riener, A. (2017). "Subliminal perception or 'can we perceive and be influenced by stimuli that do not reach us on a conscious level?'" in *Emotions and Affect in Human Factors and Human-Computer Interaction*, ed. M. Jeon (San Diego, CA: Elsevier Academic Press), 503–538.
- Riklan, M., and Levita, E. (1969). *Subcortical Correlates of Human Behavior. A Psychological Study of Thalamic and Basal Ganglia Surgery*. Baltimore, MD: Williams and Wilkins.
- Schiff, N. D. (2010). Recovery of consciousness after brain injury: a mesocircuit hypothesis. *Trends Neurosci.* 33, 1–9. doi: 10.1016/j.tins.2009.11.002
- Sheinberg, D. L., and Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. U S A* 94, 3408–3413. doi: 10.1073/pnas.94.7.3408
- Sherman, S. M., and Guillery, R. W. (2001). *Exploring the Thalamus*. San Diego, CA: Academic Press.
- Singer, W. (1979). "Central core control of visual cortex functions," in *The Neurosciences. Fourth Study Program*, eds F. O. Schmitt and F. G. Worden (Cambridge, MA: MIT), 1093–1110.
- Smirnov, V. M., Muchnik, L. S., and Shandurina, A. N. (1978). "The outline and the functions of the deep structures of the brain (in Russian)," in *Estestvennonauchnye osnovy psikhologii*, eds A. A. Smirnov, A. R. Luria and V. D. Nebylitsyn (Moscow: Pedagogika), 76–108.
- Somjen, G. (1972). *Sensory Coding in the Mammalian Nervous System*. New York, NY: Appleton-Century-Crofts.
- Steriade, M. (1981a). "Mechanisms underlying cortical activation: neuronal organization and properties of the midbrain reticular core and intralaminar thalamic nuclei," in *Brain Mechanisms and Perceptual Awareness*, eds O. Pompeiano and C. Ajmone Marsan (New York, NY: Raven), 327–377.
- Steriade, M. (1981b). EEG desynchronisation is associated with cellular events that are prerequisites for active behavioral states. *Behav. Brain Sci.* 4, 489–492. doi: 10.1017/s0140525x00010037
- Suzuki, M., and Larkum, M. E. (2017). Dendritic calcium spikes are clearly detectable at the cortical surface. *Nat. Commun.* 8:276. doi: 10.1038/s41467-017-00282-4
- Takahashi, N., Ebner, C., Sigl-Glöckner, J., Nierwetberg, S. and Larkum, M. E. (under revision). Active dendritic currents gates descending cortical outputs in perception.
- Takahashi, N., Oertner, T. G., Hegemann, P., and Larkum, M. E. (2016). Active cortical dendrites modulate perception. *Science* 354, 1587–1590. doi: 10.1126/science.aah6066
- Theyel, B. B., Llano, D. A., and Sherman, S. M. (2010). The corticothalamocortical circuit drives higher-order cortex in the mouse. *Nat. Neurosci.* 13, 84–88. doi: 10.1038/nn.2449
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.* 5:42. doi: 10.1186/1471-2202-5-42
- Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461. doi: 10.1038/nrn.2016.44
- Tononi, G., and Edelman, G. M. (1998). Consciousness and complexity. *Science* 282, 1846–1851. doi: 10.1126/science.282.5395.1846
- Tulving, E., Schacter, D. L., and Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 8, 336–342. doi: 10.1037/0278-7393.8.4.336
- Ward, L. M. (2011). The thalamic dynamic core theory of conscious experience. *Conscious. Cogn.* 20, 464–486. doi: 10.1016/j.concog.2011.01.007
- Watson, A. B. (1986). "Temporal sensitivity," in *Handbook of Perception and Human Performance. Vol. 1. Sensory Processes and Perception*, eds K. R. Boff, L. Kaufman and J. P. Thomas (New York, NY: Wiley), 6/1–6/43.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Aru, Suzuki, Rutiku, Larkum and Bachmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Sleep–Wake Cycle in Young and Older Mice

Sara Soltani<sup>1,2</sup>, Sylvain Chauvette<sup>2</sup>, Olga Bukhtiyarova<sup>1,2</sup>, Jean-Marc Lina<sup>3,4</sup>, Jonathan Dubé<sup>3,5</sup>, Josée Seigneur<sup>2</sup>, Julie Carrier<sup>3,5</sup> and Igor Timofeev<sup>1,2\*</sup>

<sup>1</sup> Department of Psychiatry and Neuroscience, Faculty of Medicine, Université Laval, Québec, QC, Canada, <sup>2</sup> CERVO Brain Research Centre, Québec, QC, Canada, <sup>3</sup> Center for Advanced Research in Sleep Medicine, Centre Intégré Universitaire de Santé et de Services Sociaux du Nord-de-l'Île de Montréal, Montreal, QC, Canada, <sup>4</sup> École de Technologie Supérieure, Montreal, QC, Canada, <sup>5</sup> Department of Psychology, Université de Montréal, Montreal, QC, Canada

Sleep plays a key role in multiple cognitive functions and sleep pattern changes with aging. Human studies revealed that aging decreases sleep efficiency and reduces the total sleep time, the time spent in slow-wave sleep (SWS), and the delta power (1–4 Hz) during sleep; however, some studies of sleep and aging in mice reported opposing results. The aim of our work is to estimate how features of sleep–wake state in mice during aging could correspond to age-dependent changes observed in human. In this study, we investigated the sleep/wake cycle in young (3 months old) and older (12 months old) C57BL/6 mice using local-field potentials (LFPs). We found that older adult mice sleep more than young ones but only during the dark phase of sleep–wake cycle. Sleep fragmentation and sleep during the active phase (dark phase of cycle), homologous to naps, were higher in older mice. Older mice show a higher delta power in frontal cortex, which was accompanied with similar trend for age differences in slow wave density. We also investigated regional specificity of sleep–wake electrographic activities and found that globally posterior regions of the cortex show more rapid eye movement (REM) sleep whereas somatosensory cortex displays more often SWS patterns. Our results indicate that the effects of aging on the sleep–wake activities in mice occur mainly during the dark phase and the electrode location strongly influence the state detection. Despite some differences in sleep–wake cycle during aging between human and mice, some features of mice sleep share similarity with human sleep during aging.

**Keywords:** sleep–wake cycle, LFP, delta power, SWS, REM, wake, sleep fragmentation

## OPEN ACCESS

### Edited by:

Ilan Lampl,  
Weizmann Institute of Science, Israel

### Reviewed by:

M. Gustavo Murer,  
University of Buenos Aires, Argentina  
Michele Bellesi,  
University of Bristol, United Kingdom

### \*Correspondence:

Igor Timofeev  
Igor.Timofeev@fmed.ulaval.ca

**Received:** 09 May 2019

**Accepted:** 09 September 2019

**Published:** 24 September 2019

### Citation:

Soltani S, Chauvette S, Bukhtiyarova O, Lina J-M, Dubé J, Seigneur J, Carrier J and Timofeev I (2019) Sleep–Wake Cycle in Young and Older Mice. *Front. Syst. Neurosci.* 13:51. doi: 10.3389/fnsys.2019.00051

## INTRODUCTION

Sleep patterns change throughout life. In humans, the highest density of slow waves moves from the occipital cortex in preschoolers to centro-parietal areas during adolescence, and to frontal areas in adulthood (Kurth et al., 2010). Middle-age subjects show a reduced sleep efficiency, duration, slow-wave density, and a reduced amplitude of slow waves compared to young adults (Carrier et al., 2011; Mander et al., 2017). However, mice studies showed conflicting results. A study compared EEG recordings above the somatosensory cortex of 6 months old to 18–24 months old C57BL/6J OlaHsd male mice and found an increase in the sleep amount in the older group as well as an increase in the delta range power (Panagiotou et al., 2017). Another group found that C57BL/6 mice show only a non-significant trend for an increase in their daily amount of non-REM sleep from 3 to

6 to 12 months old, however showing a significant decline at 2 years old compared to 1 year old (Hasan et al., 2012).

Another study performed in C57BL/6J mice with LFP recordings from motor and visual cortices in 4 months and a half, 1 year old, and 2 years old mice showed a linear increase in the total sleep time with age (McKillop et al., 2018). The same study also showed that the amplitude of slow waves was larger in the younger group compared to the two older groups, but the slow wave incidence was only significantly increased in the 2 years old group (McKillop et al., 2018). It was also demonstrated that the increase in total sleep in aging mice depend on their inability to maintain prolonged wakefulness (Wimmer et al., 2013).

As different mice lines show different sleep patterns and are differently affected by age (Franken et al., 1998; Hasan et al., 2012), it becomes important to compare results of studies performed in the same mouse line.

Sleep was long thought to be a global state; however, it was shown both in humans (Nir et al., 2011) and in rats (Vyazovskiy et al., 2011) that sleep can be local. A study performed in Long Evans rats showed a high congruence of slow-wave sleep (SWS) using frontal and parietal EEGs as well as mPFC and hippocampal LFPs; however, congruence was lower for REM sleep and even much lower during transition states (Durán et al., 2018). A study conducted in C57BL/6 mice concluded that slow waves occur regularly during REM sleep especially in primary sensory and motor cortices, and mainly in layer 4 (Funk et al., 2016). A vast majority of cortical projections in carnivores and primates, but not in rodents are patchy, local, and intracortical (Gilbert and Wiesel, 1983; Rockland, 1985; Lund et al., 1993, 1994; Schuz et al., 2006; Van Hooser et al., 2006). Such difference in corticocortical circuitry could explain regional differences in the expression of sleep in mice, specifically slow waves which should be expressed more locally throughout the cortex. While local slow waves during REM sleep were also reported in human (Bernardi et al., 2019), it would be expected that slow waves are more local in mice than in humans.

We hypothesized that the different cerebral areas reported in the above-mentioned studies done on male rodents might explain the conflicting results on aging reported in mice. We propose that mice might not be the best model of human sleep in aging, but some features will reproduce changes observed in aging humans.

Here we investigated the daily distribution of sleep–wake states, the delta power modulation, and slow waves parameters in male C57BL/6 young (3 months old) and older (12 months old) mice. We found that 1-year-old C57BL/6 mice sleep significantly more than 3 months old mice over 24 h and these differences are especially significant during the dark phase. We also report that some cerebral areas will consistently reveal less REM sleep than others as they remain in deep SWS while most areas display REM oscillations. The SWS delta power changed with aging, but these differences were topographically specific. Older mice revealed a trend of increased incidence of slow waves only in the frontal cortex. Older mice also demonstrated significant increase in sleep fragmentation. We conclude that only some features of human sleep in aging could be effectively investigated in mice.

## MATERIALS AND METHODS

### Experiments and Animals

All experiments were performed in accordance with the guideline of the Canadian Council on Animal Care and approved by the Université Laval Committee on Ethics and Animal Research. Experiments were performed on young (3 months old) male C57BL/6 mice ( $n = 8$ ) and older (12 months old) male C57BL/6 mice ( $n = 9$ ) to compare the sleep characteristics of the two groups of age. We analyzed recordings for 10–25 consecutive days in each animal obtained at least 1 week after electrode implantation and beginning of continuous recording to allow animals to recover from anesthesia and to adapt to a tethered condition.

### Sterile Surgery and Electrodes Implantation

All surgeries were performed under sterile conditions. The mice were first anesthetized with 1–2% isoflurane, the head was shaved and then fixed in the stereotaxic frame. Subcutaneous injection of buprenorphine (0.1 mg/kg) was applied for analgesia and saline (0.9% NaCl) (s.c.). The incision site and all pressure points were injected with a mixture of Bupivacaine (0.25%)/Lidocaine (0.5%). The head was cleaned with three passages of chlorhexidine (0.5%) in alternation with alcohol (70%) before the skin incision. After opening the skin above the skull, three alternating passages of a bleach solution (0.03% sodium hypochlorite) and hydrogen peroxide (3%) were used to clean the skull. Small holes were drilled for the reference, anchoring screws, and LFP electrodes. We used custom-made electrodes (stainless steel wires, 125  $\mu$ m diameter, Perfluoroalkoxy (PFA)-insulated) and in the majority of experiments they were implanted in three different cortical regions to a depth of 600  $\mu$ m from the cortical surface: frontal cortex (AP: +2.6 mm; ML: –1.5 mm); somatosensory cortex (AP: –0.94 mm; ML: –3 mm, and AP: –1.75 mm, ML: –2 mm). In separate experiments ( $n = 2$ ) we implanted 14 LFP electrodes (**Figure 6**) to obtain multisite LFP recordings from the major part of the dorsal cortical surface. One stainless steel screw over the cerebellum was used as a reference and four anchoring screws (two screws on each side of the skull) were used to secure the implant. Two 75  $\mu$ m electrodes [single-stranded stainless-steel wire (PFA-insulated)], were inserted into the neck muscle to record the electromyogram (EMG) activity. All LFPs, EMG electrodes, and the reference electrode were connected to a Nano-Miniature Omnetic Connector and were covered and fixed with dental acrylic (Dentsply Canada). Mice received subcutaneous injections of Anafen (5 mg/kg) and saline (0.5–1 ml) for 2 days post-surgery.

### Recordings

The recordings were done within the standard animal facility settings. The standard mice cages were modified to enable the passage of a tethering cable connecting the animal. Miniature custom-made buffer pre-amplifier (voltage amplification coefficient 1, but it amplifies current) was attached directly to the head and it was used to reduce movement artifacts in cables (not shown). These pre-amplifiers were connected to

the commercially available AM amplifiers (Model 3000, A-M system, Sequim, WA, United States). The signals were hardware band-pass filtered between 1–100 Hz (LFP) and 10–300 Hz (EMG), the notch filter was used at 60 Hz. The light–dark cycle was 12 h/12 h with the lights off at 19:00. LFPs and EMG were continuously recorded (sampling rate 1 kHz) with LabChart (AD instruments, Colorado Springs, CO, United States) 24 h per day for at least 3 weeks. The recordings started at the end of the surgery; however, we analyzed only data obtained at least 5 days after the end of anesthesia.

## Automatic States of Vigilance Identification

We used custom-written routine in Igor Pro (version 6, Wavemetrics, OR, United States) or MATLAB (Version R2015b) to automatically detect states of vigilances. Fast Fourier Transform (FFT) was calculated using 5 s time window for all LFP channels recorded as well as for EMG (**Figure 1A**). A brief comparison of the detection with two different time-windows (5 and 30 s, not shown) revealed that longer time-windows missed multiple microarousal periods (Watson et al., 2016). Therefore, we analyzed all data with 5 s windows with a sliding time window of 1 s for computing the power of activities of interest. First, we run Fourier analysis with 0.1 Hz resolution from which we obtained absolute values of Fourier amplitude. We then defined the delta power as the area between 0.2 and 4 Hz (**Figure 1B**), the theta power as the area between 5 and 9 Hz, and the muscle power as the area between 10 and 300 Hz. The power was initially computed in each channel, but then we realized that on multiple occasions, higher values of power of different frequencies could co-occur with lower values in other channels. To ensure the stability of state detection, in most of analysis, the delta and theta power was computed in each channel and then it was summated across channels. As the theta power alone was not always a very good indication of REM sleep (Borbély et al., 1984) and it was stronger/weaker depending on the area recorded, therefore, we divided the theta power by the delta power and also by the muscle power. This theta/(delta  $\times$  muscle) ratio was very efficient to detect REM episodes as only period with low delta and low muscle tone would give high values, corresponding to REM sleep. Thresholds for state detection were set manually. Because SWS-delta power is higher during the dark period as opposed to the light period, the thresholds for the dark vs. light periods were set to a different value. Despite the use of our custom-made preamplifier, the movement artifacts could be recorded and at the level of LFP it would produce large values in delta power frequency range. Therefore, the first step was to set a threshold for muscle power, which had bimodal distribution (**Figure 1C**). If the muscle power was high, it was considered as wake. In the remaining segments, we set thresholds for LFP delta power, that also had bimodal distribution (**Figure 1D**). The high values of LFP delta power in the absence of muscle contractions were considered as SWS. The remaining segments were analyzed with the theta/(delta  $\times$  muscle) ratio (**Figure 1E**). If it was high, we consider it as REM sleep, otherwise it was quiet wakefulness. Less than 1% of segments, typically at the transitions between states,

did not fall in these criteria [very low muscle tone power, low delta power, and low theta/(delta  $\times$  muscle) ratio] and we qualified them as undefined state, which was not analyzed further.

## Slow Waves Detection

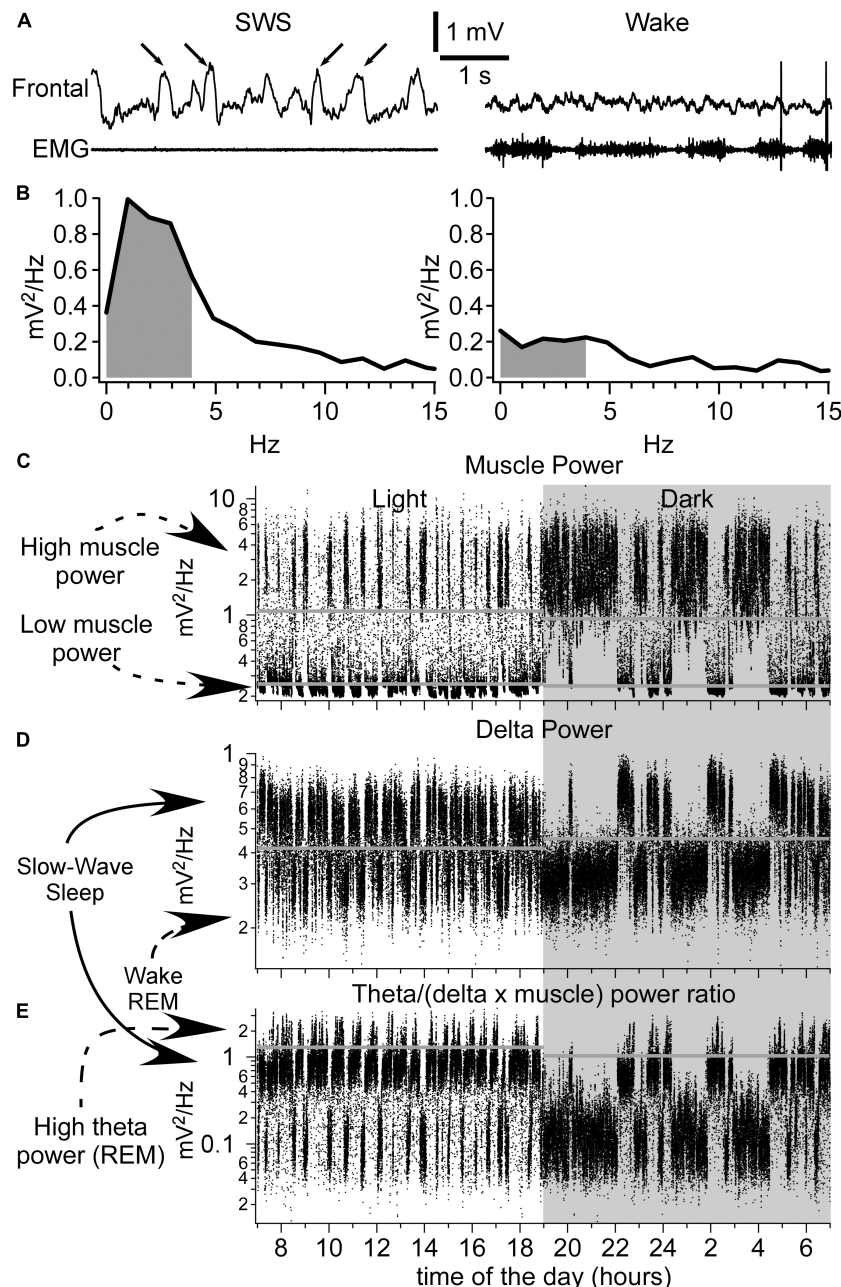
The detection of slow waves was achieved by MATLAB scripts using a pattern recognition method described previously (Bukhtiyarova et al., 2016, 2019). The method allowed us to segment the original signal, preparing input data for neural network training, training the neural network to classify the chosen templates into groups and post-processing of the results. Briefly, we created a library of 8903 LFP segments obtained from nine channels of young mice and six channels of old mice. These segments were manually labeled as “SW” or “Noise.” The library was randomly divided in three parts: 60% of it served as templates to train feed-forward artificial neural network (1 hidden layer of 20 neurons and 1 output layer with 2 possible outputs for “SW” and “Noise”), 20% was used for validation, and 20% to test the quality of detection. The precision of LFP segments classification (positive predictive rate) and recall (true positive rate) for the obtained network was 86.7 and 80% correspondingly for “SW” and 96.2 and 97.7% for “Noise,” the overall accuracy was 94.8%. This network was applied for the entire dataset after its segmentation and was followed with post-processing.

The amplitude of detected slow waves was measured from zero crossing to the maximum of depth-positive wave from traces digitally re-filtered between 1 and 100 Hz. The duration of detected slow waves was measured at half amplitude of the signal digitally filtered between 1 and 4 Hz.

## Statistics

To evaluate the effect of age on different sleep parameters, linear mixed models (LMM) were computed using SPSS25 (IBM). This approach is now recommended over mean-based tests (*t*-tests) in neuroscience to resolve the issue of independence of multiple measures in a single individual by controlling individual variation in each subject (Koerner and Zhang, 2017). Each variables-of-interest was modeled using restricted maximum-likelihood method (REML) as varying according to fixed effects of interest (which will be interpreted and discussed below) and random effects (to correct for intra-individual variability). For each variable, two random factors were taken into account: (1) the intercept across all days recorded for each mouse and (2) the slope across all recorded days. An autoregressive covariance structure was specified in the model to account for covariance between timepoints. In essence, these random effects control for intra-individual variability due to data recorded in a different number of days between mice. Convergence and significance of each model were required before interpreting results. Models converged and were significant ( $p < 0.001$ ) for all variables of interest.

Relative duration (%) of each state was averaged for each recorded hour. Two independent two-way LMMs were computed to model the hourly relative duration of each state (one model per period: light and dark). These models accounted for random effects, modeled *hour* as a repeated fixed effect (12 h) and *age* as an



**FIGURE 1 |** Identification of sleep–wake states. **(A)** A segment of LFP recorded from the frontal cortex and EMG (neck muscle) during slow-wave sleep (left panel, arrows indicate the slow waves) and wake (right panel). Below are the corresponding FFT analysis **(B)**. Gray indicates the area used to calculate the delta power (0.2–4 Hz). **(C)** Muscle power over 24 h. Shaded area indicates the dark period. **(D)** Delta power pattern for 24 h. Each dot represents 5 s of recordings. Note the bimodal distribution pattern of delta power with high delta power values corresponding to slow-wave sleep and low delta power values corresponding to either wake or REM sleep. **(E)** Calculated ratio over 24 h of theta power divided by the delta power and further divided by the muscle power. High ratio values correspond to REM sleep having a high theta power, and low delta and muscle powers. Horizontal gray lines correspond to threshold values set for shown examples. Note that the thresholds were different for recordings obtained during light phase vs. dark phase.

independent fixed effect (*two age group*). Then, relative duration (%) of each state of vigilance (wake, SWS, REM) was averaged over 24 h, dark and light periods, and modeled in independent LMMs with specified random effects and *age* as a fixed effect (*two age groups*).

Two independent three-way LMMs were computed to model hourly delta power in different channels (one model per period: light and dark). These models accounted for random effects, modeled *hour* and *electrode* as repeated fixed effects (*12 hours x 3 electrodes*) and *age* as an independent fixed effect (*two age groups*).



In addition, delta power in SWS was also averaged over 24 h, dark and light periods, and modeled in independent two-way LMMs with random effects, *age* as a fixed effect (two age groups), and *electrode* as a repeated fixed effect (*three electrodes*). Finally, slow wave characteristics (density, amplitude, and duration) were analyzed with two-way LMMs specifying *age* as a fixed effect (*two age group*) and *electrode* as a repeated effect (*three electrodes*).

Interaction terms between all fixed factors were always included in each model. When interaction terms were significant at  $p < 0.05$ , they were decomposed with simple effect tests reported in relevant figures. If not, main effects were interpreted and are reported in figures.

## Histology

The mice were deeply anesthetized with ketamine/xylazine and transcardially perfused with saline followed by 4% paraformaldehyde (PFA). The brains were kept in the PFA solution overnight and were cut coronally in 50- $\mu$ m-thick slice with a vibrotome in PBS solution or after 2 days in a 30% sucrose solution with a microtome. Sections were treated with Nissl staining for the electrode location visualization.

## RESULTS

### State of Vigilance Distribution Over 24 h

With rare exceptions, the same animal had similar daily distribution of sleep–wake states with more sleep during the light period and more wake during the dark period of the dark–light cycle. **Figure 2A** shows a distribution of automatically detected states over 23 consecutive days in one mouse implanted with electrodes at the age of 3 months. Periodical absence of sleep in 1-h segments during light phase was mainly due to environmental factors, such as unusual activities within or around the recording room rather than a natural behavior of mice. We examined the performance of our automatic method of state detection. On six arbitrarily chosen 4-h long segments, the agreement of automatic detection against experienced user detection was  $92.59 \pm 2.97\%$  (not shown). The major disagreement was in the detection of very brief states (typically microarousals) which were often not detected (not considered) by an experienced observer.

We analyzed the automatically detected state distribution for 99 days from eight young mice and 67 days from nine older mice. Out of at least 3 weeks of recording per animal, we typically analyzed 12–14 days from each animal, but up to 23 days, starting from the second week after electrode implantation and connection to recording cables. Some days were skipped because of occasional technical difficulties. The same animal typically demonstrated very similar sleep–wake pattern across days (**Figure 2A**) except occasional long wake states during the day, likely due to unusual environmental conditions (atypical noise in animal facilities). We compared the hourly distribution of state of vigilance for 24 h and the overall distribution of the states in 1 day as well as the sleep pattern according to the light/dark cycle in young and older mice (**Figures 2B–D**). As expected, young mice were more awake than older mice at the beginning of the dark phase of cycle (60–80% of time) compared

to the light phase of the light/dark cycle and with the progression of the dark phase of cycle the sleep proportion increased. The older mice had a more variable sleep–wake behavior. Out of nine older mice, five had 70–95% of wake at the beginning of the dark phase of cycle, but the four other animals had <50% of wake at the same time period (**Figure 2B**). LMM analysis showed the presence of a significant interaction between hour and age for wake time in the dark phase ( $p < 0.05$ ). Simple effects analyses show that young mice were more awake at the beginning of the dark phase of the light/dark cycle (**Figure 2C**), while the older mice were more awake during morning time (at end of the dark phase of cycle). A significant interaction between age and hour was also found for SWS and REM in the dark phase of the light/dark cycle ( $p < 0.05$ ). Simple effect analysis demonstrate that older mice had more sleep than young at the beginning of the dark phase of the light/dark cycle while old had less sleep than young at the end of the dark phase (**Figure 2C**). For the light phase, significant interactions between age and hour were found for wake and REM ( $p < 0.05$ ), but not for SWS. *Post hoc* tests in the light phase showed that old animals had less REM at the beginning of the cycle. Significance was not reached at any time points for *post hoc* of age effects on wake during light phase. Finally, main effect of age on SWS duration was non-significant in the light phase. Overall, older mice had more SWS than young mice and these differences were due to longer time spent in SWS during the dark phase of the light/dark cycle (**Figure 2D**). Overall, these results suggest that differences in sleep–wake behavior of young vs. older mice depend largely on the specific period of the dark–light cycle.

Most states in mice were short. As it can be seen from **Figures 3A–C**, at the beginning of the dark phase of cycle, when the sleep pressure is low, the animal continuously switched between SWS and waking states. The SWS was often interrupted by brief waking episodes characterized by an activated LFP in all channels and increased muscle tone (**Figure 3B**) as well as during wake, we often observed episodes of reduced muscle tone and high power of LFP activities in the low-frequency bands, characteristic for SWS (**Figure 3C**). Using 5 s resolution for state detection, very brief muscle twitches (typically less than a second), even accompanied with activated LFPs were not detected as wake (not shown). During REM sleep, the LFP was typically activated and the muscle tone was mostly absent with the exception of occasional twitches. As a mean, the older mice had more states per day ( $2002 \pm 510.5$ ,  $n = 55$  days) than young mice ( $1659 \pm 219.9$ ,  $n = 79$  days;  $p = 0.0003$ , Mann–Whitney *U*-test). The results show that except wake during the light phase of cycle ( $p = 0.3615$ ) and REM sleep during light phase of cycle ( $p = 0.2057$ ) on a 24-h scale, the older mice had more periods of wake, SWS, and REM ( $p = 0.002$ ,  $p = 0.0002$ , and  $p = 0.0237$ , correspondingly). During the dark phase of the dark/light cycle these differences were highly significant (all  $p < 0.0001$ ) indicating that the states of vigilance in older mice were less consolidated compared to young mice. Overall bimodal distribution of SWS suggests that SWS episodes lasting 10 s and less represent some separate states and can be considered as microsleep states (**Figure 3D**). For the wake, the short episodes had likely two dominant durations with a clear maximum around

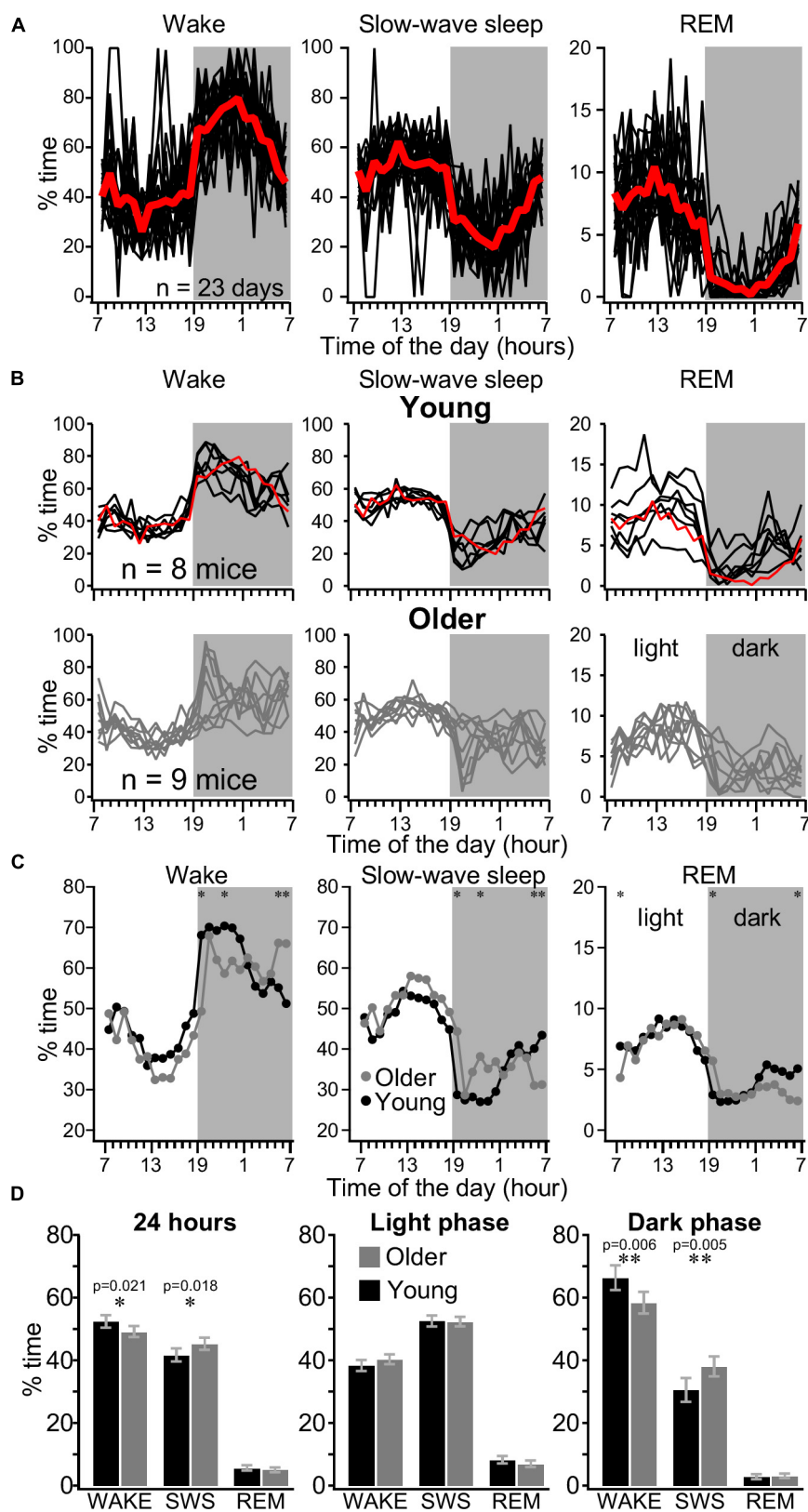


FIGURE 2 | Continued

**FIGURE 2 |** Distribution of sleep–wake states in young and older mice over 24 h. **(A)** Distribution of wake, SWS, and REM sleep recorded over 23 consecutive days with 1-h resolution in individual mice (black lines) and their average (red line). **(B)** Distribution of wake, SWS, and REM sleep in young (upper panels) and older (lower panels) mice. Each line represents averaged values of one animal over several days of recordings, the red line is an average from the example mice shown in panel **(A)**. **(C)** Hourly distribution of states of vigilance for 24 h for young (black circles,  $n = 99$  days from eight mice) and older (gray circles,  $n = 67$  days from nine mice) mice. Shaded area indicates the dark phase. Printed values are estimated marginal means computed in six independent mixed linear models (one for each state of vigilance and phase of light–dark cycle) with age and hour factors. **(D)** 24 h and light/dark cycle distribution of states of vigilance. Printed values are estimated marginal means from independent mixed linear models. Note that older mice show more SWS than young mice over a day, but this difference occurs mainly during the dark phase. Significant statistical differences in each hour are indicated by asterisk,  $*p < 0.05$ ,  $**p < 0.01$ . Error bars indicate the 95% confidence interval.

7 s and less obvious around 20 s (**Figure 3D**). It is possible that these two types of electrophysiological microarousal states represent some different behavioral states.

## Daily SWS Delta Power Dynamics Is Different in Young and Older Mice

During SWS, cortical neurons have a bi-stable behavior alternating between active and silent states and during REM sleep or wake, the membrane potential has a unimodal distribution (Steriade et al., 2001; Timofeev et al., 2001). The LFP/EEG signal is essentially generated by synchronous activities of cortical neurons (Contreras and Steriade, 1995; Chauvette et al., 2010). Therefore, during wake or REM sleep, the ratio of activities below and above 4 Hz (delta range) is nearly 1, but during SWS the activity in the delta range power is six times higher than the faster frequencies power, reflecting the bi-stable behavior of cortical neurons (Mukovski et al., 2007). In our experiments, the overall delta power distribution was bimodal, with low values corresponding to waking state and REM sleep and high values corresponding to SWS (**Figure 1D**). The origin of delta activity during wake or REM sleep is unclear; therefore, we further analyzed the delta power dynamics only during SWS. Previous experiments on cats demonstrated that intracellular sleep slow wave activity had different amplitude in different cortical areas (Chauvette et al., 2011). Therefore, we calculated the mean hourly delta power (0.2–4 Hz) values during SWS from the frontal, somatosensory anterior, and somatosensory posterior channels and computed for the entire day with a 5 s time window sliding by 1 s (**Figure 4**). As expected from sleep homeostasis process (Borbely, 1982; Borbely et al., 1989), in both groups of age, the delta power in all three investigated cortical areas declined gradually during the light period followed by gradual increase in the dark phase of the light/dark cycle in the hourly distribution (**Figure 4A**). LMM analysis did not show a three-way interaction between age, electrode, and time, but we do report the presence of a two-way interaction ( $p < 0.05$ ) between age and electrode in both dark and light period, demonstrating that age changes in delta power are topographically specific but constant across the time of measurement in each light/dark cycle. Moreover, *post hoc* tests showed that older mice have a significantly higher delta power in the frontal cortex, while there was no difference in the somatosensory anterior and posterior cortex for both light and dark phases. That was true both when calculated hourly (**Figure 4A**) and for the whole light/dark cycle or separately for the light and the dark phases (**Figure 4B**). As it is clearly seen in **Figure 4A**, in older mice, the overall delta power was the highest in the frontal cortex,

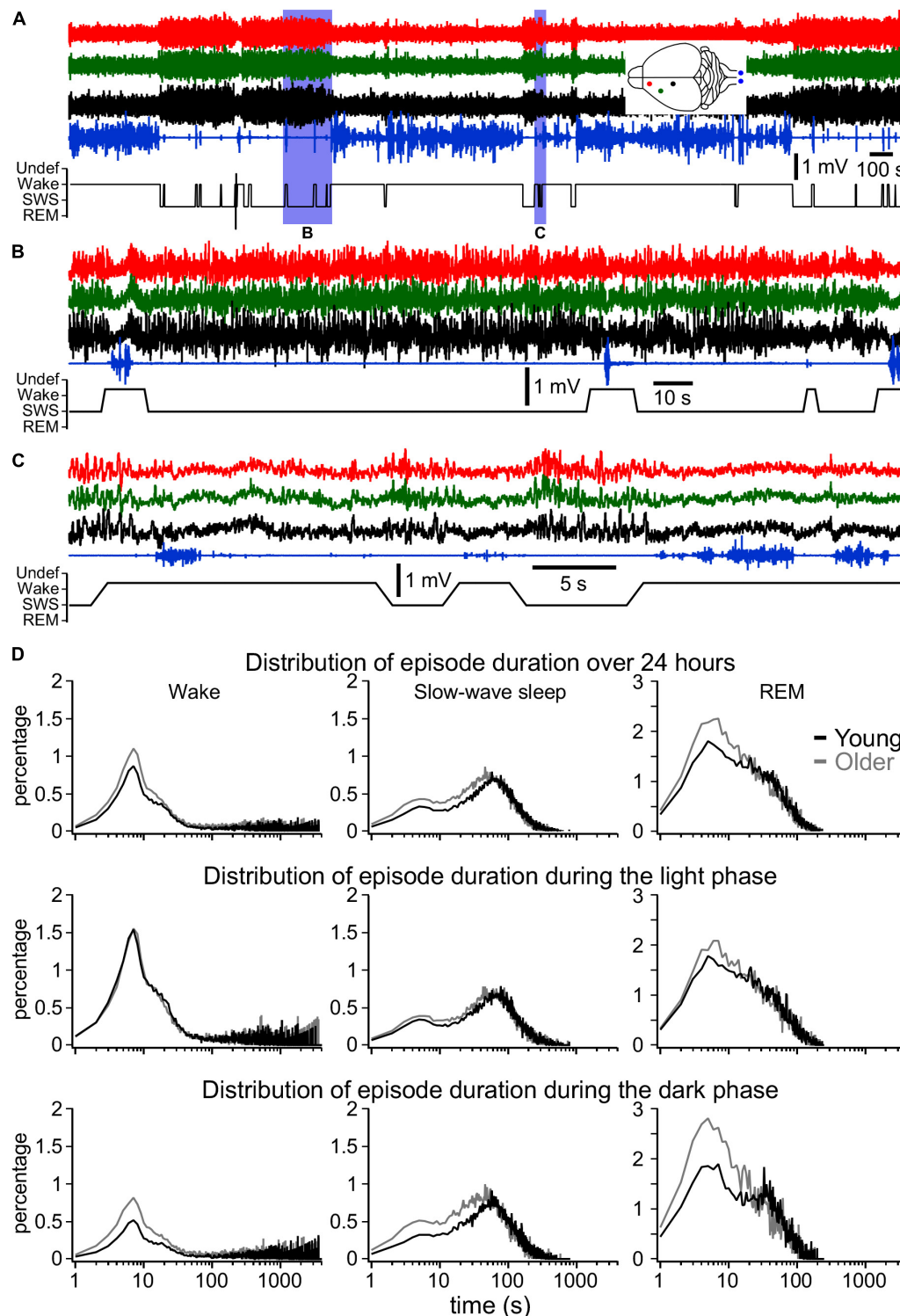
lower in anterior somatosensory, and even lower in posterior somatosensory cortices, while in young mice the delta power was the lowest in the frontal cortex and it was higher in both anterior and posterior somatosensory cortex which is the opposite of what is seen in human (Carrier et al., 2011).

## Slow-Wave Features in Young and Older Mice

The differences in SWS delta power in different ages and areas are likely mediated either by the number of waves or by their properties. Therefore, using a neural network approach (Bukhtiyarova et al., 2019), we extracted (**Figures 5A,B**) and quantified individual slow waves for 18 days from seven mice in each group and in the three different cortical areas investigated (**Figure 5**). Using LMM, a significant interaction between age and electrode for slow wave density was found ( $p = 0.05$ ), which suggest that age effect vary depending on the topography. Follow up *post hoc* tests show that slow wave density seems higher in old mice compared to young mice specifically in the frontal channel ( $p = 0.067$ , **Figure 5C**), while no age-difference approached significance for other channels (**Figure 5C**). The differences were not significant for slow wave duration or amplitude. These results suggest that a higher incidence of slow waves was likely a factor of the increase in the delta power observed in the frontal cortex of older mice.

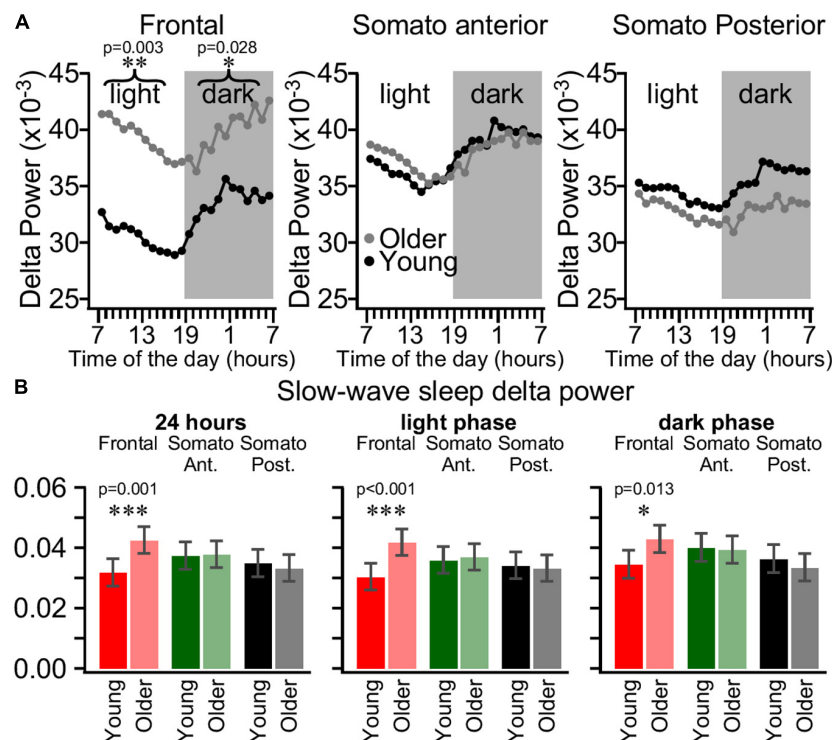
## Area Specificity for the Expression of Sleep–Wake States

It is very common that the detection of states of vigilance in rodents is done based on just one or two electrodes. Using EEG recordings in mice, it was recently shown that spontaneous sleep (Fernandez et al., 2017) or sleep deprivation differentially modulates slow wave vs. fast activities in various cortical areas (Kim et al., 2017). Combined with our data on area-specific differences in delta power, these results suggest that various areas can have different propensities for sleep and wake states. Therefore, we investigated LFP activities in 14 different cortical locations together with neck muscle EMG in intermediate age mice (6 months old, **Figure 6**). The overall idea of these experiments was to identify the given state of vigilance based on each LFP electrode separately and muscle activities, and then to compare the coincidence of state detection. Based on a formal criteria of state detection (**Figure 1**) we found that (a) the transition between states does not occur simultaneously and even in closely located areas, the delays of state transition could take up to 20 s (compare the first and second green traces, **Figure 6A**); (b) while REM sleep (high intensity theta



**FIGURE 3 |** Fragmentation of sleep and wake states. **(A)** One-hour long LFP recordings at the beginning of the dark phase (19:00–20:00 h) in a 1-year old mouse. **(B,C)** Expanded fragments from this recording demonstrating the presence of very short wake episodes during SWS [activated LFP and increased muscle tone, **(B)**] and sleep episodes during wake [slow-wave activity and reduced muscle tone, **(C)**]. **(D)** Share of all episodes from all investigated animals and days of wake, SWS, and REM sleep of different duration with 1 s resolution for 24 h, for light phase and dark phase as indicated for young and older mice. These distributions show that older mice have larger number of short-lasting states, in particular during the dark phase.





**FIGURE 4 |** Area-specific daily SWS delta power dynamics is different in young and older mice. **(A)** Estimated mean delta power in SWS at each hour of the day for the frontal electrode (left panel), the somatosensory anterior electrode (middle panel), and somatosensory posterior electrode (right panel) for young (black circles,  $n = 99$  days from eight mice) and older mice (gray circles,  $n = 67$  days from nine mice). Printed values are estimated marginal means computed in two independent mixed linear models (one per light or dark phase) using factors of age, electrode, and time of the day. **(B)** Mean SWS delta power in frontal cortex (red), somatosensory anterior cortex (green), and somatosensory posterior cortex (black) for 24 h (left panel), light phase (middle panel), and dark phase (right panel) for young (bright colors) and older (light colors) mice. Error bars indicate the 95% confidence interval. Significant statistical differences are indicated by asterisk,  $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ .

rhythm and low muscle tone) can be detected in a large number of cortical areas, some channels display slow-wave activities at the same time (**Figure 6B**). Using formal criteria of state identification, we often observed simultaneously the presence of two and sometimes three different states in different parts of neocortex (**Supplementary Movie S1**). Comparison of states (two mice, 3 days per mice) shows that during the light phase the frontal cortex spends only about 50% of time in the SWS, and in somatosensory cortex, the SWS occupies about 65% of time. The differences are even more dramatic for REM sleep, which takes 12–13% of light phase in visual and retrosplenial areas and only about 3% in lateral somatosensory area (**Figure 6C**). Therefore, electrographic activities corresponding to SWS, REM sleep, or wake can co-occur in different cortical areas.

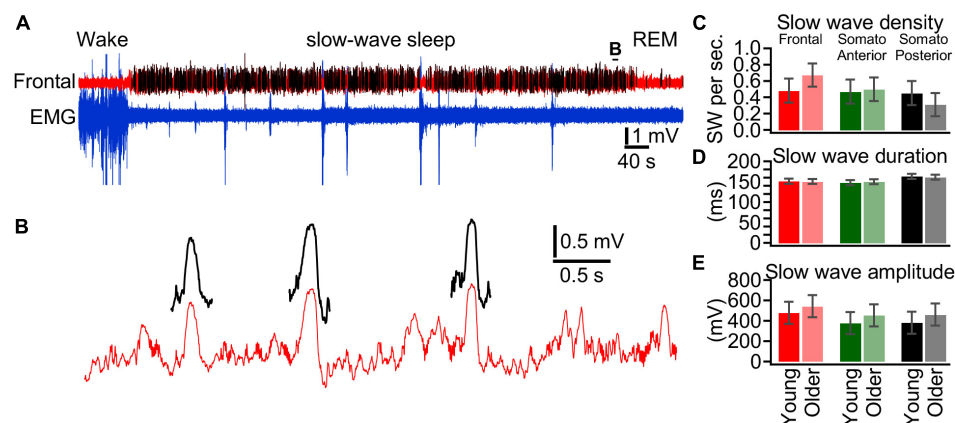
## DISCUSSION

In this study we show that age effects on sleep depend on specific period of the light/dark phases. In the light phase, 3-months old and 1-year-old mice spent similar time in sleep and wake states except that 1-year-old mice spent significantly less time in REM sleep at the beginning of each period. Similarly, 1-year-old mice spent less time in REM sleep at the end of the dark phase of

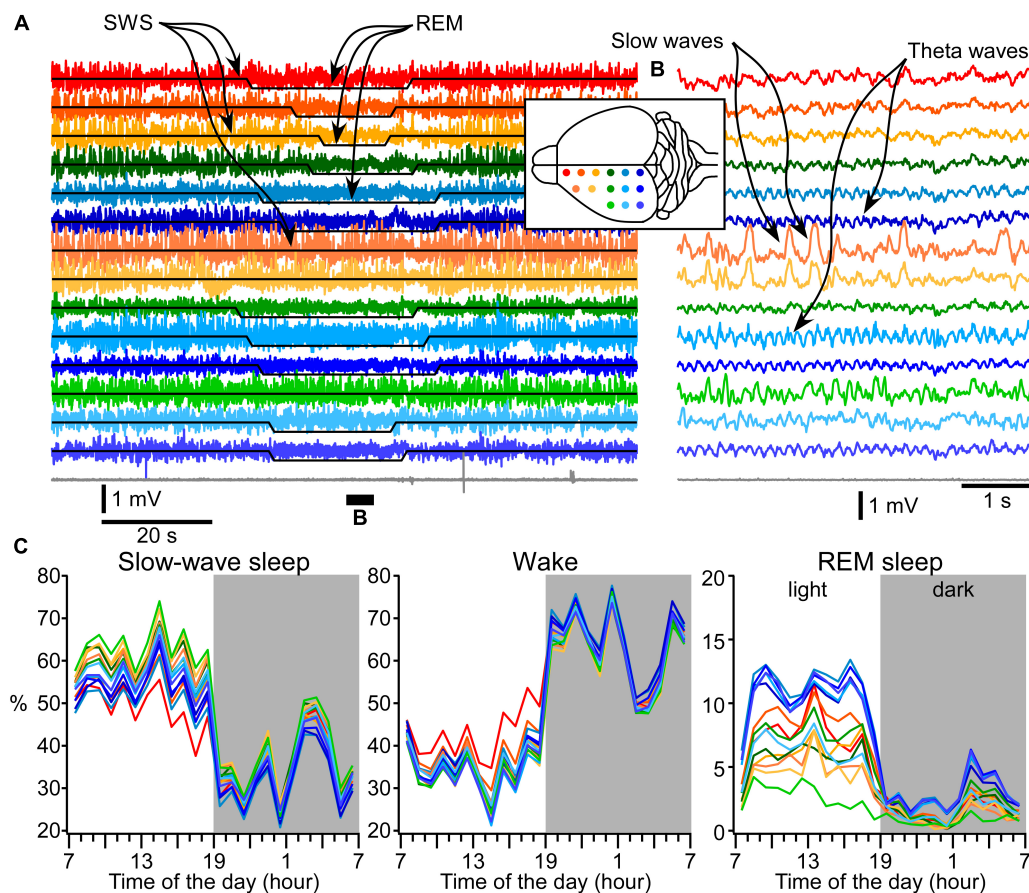
the cycle. Generally, in the dark phase, older mice spent globally less time in wake and more time in SWS. The increase in the total sleep duration was mainly due to an increase in the number of short (less than a minute) sleep episodes during the dark phase. The LFP delta power had remarkable regional specificity. Older mice showed larger LFP delta power in the frontal cortex compared to young mice, which can be explained by a tendency for an increase in slow wave density. We also investigated regional specificity of sleep–wake electrographic activities and found that using formal criteria, posterior regions of the cortex show more REM sleep activities (high theta power during low muscle tone), while somatosensory cortex displays more often SWS patterns.

## How Long Do States Last?

In agreement with multiple previous investigations, our study shows that overall, mice spent more time sleeping during the light part of the light–dark cycle and more time in waking state during the dark phase of the cycle (Tobler et al., 1997; Franken et al., 1999; El Helou et al., 2013). It is commonly used to call short states of activated EEG/LFP accompanied with an increase in the muscle tone in rodents as microarousals (Watson et al., 2016). However, the formal definition of microstates is unclear. Typically, in human, the state is arbitrarily defined as a stable pattern lasting  $> 15$  s (Iber et al., 2007). However, recent



**FIGURE 5 |** Slow wave features in young and older mice. **(A)** Segment of LFP recording from the frontal cortex (red trace) and EMG from neck muscle (blue trace). Slow waves are detected using a neural network approach and detected slow waves are depicted by black traces. **(B)** Expanded segment from **(A)** as indicated. **(C)** Slow wave density detected during slow-wave sleep over a 24 h period for young (bright colors,  $n = 18$  days from nine mice) and older (light colors,  $n = 18$  days from eight mice). **(D)** Duration of detected slow waves, **(E)** Amplitude of detected slow waves.



**FIGURE 6 |** Area-specific distribution of electrographic brain states. **(A)** A segment of 14 LFP channels and muscle (gray) recordings during SWS, REM sleep, and again SWS. Signals from electrodes are color coded and the location of electrodes is indicated in the inserted drawing. **(B)** A short segment that was overall qualified as REM sleep, but two fronto-laterally located electrodes show clear slow-wave activity. **(C)** 24-h distribution of states detected based on muscle activity and each electrode individually.

advances in automatic methods of state detection demonstrated that the conventional states can be much more fragmented than it was previously assumed (Koch et al., 2019). It is known that brain states in animals undergo rapid changes, which are associated with marked changes in neuronal activities in different structures (Bezdudnaya et al., 2006; Reimer et al., 2016). Human studies indicate that 500 ms is the minimal time to mediate conscious perception (Libet et al., 1967; Edelman, 2003). An essential condition for a conscious state to be generated is spatio-temporally coordinated neuronal firing (Konorski, 1967). The presence of EEG slow wave activities is a key factor for the loss of consciousness (Purdon et al., 2013; Mukamel et al., 2014) because it interrupts firing. Sleep or anesthesia-induced slow waves in cats and those measured intracellularly are mediated by neuronal hyperpolarization and silence that lasts 150–300 ms (Timofeev et al., 2001; Volgushev et al., 2006; Chauvette et al., 2010, 2011). In sleeping rats and mice, the LFP slow waves or reduced/abolished firing associated with slow waves are variable in duration, and they typically last >100 ms (Vyazovskiy et al., 2009; Bukhtiyarova et al., 2019). Therefore, the active states during sleep and anesthesia last typically longer than 500 ms and the animals are still unconscious. Thus, the minimal duration of an arousal should be longer than 500 ms and likely longer than 1 s in order to be considered as a state. On the contrary, the presence of local slow waves in some spatially restricted parts of the cortex alters behavioral responses (Vyazovskiy et al., 2011) suggesting that locally synchronized silent periods, lasting a few hundreds of milliseconds, could be considered as a state. Our study demonstrates the presence of a large number of events that follow a lognormal distribution and were identified as wake or SWS that last <10 s (**Figure 3D**). This suggests that they represent isolated states, which likely can be called microarousal or microsleep states. Similar activities exceeding this time likely represent either short or long states, not microstates. As our detection method occasionally leads to undefined states at state transitions (see the section “Materials and Methods”), we cannot exclude that a significant part of detected short states are due to the detection method itself. Because the delta power varied not only between cycles, but also between the beginning and the end of the cycle, and we used the same threshold throughout each phase (light or dark), we could artificially increase the detection of short states. However, there are no reasons to believe that the method would affect more the results of one group in particular; therefore, our observation of a higher number of short-lasting states (wake or SWS) in older mice compared to young mice is very solid. The state duration in mice living in natural environment could be different. For example, it is unlikely that mice foraging in the forest at the beginning of the dark phase of the sleep–wake cycle will have brief sleep episodes.

## Delta Power Differences

The anterior somatosensory electrode in our study was located in the barrel field and we did not observe age-specific differences in the sleep delta power in young vs. older mice (**Figure 4**). This suggests that there are very little age-dependent changes in the most important sensory system of mice. While no differences were found in the posterior somatosensory cortex (trunk area),

there was an increase in delta power and a trend for an increased slow wave density in frontal cortex of older mice. Increase in delta power in frontal cortex in older mice are opposed to what is seen in humans (Carrier et al., 2011). Mechanistic aspects of such differences are unclear and will require further studies. It was demonstrated in human that a reduction of cortical thickness was responsible for an age-dependent reduction in slow wave activities (Dube et al., 2015). There is no difference in frontal cortical thickness between 1- and 3.5-month-old mice, but somatosensory cortex thickness decreases (Grand'Maison et al., 2013), and at later ages (12, 18 months) the decrease was minor (Hébert et al., 2013). This might explain the relative stability of delta power in the barrel cortex in our experiments.

## Wake–Sleep Cycle and Aging in Mice

Mice seemed to be a very attractive model to study aging. By the age of 3 months, they become mature adults, and at the age 10–15 months they belong to middle-aged group and after 18 months they are considered as being old (Flurkey et al., 2007). The aging in mice occurs substantially faster than in other laboratory animals and if the model is good, it would give multiple practical advances to study aging. Here we investigated several parameters of sleep–wake cycle in C57bL/6 mice in aiming to evaluate appropriateness of this mice strain to study in future the mechanistic aspects of human sleep.

During aging there are major alterations in human sleep. We took several appropriate parameters of human sleep affected by aging (Mander et al., 2017) and compared them to our results (1) with aging, humans have overall shorter sleep duration but some reports indicated that they sleep more during the day. *In our study*, the overall duration of SWS was higher in older mice, due longer SWS time only during the dark phase of the cycle (**Figure 3D**). Thus, while overall time spent in sleep is *different* between species, both show a *similar* increase of sleep during their active phase (light for humans, dark for mice). (2) *Similar* to human that reveal increased sleep fragmentation with aging, sleep fragmentation was much higher in older mice compared to young that is evidenced by a larger number of short states. (3) Older human have increased time spent awake throughout the night; however, older mice showed just a trend for increased time spent in a wake state throughout the light period, that is a period at which sleep dominates. (4) Older humans display an increased frequency of diurnal naps. In older mice, in the dark phase, there was not only an increase in the number of microarousals/micro sleep states, but also in the number of SWS episodes lasting between 10 and 60 s, which might be considered as “naps,” and there was an increased total duration of sleep, and is therefore similar to human. (5) Humans show an overall reduced slow-wave activity with aging. *Different* from human, in our mice experiments, the overall slow-wave activity did not undergo systematic changes. We found an increase with aging in the delta power and a trend for an increased slow wave density in frontal cortex, but stable delta power in the posterior part of somatosensory cortex. (6) Our data demonstrate that older mice showed a decrease in REM sleep, in particular at the end of dark phase and the beginning of the light phase of the light/dark cycle (**Figure 2D**). Although effects of aging in humans are more

prominent in NREM sleep, reductions in REM sleep are also reported in healthy older participants (Carrier et al., 1997).

One aspect of our study investigated the spatial distribution of formally defined states over the cortical surface. We found that electrodes implanted in frontal cortex would identify more of wake, in motor and somatosensory cortex more SWS and in visual and retrosplenial cortex more REM sleep (Figure 6). REM sleep-related theta activity was more often seen in posterior parts of the dorsal cortical surface and slow-wave activity during otherwise REM sleep was often recorded over somatosensory and motor cortical areas (Figure 6). This supports a recent discovery of REM sleep-dependent slow wave activity in mice (Funk et al., 2016) except that we did not see REM sleep slow wave activities in visual cortex, and our recordings identified clear slow waves in layer 5, but not restricted to layer 4 as in the Funk et al.'s (2016) study. The observed dominance of wake, SWS, and REM sleep regions in our study largely overlaps with the anatomical localization of cortical clusters belonging to the default mode network identified in awake mice (Vanni et al., 2017) suggesting that on a daily scale, sleep and wake activities interact over large cortical areas. Two types of slow-wave activities during REM sleep were also found in human (Bernardi et al., 2019). Frontal-central slow waves, which authors believe are analogous with PGO waves, had increased gamma activity during slow waves (Bernardi et al., 2019, their Figure 4E). It is well-demonstrated that sleep slow waves in both human and animals are associated with a reduction of gamma activities (Mukovski et al., 2007; Cash et al., 2009; Csécsa et al., 2010; Bukhtiyarova et al., 2019). Some field potential deflections during REMs are expected because of synchronous increase in inhibitory activities in the cortex (Timofeev et al., 2001), which typically mediate gamma oscillations (Whittington et al., 2000; Buzsáki and Wang, 2012). Medial occipital slow waves although show a reduction in gamma power, but in the shown example (Bernardi et al., 2019, their Figure 7A), consist of an oscillation resembling spindle, not slow wave. It should be noted that a reduction in gamma frequency activities occurs more often than slow waves and often it cannot be used as a solo mean of slow wave detection (Bukhtiyarova et al., 2019). Therefore, the low frequency activity recorded during REM sleep by Bernardi et al. (2019) does not share the known feature of sleep slow waves.

## CONCLUSION

In this study we investigated the sleep–wake cycle and slow-wave activities of young and older mice during sleep. The ultimate goal

was to evaluate the appropriateness of mice as a model to study human sleep during aging. We suggest that overall the aging mice sleep shares several common features human sleep. The major difference is an increase in older mice of the delta power in frontal cortex, while it decreases in human.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

All experiments were performed in accordance with the guideline of the Canadian Council on Animal Care and approved by the Université Laval Committee on Ethics and Animal Research.

## AUTHOR CONTRIBUTIONS

SS, SC, and JS performed the experiments. SS, SC, and IT wrote the manuscript. SC, OB, J-ML, and JD developed the analytical tools. SS, SC, JS, OB, JD, and IT analyzed the data. JC contributed to the experimental design. IT supervised the experiments and analysis, and planned the research. All authors corrected and approved the final version of the manuscript.

## ACKNOWLEDGMENTS

We thank Sergiu Ftomov for excellent technical assistance. This study was supported by the Canadian Institutes of Health Research (CIHR, MOP-136969 and MOP-136967), the Fonds de Recherche du Québec – Nature et Technologies (FRQNT), the National Institutes of Health (NIH, NS104368) and the Savoy Foundation.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnsys.2019.00051/full#supplementary-material>

**MOVIE S1 |** Dynamic of sleep–wake states detected in individual cortical electrodes based on delta power and theta power in each cortical electrode and neck muscle tone. The movie is played at a faster speed and the time is indicated in the top left corner.

## REFERENCES

- Bernardi, G., Betta, M., Ricciardi, E., Pietrini, P., Tononi, G., and Siclari, F. (2019). Regional delta waves in human rapid eye movement sleep. *J. Neurosci.* 39, 2686–2797. doi: 10.1523/JNEUROSCI.2298-18.2019
- Bezdudnaya, T., Cano, M., Bereshpolova, Y., Stoelzel, C. R., Alonso, J. M., and Swadlow, H. A. (2006). Thalamic burst mode and inattention in the awake LGNd. *Neuron* 49, 421–432. doi: 10.1016/j.neuron.2006.01.010
- Borbély, A. A. (1982). A two process model of sleep regulation. *Hum. Neurobiol.* 1, 195–204.
- Borbély, A. A., Achermann, P., Trachsel, L., and Tobler, I. (1989). Sleep initiation and initial sleep intensity: interactions of homeostatic and circadian mechanisms. *J. Biol. Rhythms* 4, 149–160.
- Borbély, A. A., Tobler, I., and Hanagasioglu, M. (1984). Effect of sleep deprivation on sleep and EEG power spectra in the rat. *Behav. Brain Res.* 14, 171–182. doi: 10.1016/0166-4328(84)90186-4



- Bukhtiarova, O., Soltani, S., Chauvette, S., and Timofeev, I. (2016). Supervised semi-automatic detection of slow waves in non-anaesthetized mice with the use of neural network approach. *Transl. Brain Rhythm.* 1, 14–18.
- Bukhtiarova, O., Soltani, S., Chauvette, S., and Timofeev, I. (2019). Slow wave detection in sleeping mice: comparison of traditional and machine learning methods. *J. Neurosci. Methods* 316, 35–45. doi: 10.1016/j.jneumeth.2018.08.016
- Buzsaki, G., and Wang, X. J. (2012). Mechanisms of gamma oscillations. *Annu. Rev. Neurosci.* 35, 203–225. doi: 10.1146/annurev-neuro-062111-150444
- Carrier, J., Monk, T. H., Buysse, D. J., and Kupfer, D. J. (1997). Sleep and morningness-eveningness in the 'middle' years of life (20–59 y). *J. Sleep Res.* 6, 230–237. doi: 10.1111/j.1365-2869.1997.00230.x
- Carrier, J., Viens, I., Poirier, G., Robillard, R., Lafortune, M., Vandewalle, G., et al. (2011). Sleep slow wave changes during the middle years of life. *Eur. J. Neurosci.* 33, 758–766. doi: 10.1111/j.1460-9568.2010.07543.x
- Cash, S. S., Halgren, E., Dehghani, N., Rossetti, A. O., Thesen, T., Wang, C., et al. (2009). The human K-complex represents an isolated cortical down-state. *Science* 324, 1084–1087. doi: 10.1126/science.1169626
- Chauvette, S., Crochet, S., Volgushev, M., and Timofeev, I. (2011). Properties of slow oscillation during slow-wave sleep and anesthesia in cats. *J. Neurosci.* 31, 14998–15008. doi: 10.1523/JNEUROSCI.2339-11.2011
- Chauvette, S., Volgushev, M., and Timofeev, I. (2010). Origin of active states in local neocortical networks during slow sleep oscillation. *Cereb. Cortex* 20, 2660–2674. doi: 10.1093/cercor/bhq009
- Contreras, D., and Steriade, M. (1995). Cellular basis of EEG slow rhythms: a study of dynamic corticothalamic relationships. *J. Neurosci.* 15, 604–622. doi: 10.1523/jneurosci.15-01-00604.1995
- Csercsa, R., Dombovári, B., Fabó, D., Wittner, L., Eross, L., Entz, L., et al. (2010). Laminar analysis of slow wave activity in humans. *Brain* 133, 2814–2829. doi: 10.1093/brain/awq169
- Dube, J., Lafortune, M., Bedetti, C., Bouchard, M., Gagnon, J. F., Doyon, J., et al. (2015). Cortical thinning explains changes in sleep slow waves during adulthood. *J. Neurosci.* 35, 7795–7807. doi: 10.1523/JNEUROSCI.3956-14.2015
- Durán, E., Oyanedel, C. N., Niethard, N., Inostroza, M., and Born, J. (2018). Sleep stage dynamics in neocortex and hippocampus. *Sleep* 41:zsy060. doi: 10.1093/sleep/zsy060
- Edelman, G. M. (2003). Naturalizing consciousness: a theoretical framework. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5520–5524. doi: 10.1073/pnas.0931349100
- El Helou, J., Bélanger-Nelson, E., Freyburger, M., Dorsaz, S., Curie, T., La Spada, F., et al. (2013). Neuroligin-1 links neuronal activity to sleep-wake regulation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9974–9979. doi: 10.1073/pnas.1221381110
- Fernandez, L. M. J., Comte, J.-C., Le Merre, P., Lin, J.-S., Salin, P.-A., and Crochet, S. (2017). Highly dynamic spatiotemporal organization of low-frequency activities during behavioral states in the mouse cerebral cortex. *Cereb. Cortex* 27, 5444–5462. doi: 10.1093/cercor/bhw311
- Flurkey, K., Currer, J., and Harrison, D. (2007). "The mouse in aging research," in *The Mouse in Biomedical Research*, 2nd Edn, ed. J. G. Fox (Burlington, MA: American College Laboratory Animal Medicine), 637–672. doi: 10.1016/b978-012369454-6/50074-1
- Franken, P., Malafosse, A., and Tafti, M. (1998). Genetic variation in EEG activity during sleep in inbred mice. *Am. J. Physiol.* 275, R1127–R1137. doi: 10.1152/ajpregu.1998.275.4.R1127
- Franken, P., Malafosse, A., and Tafti, M. (1999). Genetic determinants of sleep regulation in inbred mice. *Sleep* 22, 155–169.
- Funk, C. M., Honjoh, S., Rodriguez, A. V., Cirelli, C., and Tononi, G. (2016). Local slow waves in superficial layers of primary cortical areas during REM sleep. *Curr. Biol.* 26, 396–403. doi: 10.1016/j.cub.2015.11.062
- Gilbert, C., and Wiesel, T. (1983). Clustered intrinsic connections in cat visual cortex. *J. Neurosci.* 3, 1116–1133. doi: 10.1523/jneurosci.03-05-01116.1983
- Grand'Maison, M., Zehntner, S. P., Ho, M.-K., Hébert, F., Wood, A., Carbonell, F., et al. (2013). Early cortical thickness changes predict  $\beta$ -amyloid deposition in a mouse model of Alzheimer's disease. *Neurobiol. Dis.* 54, 59–67. doi: 10.1016/j.nbd.2013.02.005
- Hasan, S., Dauvilliers, Y., Mongrain, V., Franken, P., and Tafti, M. (2012). Age-related changes in sleep in inbred mice are genotype dependent. *Neurobiol. Aging* 33, e13–e195. doi: 10.1016/j.neurobiolaging.2010.05.010
- Hébert, F., Grand'Maison, M., Ho, M.-K., Lerch, J. P., Hamel, E., and Bedell, B. J. (2013). Cortical atrophy and hypoperfusion in a transgenic mouse model of Alzheimer's disease. *Neurobiol. Aging* 34, 1644–1652. doi: 10.1016/j.neurobiolaging.2012.11.022
- Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. (2007). *Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Westchester, NY: American Academy of Sleep Medicine.
- Kim, B., Kocsis, B., Hwang, E., Kim, Y., Strecker, R. E., McCarley, R. W., et al. (2017). Differential modulation of global and local neural oscillations in REM sleep by homeostatic sleep regulation. *Proc. Natl. Acad. Sci. U.S.A.* 114, E1727–E1736. doi: 10.1073/pnas.1615230114
- Koch, H., Jennum, P., and Christensen, J. A. E. (2019). Automatic sleep classification using adaptive segmentation reveals an increased number of rapid eye movement sleep transitions. *J. Sleep Res.* 28:e12780. doi: 10.1111/jsr.12780
- Koerner, K. T., and Zhang, Y. (2017). Application of linear mixed-effects models in human neuroscience research: a comparison with pearson correlation in two auditory electrophysiology studies. *Brain Sci.* 7, 1–11. doi: 10.3390/brainsci7030026
- Konorski, J. (1967). *Integrative Activity of the Brain: An Interdisciplinary Approach*. Chicago, IL: The University of Chicago Press.
- Kurth, S., Ringli, M., Geiger, A., LeBourgeois, M., Jenni, O. G., and Huber, R. (2010). Mapping of cortical activity in the first two decades of life: a high-density sleep electroencephalogram study. *J. Neurosci.* 30, 13211–13219. doi: 10.1523/JNEUROSCI.2532-10.2010
- Libet, B., Alberts, W. W., Wright, E. W. Jr., and Feinstein, B. (1967). Responses of human somatosensory cortex to stimuli below threshold for conscious sensation. *Science* 158, 1597–1600. doi: 10.1126/science.158.3808.1597
- Lund, J. P., Sun, G.-D., and Lamarque, Y. (1994). Cortical reorganization and deafferentation in adult macaques. *Science* 265, 546–548. doi: 10.1126/science.8036500
- Lund, J. S., Yoshioka, T., and Levitt, J. B. (1993). Comparison of intrinsic connectivity in different areas of macaque monkey cerebral cortex. *Cereb. Cortex* 3, 148–162. doi: 10.1093/cercor/3.2.148
- Mander, B. A., Winer, J. R., and Walker, M. P. (2017). Sleep and human aging. *Neuron* 94, 19–36. doi: 10.1016/j.neuron.2017.02.004
- McKillop, L. E., Fisher, S. P., Cui, N., Peirson, S. N., Foster, R. G., Wafford, K. A., et al. (2018). Effects of aging on cortical neural dynamics and local sleep homeostasis in mice. *J. Neurosci.* 38, 3911–3928. doi: 10.1523/JNEUROSCI.2513-17.2018
- Mukamel, E. A., Pironi, E., Babadi, B., Wong, K. F. K., Pierce, E. T., Harrell, P. G., et al. (2014). A transition in brain state during propofol-induced unconsciousness. *J. Neurosci.* 34, 839–845. doi: 10.1523/JNEUROSCI.5813-12.2014
- Mukovski, M., Chauvette, S., Timofeev, I., and Volgushev, M. (2007). Detection of active and silent states in neocortical neurons from the field potential signal during slow-wave sleep. *Cereb. Cortex* 17, 400–414. doi: 10.1093/cercor/bhj157
- Nir, Y., Staba, R. J., Andrillon, T., Vyazovskiy, V. V., Cirelli, C., Fried, I., et al. (2011). Regional slow waves and spindles in human sleep. *Neuron* 70, 153–169. doi: 10.1016/j.neuron.2011.02.043
- Panagiotou, M., Vyazovskiy, V. V., Meijer, J. H., and Deboer, T. (2017). Differences in electroencephalographic non-rapid-eye movement sleep slow-wave characteristics between young and old mice. *Sci. Rep.* 7:43656. doi: 10.1038/srep43656
- Purdon, P. L., Pierce, E. T., Mukamel, E. A., Prerau, M. J., Walsh, J. L., Wong, K. F. K., et al. (2013). Electroencephalogram signatures of loss and recovery of consciousness from propofol. *Proc. Natl. Acad. Sci. U.S.A.* 110, E1142–E1151. doi: 10.1073/pnas.1221180110
- Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., et al. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat. Commun.* 7:13289. doi: 10.1038/ncomms13289
- Rockland, K. S. (1985). Anatomical organization of primary visual cortex. (area 17) in the ferret. *J. Comp. Neurol.* 241, 225–236. doi: 10.1002/cne.902410209
- Schuz, A., Chaimow, D., Liewald, D., and Dörtenman, M. (2006). Quantitative aspects of corticocortical connections: a tracer study in the mouse. *Cereb. Cortex* 16, 1474–1486. doi: 10.1093/cercor/bhj085
- Steriade, M., Timofeev, I., and Grenier, F. (2001). Natural waking and sleep states: a view from inside neocortical neurons. *J. Neurophysiol.* 85, 1969–1985. doi: 10.1152/jn.2001.85.5.1969

- Timofeev, I., Grenier, F., and Steriade, M. (2001). Disfacilitation and active inhibition in the neocortex during the natural sleep-wake cycle: an intracellular study. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1924–1929. doi: 10.1073/pnas.041430398
- Tobler, I., Deboer, T., and Fischer, M. (1997). Sleep and sleep regulation in normal and prion protein-deficient mice. *J. Neurosci.* 17, 1869–1879. doi: 10.1523/jneurosci.17-05-01869.1997
- Van Hooser, S. D., Heimel, J. A., Chung, S., and Nelson, S. B. (2006). Lack of patchy horizontal connectivity in primary visual cortex of a mammal without orientation maps. *J. Neurosci.* 26, 7680–7692. doi: 10.1523/jneurosci.0108-06.2006
- Vanni, M. P., Chan, A. W., Balbi, M., Silasi, G., and Murphy, T. H. (2017). Mesoscale mapping of mouse cortex reveals frequency-dependent cycling between distinct macroscale functional modules. *J. Neurosci.* 37, 7513–7533. doi: 10.1523/JNEUROSCI.3560-16.2017
- Volgushev, M., Chauvette, S., Mukovski, M., and Timofeev, I. (2006). Precise long-range synchronization of activity and silence in neocortical neurons during slow-wave sleep. *J. Neurosci.* 26, 5665–5672. doi: 10.1523/jneurosci.0279-06.2006
- Vyazovskiy, V. V., Olcese, U., Hanlon, E. C., Nir, Y., Cirelli, C., and Tononi, G. (2011). Local sleep in awake rats. *Nature* 472, 443–447. doi: 10.1038/nature10009
- Vyazovskiy, V. V., Olcese, U., Lazimy, Y. M., Faraguna, U., Esser, S. K., Williams, J. C., et al. (2009). Cortical firing and sleep homeostasis. *Neuron* 63, 865–878. doi: 10.1016/j.neuron.2009.08.024
- Watson, B. O., Levenstein, D., Greene, J. P., Gelinas, J. N., and Buzsáki, G. (2016). Network homeostasis and state dynamics of neocortical sleep. *Neuron* 90, 839–852. doi: 10.1016/j.neuron.2016.03.036
- Whittington, M. A., Traub, R. D., Kopell, N., Ermentrout, B., and Buhl, E. H. (2000). Inhibition-based rhythms: experimental and mathematical observations on network dynamics. *Int. J. Psychophysiol.* 38, 315–336. doi: 10.1016/S0167-8760(00)00173-2
- Wimmer, M. E., Rising, J., Galante, R. J., Wyner, A., Pack, A. I., and Abel, T. (2013). Aging in mice reduces the ability to sustain sleep/wake states. *PLoS One* 8:e81880. doi: 10.1371/journal.pone.0081880

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Soltani, Chauvette, Bukhtiyarova, Lina, Dubé, Seigneur, Carrier and Timofeev. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Naturalistic Approach to the Hard Problem of Consciousness

**Wolf Singer<sup>1,2,3\*</sup>**

<sup>1</sup>Max Planck Institute for Brain Research (MPI), Frankfurt am Main, Germany, <sup>2</sup>Ernst Strüngmann Institute for Neuroscience in Cooperation with the Max Planck Society (ESI), Frankfurt am Main, Germany, <sup>3</sup>Frankfurt Institute for Advanced Studies (FIAS), Frankfurt am Main, Germany

Following a brief review of current efforts to identify the neuronal correlates of conscious processing (NCCP) an attempt is made to bridge the gap between the material neuronal processes and the immaterial dimensions of subjective experience. It is argued that this “hard problem” of consciousness research cannot be solved by only considering the neuronal underpinnings of cognition. The proposal is that the hard problem can be treated within a naturalistic framework if one considers not only the biological but also the socio-cultural dimensions of evolution. The argument is based on the following premises: perceptions are the result of a constructivist process that depends on priors. This applies both for perceptions of the outer world and the perception of oneself. Social interactions between agents endowed with the cognitive abilities of humans generated immaterial realities, addressed as social or cultural realities. This novel class of realities assumed the role of priors for the perception of oneself and the embedding world. A natural consequence of these extended perceptions is a dualist classification of observables into material and immaterial phenomena nurturing the concept of ontological substance dualism. It is argued that perceptions shaped by socio-cultural priors lead to the construction of a self-model that has both a material and an immaterial dimension. As priors are implicit and not amenable to conscious recollection the perceived immaterial dimension is experienced as veridical and not derivable from material processes—which is the hallmark of the hard problem. These considerations let the hard problem appear as the result of cognitive constructs that are amenable to naturalistic explanations in an evolutionary framework.

**Keywords:** consciousness, dualism, social realities, qualia, emergence, self-model

## OPEN ACCESS

### Edited by:

Maurizio Mattia,  
National Institute of Health (ISS), Italy

### Reviewed by:

Christopher I. Petkov,  
Newcastle University,  
United Kingdom  
José M. Delgado-García,  
Universidad Pablo de Olavide, Spain

### \*Correspondence:

Wolf Singer  
wolf.singer@brain.mpg.de

**Received:** 04 June 2019

**Accepted:** 07 October 2019

**Published:** 25 October 2019

### Citation:

Singer W (2019) A Naturalistic  
Approach to the Hard Problem of  
Consciousness.  
*Front. Syst. Neurosci.* 13:58.  
doi: 10.3389/fnsys.2019.00058

## INTRODUCTION

Attempts to provide naturalistic explanations for the phenomenon of consciousness are confronted with at least three major difficulties. The first arises from the fact that the explanandum is not well defined. The second results from the still rudimentary understanding of neuronal processes underlying higher cognitive functions. And the third is related to the “hard problem” of consciousness research (for review, see Dennett, 2018), the intuition that even if we had

a comprehensive account of the neuronal correlates of consciousness (NCCP) we would still be unable to explain how the first-person experiences of the results of conscious processing, the qualia, emerge from the neuronal interactions described from a third-person perspective.

In this essay I shall briefly address the first two problems by reviewing recent developments in the search for the NCCP and then propose a strategy to soften the hard problem. The proposal is that the immaterial nature of the qualia can perhaps be accounted for within a naturalistic framework if one considers that not only the perception of the world around us but also the perception of ourselves is the result of a constructivist process that depends on priors. The core assumption is that the priors for our self-model are provided by the socio-cultural environments, the immaterial social realities, that humans have created once cultural and biological dimensions became interactive in evolution. Some of the experimental results pertinent to the NCCP have been reviewed by this author in previous articles and it is, therefore, likely that fragments of formulations are repeated here.

## AN ILL-DEFINED EXPLANANDUM

The terms consciousness, conscious and consciously are associated with many different connotations and therefore discussions on neuronal correlates of consciousness are carried out on widely differing levels. In the most straight forward sense, the adjective conscious is used to simply designate brain states enabling subjects to be aware of their actions and their environment. In this case, consciousness is contrasted with sleep or coma. Defined in this way consciousness is a phenomenon that humans share with many different species of widely differing complexity. In another context, consciousness refers to a processing mode that is associated with verbal reportability of perceived stimuli or with storage of perceived items in working and/or declarative memory. The contrasting processing mode is subconscious or non-conscious processing in which stimuli are readily analyzed by the brain and can even control responses without the subject being aware of having been engaged in a cognitive operation. As far as can be assessed from a third-person perspective, all mammals and probably all vertebrates seem to be able to exploit these different processing modes. Although verbal reportability cannot be used as criterion, the brains of these animals possess all the mechanisms qualifying for conscious processing such as the control of attention, the storage of attended contents in working and episodic memory, the evaluation of context, the intentional selection of appropriate behavioral reactions and the ability to purposefully navigate in complex foraging grounds. In yet another reference frame consciousness is equated with a form of meta-awareness and denotes a condition, in which subjects are aware of their body scheme, of having emotions and memories or being in a particular state or having performed an action. This connotation of consciousness requires a form of self-awareness and is often assessed with the mirror test. A test that human babies pass beyond the age of two and certain animals such as crows, monkeys and

apes. Yet another connotation of consciousness is the ability to generate a theory of mind, the ability to imagine what is in the mind of the respective other. For long this has been considered as a specific human ability but there is now robust evidence that at least birds, dogs and monkeys have this capacity as well. Finally, there are higher forms of meta-awareness that are thought to exist only in humans and are associated with self-reflection and self-control, leading to the attribution of responsibility, morality and free will. Here the implicit assumption is that only contents that enter consciousness are amenable to rational deliberations and decisions. And last but not least humans are conscious of being conscious, which might be considered the highest form of meta-awareness. Experts in contemplative practices claim in addition, that such states of meta-consciousness can be devoid of content, the conscious state being the only object of the “inner eye.” In view of these very diverse meanings attributed to the term consciousness it is natural that the strategies applied in the search for the NCCP are heterogeneous and address only selected aspects of consciousness.

For the sake of brevity, I shall not review work aimed at the identification of mechanisms controlling brain states that permit or prevent the manifestation of conscious behavior. In this field of research, there is broad consensus that a critical level of network excitability is required to enable conscious processing and that these states are characterized by dynamics with specific electrographic signatures. The mechanisms controlling these states are closely related with ascending modulatory systems that regulate the sleep-waking cycle and global levels of excitability. Rather I shall concentrate on the discussion of neuronal processes that distinguish conscious from non-conscious processing in the awake brain.

## CONSCIOUS AND SUBCONSCIOUS PROCESSING

At any one moment, subjects are only aware of a small fraction of their cognitive and executive operations. Still, signals that subjects are not aware of can be processed in considerable depth and impact behavior (Dehaene et al., 1998). Thus, there must be gating mechanisms that determine which signals are processed consciously, which are processed and control behavior but remain unconscious and which are not processed at all.

In animal experiments, one of the methodological problems is to assess from a third-person perspective whether a content had been processed consciously or subconsciously. It is commonly held that conscious processing is distinguished from subconscious processing by the ability of subjects to be *aware* of the consciously processed content and to report this fact. These contents can be percepts, thoughts, decisions, intentions and actions or, in the case of contemplative practices, the awareness of pure presence. Thus, the experimenter has to rely on reports from the subjects' first-person perspective. In human subjects, this problem can be mitigated by requesting verbal reports or by instructing subjects to grade their operant responses according to the experienced degree of awareness. This allows one to overcome the ambiguity introduced by



the fact that in most forced-choice paradigms subjects give correct responses well above chance even if they have not been aware of having perceived the stimuli. Experiments on blind sight and investigations of split-brain patients impressively document this fact (see below). As in animals only operant responses can be obtained, it is arguable whether distinctions between conscious and unconscious processing can be made in the same way as in experiments with human subjects. The mere difficulty or complexity of an accomplished task is only a weak indicator for the involvement of conscious processes because subconscious computations can also rely on highly sophisticated heuristics, semantic interpretations and logical deductions.

In depth exploration of neuronal mechanisms requires assessment of neuronal responses with high temporal and spatial resolution, and this is also the case for the investigation of the NCCP. With the exception of utilizing data from patients implanted with intracranial electrodes for diagnostic reasons, such high resolution data can only be obtained in animal studies. Yet, these approaches are hampered by the ambiguities associated with operant responses. Thus, results from animal studies are usually interpreted on the basis of the assumption that the distinction between conscious and subconscious processing also holds at least for higher vertebrates and mammals. The reason is the remarkable cross species similarity of brain organization. However, as pointed out by the philosopher Nagel (1974) we cannot know whether animals are aware of stimuli and responses in the same way as human subjects.

For this very reason, most studies on the NCCP are performed with non-invasive measurements in human subjects. However, because of the limited spatial and/or temporal resolution of these techniques, mechanistic interpretations often have to rely on analogies with neuronal processes supporting cognitive functions in animals that are only indirectly related to consciousness.

One important mechanism gating the access of information to conscious processing, most likely shared with other mammals, is attention. As suggested by the phenomena of change blindness (Simons and Chabris, 1999), sensory neglect (Doricchi et al., 2008), attentional blink (Fu and Rutishauser, 2018), and certain masking paradigms, non-attended stimuli usually fail to be processed consciously and escape the subjects' awareness. Whether these limitations are due to the inability to attend to large numbers of items simultaneously or whether they result from the restricted capacity of working memory or the workspace of consciousness is subject to intense scientific investigation. In any case, capacity constraints limit the number of items simultaneously amenable to conscious processing. Which contents eventually reach the level of conscious awareness depends either on external cues that attract attention or on internal selection processes that direct attention either to external inputs or to material stored in memory. Most of the time subjects are not aware of performing such selections which gives rise to the impression that what surfaces in consciousness is all there is. Interestingly even conscious, intentional search for a content safely stored in declarative memory may fail to move that content into the workspace of consciousness. It is often a persistent non-conscious search process that suddenly lifts the searched

items into the workspace of consciousness. This indicates that access to consciousness is only partly under the control of the conscious agent itself.

## THE CLASSICAL EXPERIMENTAL PARADIGMS: RESULTS AND LIMITATIONS

The most frequently applied strategy for the identification of the NCCP consists of creating conditions in which physically identical stimuli are processed consciously only in a fraction of trials and then to subtract neuronal activation patterns associated with non-conscious processing from those generated during conscious processing. The assumption is that the remaining activation patterns are characteristic for conscious processing. However, as discussed by Aru et al. (2012b), this approach is fraught with numerous ambiguities.

The subtractive procedure uncovers not only the hypothetical NCCP proper. It reveals also the various processes that gate access to consciousness and the many processes that follow once subjects became aware of a stimulus. Among the latter are the transfer of information into working and episodic memory, the covert preparation of motor responses and in case of human subjects the covert verbalization of perceived contents. Because of dense reciprocal coupling between brain areas and the prevalence of parallel processing, segregation of these confounding factors is notoriously difficult with non-invasive recording techniques. Thus, there is the caveat that data obtained with this method may reflect not only the NCCP proper but also prerequisites for and consequences of conscious processing.

Here is an example: patients implanted with subdural electrodes over the visual cortex performed a recognition task in which the visibility of faces was manipulated either by increasing sensory evidence or providing a-priori knowledge (Aru et al., 2012a). The reasoning was that activity patterns specific for the NCCP proper should be the same irrespective of whether stimuli were consciously perceived because of enhanced sensory evidence or because of top-down facilitation. In trials in which conscious perception was caused by increasing sensory evidence there was indeed a category specific enhancement of gamma oscillations in the fusiform face area, suggesting that this increase in synchrony of neuronal responses had to do with conscious perception. However, this increase was lacking when sensory evidence was kept constant and visibility enhanced by prior knowledge. This suggested the conclusion "that the differential activation of specific areas of the visual cortex is a necessary but not a sufficient condition for conscious processing" (Aru et al., 2012a).

Other frequently applied paradigms for the identification of the NCCP manipulate the context of stimulus presentation with the aim to abolish conscious perception (reportability) of subsets of identical stimuli. This is achieved by exploiting interocular rivalry, masking paradigms, priming and variations of signal to noise ratios. So far, these approaches have yielded inconclusive results. Exploiting binocular rivalry, Leopold and Logothetis (1996) found that responses to perceived and non-perceived stimuli differ only at higher processing stages of the ventral stream. They concluded that activation of

neurons in the inferotemporal cortex, one of the highest levels in the visual processing hierarchy, qualified as NCCP. However, Fries et al. (1997) discovered in the primary visual cortex of cats that responses of cells to perceived stimuli differed from those to non-perceived stimuli because of increased synchronization of oscillatory responses in the gamma frequency band (Fries et al., 1997). This suggests that at this early stage of processing increased synchronization rather than increased discharge rate allows these responses to compete successfully with the conflicting inputs from the suppressed eye in the respective upstream target areas. These results agree with psychophysical and non-invasive tract tracing studies in human subjects, which indicate that interocular rivalry and hence the gating of access to consciousness involves already mechanisms in primary visual cortex (Genç et al., 2011). Finally, combining rivalry experiments with fMRI measurements in human subjects revealed reduced responses to the respective suppressed eye as early as in the lateral geniculate body, the thalamic relay for retinal signals (Haynes et al., 2005).

Thus, conscious perception seems to involve also early stages of sensory processing. This view is supported by the evidence that imagery, the visualization of imagined contents, is associated with increased BOLD activity in primary visual cortex and that lesions of primary visual cortex lead to blindsight. Patients with such lesions can still use visual information for orienting responses and avoidance of obstacles but they cannot consciously perceive visual stimuli (for review, see Goebel et al., 2001). These results indicate that appropriate activation of primary sensory areas of the cerebral cortex is a necessary prerequisite for the mediation of conscious perception but they do not allow the conclusion that this is a sufficient condition.

Another class of experiments makes use of clinical syndromes that go along with disturbances of conscious perception as is the case in patients with blind sight (Weiskrantz, 2004), neglect, agnosia, section of the commissures (split brain) or reduced states of consciousness (for discussion of some of these approaches see the other contributions of this volume).

Signals amenable to conscious processing can obviously also originate within the brain itself. Examples are signals associated with the recall of memories, imagery, decision making, planning, deliberating and reasoning. Thus, conscious experience appears to result from very versatile cognitive processes that can recruit neuronal activation patterns from many different sources and bind them together in a unified format.

Finally, indications for the substrate of the NCCP are derived from the evidence, that a plethora of signals from specialized receptor systems are excluded from conscious perception even though these signals are processed in great depth by the brain and exert strong control over behavior. Examples are interoceptive signals that maintain metabolic homeostasis, pheromone signals controlling reproductive behavior and signals for the synchronization of circadian rhythms. Unlike the classical five senses, these signaling systems are not represented by devoted cortical areas, supporting the view that cortical structures are involved in the mediation of conscious experience.

## TWO NON-EXCLUSIVE HYPOTHESES: ANATOMICAL SUBSTRATE VS. DYNAMICAL STATE

Current theories about the nature of the NCCP can be grouped into two major, non-exclusive clusters. The first assumes that particular brain structures have to be engaged to permit conscious processing. The idea is that these structures subserve what is sometimes addressed as the “inner eye function.” In this case it is assumed that the contents of conscious experience are represented and bound together by a distinct structure onto which the various processing streams would converge. This structure would have to be positioned at the top of the processing hierarchy. The second group of theories assume that conscious and non-conscious processes could involve the same anatomical substrate but differ with respect to dynamic states reflecting the degree of integration of distributed processes. As candidates for such state variables have been proposed temporal coherence, synchrony, correlation length and dimensionality.

### Binding in the Spatial Domain

Identification of brain structures whose activation is crucial for conscious processing is problematic if the distinguishing criterion is reportability. In this case a considerable number of brain structures and networks would qualify: the entire dominant hemisphere in split-brain patients, the parietal cortex in case of neglect, multiple sensory areas in case of agnosia, and ultimately the language system itself or structures required to access the language system. Split-brain experiments illustrate this problem. Stimulus material presented to the sensory space contralateral to the non-dominant hemisphere is often not reportable even though patients readily process the respective information and generate adapted motor responses. Rather than taking this as evidence that conscious processing is tied to the dominant, speech-competent hemisphere one could argue that the disconnection simply prevents the dominant, speech competent hemisphere from reporting. Although this disconnection jeopardizes language-dependent post-processing steps such as rational deliberation it would seem strange to deny the otherwise intact and awake non-dominant hemisphere the ability to sustain consciousness. If one were to reach this conclusion, one would have to deny that animals are conscious which is clearly untenable. The proposal that there is a special work space for conscious processing, promoted by Baars (1997) and later by Dehaene et al. (2006) also makes assumptions on the involvement of specific structures, in this case the ensemble of reciprocally coupled neuronal groups located in the supragranular layers of the cerebral cortex. However, it is difficult to provide causal evidence for this hypothesis because inactivation of supragranular layers would also jeopardize all the other functions of the cerebral cortex. The intuitively plausible hypothesis that the contents of conscious experiences are represented and bound together in a distinct structure at the top of the processing hierarchy is thus not well supported by experimental evidence. As argued by Dennet (1992), a region with such universal

“observer functions” would be theoretically implausible. Moreover, behavioral and brain imaging studies have shown that unconscious and conscious processing engage very much the same brain regions, including frontal and prefrontal cortex (Lau and Passingham, 2007; van Gaal et al., 2008). Which of the respective areas get recruited into functional networks depends more on the nature of the task than on the mode of processing.

## Binding in the Temporal Domain

Probably the first to propose that conscious and non-conscious processes could involve the same anatomical substrate but differ with respect to dynamic states reflecting the degree of integration of distributed processes was Sherrington (1906). He proposed that the unity of consciousness could be achieved by binding the contents of conscious experience in the temporal domain. In his book “The Integrative Action of the Nervous System” he stated: “Pure conjunction in time without necessarily cerebral conjunction in space lies at the root of the solution of the problem of the unity of mind.” He proposed that the unity of consciousness does not necessarily require anatomical convergence but could be achieved by convergence in time. The idea that temporal rather than spatial integration is a necessary prerequisite for conscious processing is at the basis of numerous recent theories and experimental evidence in favor of this notion keeps increasing.

Baars (1997) proposed that there is a special workspace for conscious processing and that subjects become aware of signals if these are sufficiently salient to ignite coordinated activity within this workspace. As mentioned above, Dehaene et al. (2006) and Gaillard et al. (2009) proposed the neuronal correlate of this workspace to be a widely distributed network of neurons located in the superficial layers of the cortical mantle. This network, so the assumption, would be “ignited” if a sufficient number of nodes were activated together.

Others have suggested that this “workspace” should be seen not so much as a sub-compartment of the cerebral cortex but as a special dynamic state of the brain that favors large scale *binding* of the results of widely distributed cortical computations (e.g., Varela et al., 2001; Melloni et al., 2007; Gaillard et al., 2009; Oizumi et al., 2014). According to the binding by synchrony hypothesis (Singer, 1993, 1999) it has been proposed that the respective dynamic state should be characterized by enhanced coherence of oscillatory activity in the beta or gamma frequency range. Using the subtraction method (see above) experimental evidence could be provided that processing of consciously perceived stimuli was indeed associated with better synchronization of large cortical networks in the beta and gamma frequency range than processing of identical stimuli that had failed to reach the threshold for awareness and remained invisible (Melloni and Rodriguez, 2007; Melloni et al., 2007; Gaillard et al., 2009; Melloni and Singer, 2010).

Contents that one is aware of are experienced as simultaneously present and related to each other. Thus, a mechanism is required that permits flexible and fast association of the ever-changing contents of conscious experience into a coherent whole. Dynamic binding by transient synchronization

of widely distributed processes could in principle fulfill such a function.

More recently a related hypothesis, the “Information Integration Theory” has been formulated by Tononi (2004; see also Oizumi et al., 2014). This theory also posits that conscious processing is associated with particularly effective and global integration of information from different sources. Supportive evidence for this conjecture comes from several independent observations. First, dynamic states that favor conscious processing such as arousal and attention facilitate the propagation of excitatory perturbations over larger cortical distances, which is likely to enhance interactions between distributed processes (Massimini et al., 2005). Second, arousal and attention facilitate synchronous oscillations in the gamma and high beta frequency range (Herculano-Houzel et al., 1999; Fries et al., 2001; Lima et al., 2011). This oscillatory patterning of activity, in turn, facilitates long-range synchronization (Roelfsema et al., 1997; for review, see Singer, 1999) and thereby enhances communication between remote groups of neurons and the formation of large scale functional networks (Roelfsema et al., 1997). A mechanistic account for the “binding” function of synchronization has been formulated in the “communication by coherence” (CTC) hypothesis (Fries, 2005) that has in the meantime received experimental support (Womelsdorf et al., 2007; Bastos et al., 2015).

The reportability that is considered as such a critical feature of conscious processing could thus be a natural consequence of a highly integrated processing mode. In human subjects integration of widely distributed processes would automatically involve the language network because of its particularly strong interconnections with both sensory and executive systems. Thus, reportability would simply be a consequence of processing modes characterized by a particularly high degree of integration but not a necessary requirement for conscious processing. This is a further argument for the notion that animals, at least those with highly evolved brains such as vertebrates, can switch between conscious and unconscious processing modes. The circuitry of their brains and the ensuing dynamics can certainly sustain highly integrated processes.

## THE HARD PROBLEM

Even if we had a full account of the NCCP, of the neuronal mechanisms whose involvement distinguishes conscious from non-conscious processing, the “hard problem” in consciousness research would persist. We still would have no explanation for the phase transition from material neuronal processes, described from a third-person perspective, to the immaterial mental phenomena, that we experience from our first-person perspective. In the following I shall attempt to narrow this explanatory gap by attempting a naturalistic, evolutionary explanation for the fact that many humans experience themselves as having both a material and an immaterial mental or spiritual dimension. The core of the argument is that the gap can probably not be closed if one considers only the cognitive functions of individual brains but that in addition the phenomena emerging from *social interactions* have to be taken into account. This

extension requires joint consideration of evidence and analyses from philosophy, cognitive neuroscience, cultural anthropology, developmental psychology and social science. The arguments are based on the assumptions that: (i) perceptions are the result of a constructivist process that depends on priors; (ii) social interactions lead to the emergence of a novel class of realities, the immaterial social and cultural realities; (iii) these cultural realities assume the role of priors for perception; and (iv) the construction of the self-model is based on experiences that are shaped by cultural priors.

## THE CONSTRUCTIVIST NATURE OF PERCEPTIONS

Abundant psychophysical and neuroscientific evidence indicates, that our perceptions are the result of complex computations in which sparse sensory evidence is interpreted on the basis of a huge amount of prior information about the world (for review, see Spratling, 2017). This information is contained in the functional architecture of the brain. Part of this knowledge has been acquired through evolutionary selection and is stored in the genes. This inherited knowledge is then complemented by experience-dependent development and adult learning. The knowledge acquired during evolution and early development is implicit, i.e., the perceiving agent is not aware of its existence although it plays an essential role in determining the agent's perception. Hence, the validity of the perceptions shaped by these implicit priors cannot be questioned. The perceiving agent cannot but take as real what she/he perceives. By contrast, priors acquired by learning later in life are to some extent amenable to conscious recollection and perceptions shaped by these explicit priors can be challenged by reasoning (conscious deliberations). A core assumption of the present proposal is, that all perceptions, regardless as to whether they result from stimuli in the external world or from introspection depend on priors.

## THE EMERGENCE OF SOCIAL REALITIES

Biological evolution has brought forth organisms with increasingly refined cognitive functions: the ability to develop a theory of mind by taking the perspective of the respective other, to generate abstract descriptions by recognizing similarities among seemingly different appearances through poly-sensory integration, to represent these abstractions in symbolic form and to eventually communicate the results of these cognitive functions. Recent studies suggest that evolved animals such as birds and mammals possess some of these abilities in various combinations and often in rudimentary forms. Humans excel because they possess all of these functions and in addition have developed language which allows them to communicate the results of their cognitive operations in a highly abstract and symbolic way. Once agents endowed with this unique combination of cognitive abilities began to cooperate and to communicate with each other, they began to create a new class of phenomena that the philosopher John Searle addressed as "social realities." These are immaterial realities that evolution brought into this world once development of sophisticated cognitive

abilities allowed organisms to engage in social interactions. The rules governing the coexistence of social animals can be regarded as rudimentary forms of such social realities. However, in human societies these immaterial realities are no longer just implicit forces that coordinate cooperativity but assumed the status of realities that became consciously perceivable as integral part of the world. Examples of social realities created by human societies are empathy, fairness, greed, love, devotion, shame, norms, vows, commitments, social status, values, belief systems, laws, regulations and moral imperatives. These are realities that cannot emerge from the cognitive abilities of individual brains alone but require for their creation the interaction of at least two cognitive agents. They are immaterial, mental constructs but they are real in the sense that they can readily be perceived and strongly influence behavior. Believes in the reality of these immaterial constructs erect cathedrals and motivate suicidal behavior.

How then could such immaterial realities have emerged? Here is a likely scenario. A group of cave dwellers sits around a fireplace and shares food. Sooner or later members of this group will discover that there are certain subjects who are more generous, or greedier than others. If these observations are shared by a sufficient number of group members, generosity and greed will eventually acquire the status of perceivable realities and then can be symbolically represented by a common description. This scenario suggests as necessary prerequisites for the emergence of novel, immaterial realities: the collective experience of intangible, immaterial phenomena, the mutual affirmation of the reality/existence of these phenomena through shared attention and experience, the naming of these phenomena and finally the representation of these immaterial realities in rituals and artistic creations. These artistic activities retranslate the immaterial realities into concrete symbols that are then perceivable with the classical five senses.

## SOCIAL REALITIES AS PRIORS OF PERCEPTION

The second core assumption is that these immaterial realities assume the function of priors for perception in very much the same way as all the other inherited and acquired a priori assumptions about the world and ourselves. As a consequence of this cognitive embedding in a world in which tangible and immaterial realities coexist humans are bound to perceive reality as consisting of two classes of phenomena, objects perceivable with one of the five senses and immaterial phenomena that cannot be directly perceived. Priors of perception such as e.g., the Gestaltrules are usually implicit, i.e., subjects are not aware of the priors that shape their perceptions. Hence, humans—and animals as well—are bound to take what they perceive as evident and real. This is likely the case also for the perceptions shaped by cultural priors. As a consequence, individuals perceive cultural realities as equally evident and concrete as objects of the material world. What is perceived is taken for granted and experienced as true—and since cultural priors are shared by the members of communities there is usually a broad consensus about the reality of the perceived. A natural consequence of this shared



perception of a dichotomous reality is the construction of an ontological substance dualism with its many different, culture-specific flavors.

## SUBSTANCE DUALISM AND THE SELF-MODEL

The third core assumption of the attempt to soften the hard problem of consciousness research is that the self-perception of the conscious Self with all its immaterial connotations is a consequence of perceptions shaped by priors provided by social realities. Humans experience/perceive themselves as autonomous, cognitive and intentional agents and observations of the actions of the respective other do not contradict but confirm this perception. Humans are aware of being able to perceive, to reason, to decide and to act and they can share through language the contents of this meta-awareness and through comparison and observation of the respective other assure themselves of the reality of these immaterial properties of the conscious Self. However, humans have neither access to the priors that shape the perception of an immaterial Self nor do they have access to the material neuronal processes that underlie their cognitive and executive functions. All that humans can experience are the consequences of their brains' actions and because they have been familiarized with the existence of immaterial realities they naturally postulate as cause of their actions an invisible, seemingly immaterial agent that is not constrained by the laws of nature and whom they equate with the conscious, intentional and responsible Self.

Thus, very much in the same way as proposed above for the construction of social realities, humans have shared their experiences on the existence of an immaterial agent, came to a similar conclusion, invented names for the various manifestations of this agent and assigned it to the realm of the intangible immaterial realities. Once this concept was commonly shared it likely assumed the status of a prior that henceforth shaped not only the perception of others but also the perception of one-self. In that way, it could have become seamlessly integrated in the self-model. This, in turn, could account for the fact that humans perceive themselves as existing both in a material and in an immaterial dimension. As perceptions based on implicit priors cannot be questioned and are experienced as true, human subjects are bound to take their dual nature, their existence in both a material and a spiritual domain for granted. Consequently all the properties of this immaterial agent, its feelings, intentions, beliefs, wishes and the contents of consciousness are attributed to the immaterial domain—in perfect agreement with the traditional view of ontological substance dualism.

The perception of a dualist reality is further reinforced by education, religions and shared belief systems. These normative systems emphasize the autonomy and independence of an immaterial agent that is unconstrained by the material, biological domain. They instrumentalize cultural priors in order to establish a self-model that permits to experience freedom and hence responsibility.

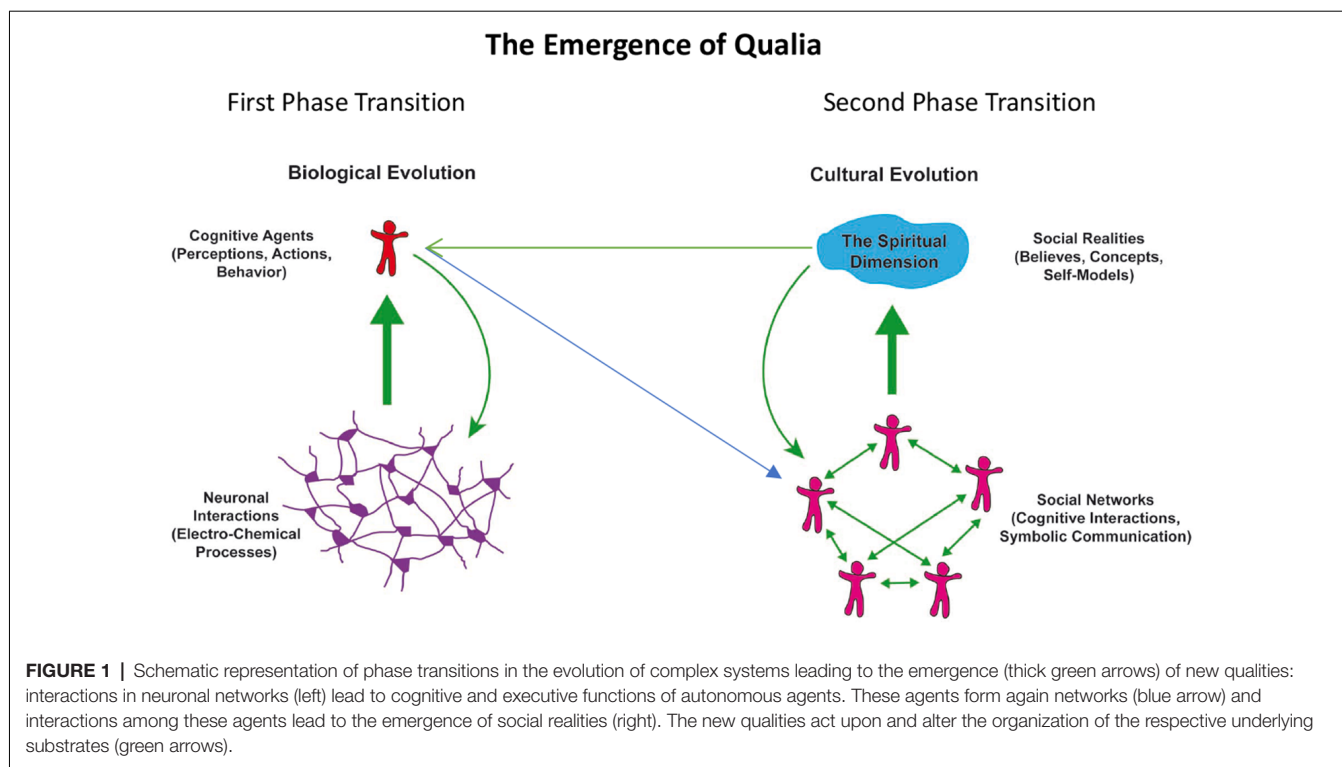
However, problems arose for this self-model once the natural sciences, in particular the neurosciences, set out to study the biological underpinnings of behavior and mental phenomena. One of these problems is the postulate of mental causation. If one adheres to ontological substance dualism, if one believes in the existence of an immaterial agent, the Self, that is independent of the neuronal processes in the brain and endowed with consciousness, intentions and free will, one needs to assume that this immaterial agent interacts with the brain so that the brain translates intentions and decisions into action. Such a scenario violates the known laws of nature, in particular, the law for the conservation of energy, because interactions with a material substrate require the exchange of energy. Per definition, however, the immaterial domain should be devoid of energy, otherwise it would again be part of the physical, the material domain.

The other, closely related problem is the “hard problem”, the explanatory gap between neuronal processes and the qualia of our experiences. The first problem vanishes if one is prepared to accept the overwhelming neurobiological and neuropsychological evidence that mental phenomena are the consequence and not the cause of neuronal interactions. The second problem is at least alleviated if one accepts: (i) that our perceptions are the result of a constructivist process that depends on priors; (ii) that this applies not only for perceptions of the outer world but also for perceptions based on introspection; and (iii) that social realities assume the role of priors for self-perception.

## THE EMERGENCE OF NEW QUALITIES FROM INTERACTIONS IN COMPLEX SYSTEMS

From an evolutionary perspective, the awareness/experience of a mental or spiritual dimension and its integration in our self-model can be understood as a sequence of evolutionary phase transitions that characterize complex, self-organizing systems. For illustration of this concept see **Figure 1**.

The first phase transition is the emergence of cognitive functions from complex interactions in neuronal networks. The second phase transition is the emergence of social realities from the complex interactions in networks of cognitive agents. In both cases the emergent phenomena, the cognitive functions generated by neural networks and the social realities generated by social networks, transcend the properties of the components of the respective networks. The emergent phenomena cannot be understood by only considering the properties of the respective network nodes and they cannot be described with the terms used for the description of the nodes. Hence different language systems had to be developed to capture the properties of the emergent phenomena. There is a language for the description of neuronal processes, another one for the characterization of the emergent cognitive and executive functions, yet another for the description of cognitive agents in their role as nodes in social networks and finally there is a language to capture the emergent social realities. These language systems are each represented by



different scientific disciplines, the second and third straddling the border between the natural sciences and the humanities and the fourth being entirely a domain of the humanities. Past attempts to relate consciousness to neuronal processes and to narrow the explanatory gap between the material neuronal processes and the qualia of subjective experience have considered only the first phase transition and by and large neglected the second—which is the likely reason why the gap is perceived as too large to be closed.

The arguments exposed in this contribution suggest that the attempts to identify the underpinnings of consciousness must not be confined to the analysis of the neuronal functions of individual brains but must include the domain of socio-cultural phenomena that are traditionally dealt with by the humanities. The present approach is partly based on assumptions whose validation is beyond my competence. These concern, in particular, the emergence of social realities, the function of social realities as priors of perception, the influence of these priors on self-perception and the constructivist nature of the processes that lead to the self-model. Some of these assumptions may have been verified already by empirical evidence in studies on cultural evolution, evolutionary anthropology and developmental psychology but in principle, they should all be amenable to empirical testing. Thus, research on the ill-defined explanandum “consciousness” seems ideally suited to bridge the still wide gap between the natural sciences and the humanities. If successful, such a comprehensive research agenda might be able to eventually settle the epistemic disputes on the nature of consciousness, on the problem of mental causation and on the relation between mind

and matter. As I have tried to show, this synthesis should be realizable within a naturalistic, evolutionary framework that is based on empirical evidence. However, it requires joint consideration of phenomena that emerged from phase transitions in a continuous evolutionary process that comprises both biological *and* cultural evolution. Although this approach is incompatible with ontological dualism and qualifies the spiritual dimension as a cognitive construct it leaves sufficient space for the precious immaterial entities that are constitutive for human identity.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## FUNDING

This work has been supported by the Max Planck Society (Max-Planck-Gesellschaft), the Ernst Strüngmann Institute for Neuroscience in Cooperation with the Max Planck Society and a Koselleck Grant from the German Research Foundation (Deutsche Forschungsgemeinschaft), administered by the Frankfurt Institute for Advanced Studies.

## ACKNOWLEDGMENTS

Special thanks go to Dr. Kathinka Evers for critical comments on an earlier version of this manuscript.

## REFERENCES

- Aru, J., Axmacher, N., Do Lam, A. T. A., Fell, J., Elger, C. E., Singer, W., et al. (2012a). Local category-specific gamma band responses in the visual cortex do not reflect conscious perception. *J. Neurosci.* 32, 14909–14914. doi: 10.1523/jneurosci.2051-12.2012
- Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012b). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746. doi: 10.1016/j.neubiorev.2011.12.003
- Baars, B. J. (1997). In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. *J. Conscious. Stud.* 4, 292–309.
- Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., et al. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85, 390–401. doi: 10.1016/j.neuron.2014.12.018
- Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., and Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn. Sci.* 10, 204–211. doi: 10.1016/j.tics.2006.03.007
- Dehaene, S., Naccache, L., Le Clec, H. G., Koehlin, E., Mueller, M., Dehaene-Lambertz, G., et al. (1998). Imaging unconscious semantic priming. *Nature* 395, 597–600. doi: 10.1038/26967
- Dennet, D. C. (1992). *Consciousness Explained*. London: Penguin.
- Dennett, D. C. (2018). Facing up to the hard question of consciousness. *Phil. Trans. R. Soc. B Biol. Sci.* 373:20170342. doi: 10.1098/rstb.2017.0342
- Doricchi, F., Thiebaut de Schotten, M., Tomaiuolo, F., and Bartolomeo, P. (2008). White matter (dis)connections and gray matter (dys)functions in visual neglect: gaining insights into the brain networks of spatial awareness. *Cortex* 44, 983–995. doi: 10.1016/j.cortex.2008.03.006
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011
- Fries, P., Reynolds, J. H., Rorie, A. E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291, 1560–1563. doi: 10.1126/science.1055465
- Fries, P., Roelfsema, P. R., Engel, A. K., König, P., and Singer, W. (1997). Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proc. Natl. Acad. Sci. U S A* 94, 12699–12704. doi: 10.1073/pnas.94.23.12699
- Fu, Z., and Rutishauser, U. (2018). Single-neuron correlates of awareness during attentional blinks. *Trends Cogn. Sci.* 22, 5–7. doi: 10.1016/j.tics.2017.10.005
- Gaillard, R., Dehaene, S., Adam, C., Clemenceau, S., Hasboun, D., Baulac, M., et al. (2009). Converging intracranial markers of conscious access. *PLoS Biol.* 7:e61. doi: 10.1371/journal.pbio.1000061
- Genç, E., Bergmann, J., Tong, F., Blake, R., Singer, W., and Kohler, A. (2011). Callosal connections of primary visual cortex predict the spatial spreading of binocular rivalry across the visual hemifields. *Front. Hum. Neurosci.* 5:161. doi: 10.3389/fnhum.2011.00161
- Goebel, R., Muckli, L., Zanella, F. E., Singer, W., and Stoerig, P. (2001). Sustained extrastriate cortical activation without visual awareness revealed by fMRI studies of hemianopic patients. *Vision Res.* 41, 1459–1474. doi: 10.1016/s0042-6989(01)00069-4
- Haynes, J.-D., Deichmann, R., and Rees, G. (2005). Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature* 438, 496–499. doi: 10.1038/nature04169
- Herculano-Houzel, S., Munk, M. H. J., Neuenschwander, S., and Singer, W. (1999). Precisely synchronized oscillatory firing patterns require electroencephalographic activation. *J. Neurosci.* 19, 3992–4010. doi: 10.1523/jneurosci.19-10-03992.1999
- Lau, H. C., and Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *J. Neurosci.* 27, 5805–5811. doi: 10.1523/JNEUROSCI.4335-06.2007
- Leopold, D. A., and Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379, 549–553. doi: 10.1038/379549a0
- Lima, B., Singer, W., and Neuenschwander, S. (2011). Gamma responses correlate with temporal expectation in monkey primary visual cortex. *J. Neurosci.* 31, 15919–15931. doi: 10.1523/JNEUROSCI.0957-11.2011
- Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. doi: 10.1126/science.1117256
- Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W., and Rodriguez, E. (2007). Synchronization of neural activity across cortical areas correlates with conscious perception. *J. Neurosci.* 27, 2858–2865. doi: 10.1523/jneurosci.4623-06.2007
- Melloni, L., and Rodriguez, E. (2007). Non-perceived stimuli elicit local but not large-scale neural synchrony. *Perception* 36:116.
- Melloni, L., and Singer, W. (2010). “Distinct characteristics of conscious experience are met by large-scale neuronal synchronization,” in *New Horizons in the Neuroscience of Consciousness*, eds E. Perry, D. Collerton, F. LeBeau and H. Ashton (Amsterdam: John Benjamins, B.V.), 17–28.
- Nagel, T. (1974). What is it like to be a bat? *Philosoph. Rev.* 83, 435–450. doi: 10.2307/2183914
- Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput. Biol.* 10:e1003588. doi: 10.1371/journal.pcbi.1003588
- Roelfsema, P. R., Engel, A. K., König, P., and Singer, W. (1997). Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature* 385, 157–161. doi: 10.1038/385157a0
- Sherrington, C. S. (1906). *The Integrative Action of the Nervous System*. New York, NY: Charles Scribner's Sons.
- Simons, D. J., and Chabris, C. F. (1999). Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception* 28, 1059–1074. doi: 10.1068/p2952
- Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. *Annu. Rev. Physiol.* 55, 349–374. doi: 10.1146/annurev.ph.55.030193.002025
- Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24, 49–65. doi: 10.1016/s0896-6273(00)80821-1
- Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain Cogn.* 112, 92–97. doi: 10.1016/j.bandc.2015.11.003
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.* 5:42. doi: 10.1186/1471-2202-5-42
- van Gaal, S., Ridderinkhof, K. R., Fahrenfort, J. J., Scholte, H. S., and Lamme, V. A. (2008). Frontal cortex mediates unconsciously triggered inhibitory control. *J. Neurosci.* 28, 8053–8062. doi: 10.1523/jneurosci.1278-08.2008
- Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronisation and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239. doi: 10.1038/35067550
- Weiskrantz, L. (2004). Roots of blindsight. *Prog. Brain Res.* 144, 229–241. doi: 10.1016/s0079-6123(03)14416-0
- Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., et al. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science* 316, 1609–1612. doi: 10.1126/science.1139597

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Singer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# COALIA: A Computational Model of Human EEG for Consciousness Research

Siouar Bensaid<sup>†</sup>, Julien Modolo<sup>†</sup>, Isabelle Merlet, Fabrice Wendling\* and Pascal Benquet

INSERM, Laboratoire Traitement du Signal et de l'Image (LTSI)—U1099, University of Rennes, Rennes, France

## OPEN ACCESS

### Edited by:

Igor Timofeev,  
Laval University, Canada

### Reviewed by:

Marco Atzori,  
Universidad Autónoma de San Luis  
Potosí, Mexico  
Jamie Sleigh,  
The University of Auckland,  
New Zealand

### \*Correspondence:

Fabrice Wendling  
fabrice.wendling@inserm.fr

<sup>†</sup>Co-first authors

**Received:** 08 March 2019

**Accepted:** 07 October 2019

**Published:** 13 November 2019

### Citation:

Bensaid S, Modolo J, Merlet I,  
Wendling F and Benquet P  
(2019) COALIA: A Computational  
Model of Human EEG for  
Consciousness Research.  
*Front. Syst. Neurosci.* 13:59.  
doi: 10.3389/fnsys.2019.00059

Understanding the origin of the main physiological processes involved in consciousness is a major challenge of contemporary neuroscience, with crucial implications for the study of Disorders of Consciousness (DOC). The difficulties in achieving this task include the considerable quantity of experimental data in this field, along with the non-intuitive, nonlinear nature of neuronal dynamics. One possibility of integrating the main results from the experimental literature into a cohesive framework, while accounting for nonlinear brain dynamics, is the use of physiologically-inspired computational models. In this study, we present a physiologically-grounded computational model, attempting to account for the main micro-circuits identified in the human cortex, while including the specificities of each neuronal type. More specifically, the model accounts for thalamo-cortical (vertical) regulation of cortico-cortical (horizontal) connectivity, which is a central mechanism for brain information integration and processing. The distinct neuronal assemblies communicate through feedforward and feedback excitatory and inhibitory synaptic connections implemented in a template brain accounting for long-range connectome. The EEG generated by this physiologically-based simulated brain is validated through comparison with brain rhythms recorded in humans in two states of consciousness (wakefulness, sleep). Using the model, it is possible to reproduce the local disinhibitory disinhibition of basket cells (fast GABAergic inhibition) and glutamatergic pyramidal neurons through long-range activation of vasoactive intestinal-peptide (VIP) interneurons that induced inhibition of somatostatin positive (SST) interneurons. The model (COALIA) predicts that the strength and dynamics of the thalamic output on the cortex control the local and long-range cortical processing of information. Furthermore, the model reproduces and explains clinical results regarding the complexity of transcranial magnetic stimulation TMS-evoked EEG responses in DOC patients and healthy volunteers, through a modulation of thalamo-cortical connectivity that governs the level of cortico-cortical communication. This new model provides a quantitative framework to accelerate the study of the physiological mechanisms involved in the emergence, maintenance and disruption (sleep, anesthesia, DOC) of consciousness.

**Keywords:** computational modeling, brain connectivity, feedforward inhibition, GABA, disinhibition, TMS-EEG, disorders of consciousness (DOC)



## INTRODUCTION

The characterization and understanding of the mechanisms underlying consciousness is one, if not the greatest challenge that contemporary neuroscience is facing. Beyond the purely fundamental interest of this question, a major clinical issue is also at stake: evaluating residual consciousness in patients suffering from Disorders of Consciousness (DOC), which can be extremely difficult, and have crucial implications in terms of clinical care. For example, motor imagery paradigms can reveal covert consciousness in coma patients, using functional magnetic resonance imaging (fMRI; Owen et al., 2006) for instance. This illustrates the pressing need for an improved characterization of the mechanisms that underlie consciousness, which could be exploited to propose novel quantified measures, or *metrics*, of consciousness.

Many theories have attempted, at various levels of description, to integrate the multifaceted aspects of consciousness. One of the first theories that found a significant echo in the neuroscience community is the Dynamic Core Hypothesis (DCH; Tononi and Edelman, 1998), which was the first to relate the concept of information with consciousness. In this theory, functional clusters in the thalamocortical system are central, and involve fast re-entrant interactions, as well as a high level of integration and differentiation giving rise to complex patterns of neuronal activity. Another popular theory that has been gradually expanded over the years, and that has solid ties with neurophysiology, is the Global Workspace Theory (GWT; Dehaene et al., 1998; Dehaene and Changeux, 2011). In short, GWT states that conscious information is globally available within the brain, and that the “ignition” of large-scale networks, i.e., the sudden communication between distant brain regions to engage into the processing of information, enables a stimulus to reach the global workspace, hence consciousness. Ignition is thought to involve long-range glutamatergic fibers that enable long-distance communication between cortical regions. Several experiments have supported GWT, for example that non-masked words involve the activation of much wider networks as compared to masked words (Dehaene et al., 2001), with a similar result found for sub-liminal vs. supra-threshold visual stimuli (Modolo et al., 2018; van Vugt et al., 2018). The Integrated Information Theory (IIT) is based on a different approach. Instead of beginning from the large-scale structure of the thalamo-cortical system as in the DCH and GWT, IIT introduces several axioms to derive general principles of consciousness. One of the leading ideas of IIT is that consciousness involves the *integration* of information between distant areas (reminiscent of ignition in GWT), which increases the complexity of the processed information. Segregation, or *differentiation* of information, is also key in IIT: for example, large-scale synchronization of several regions with the same activity is indeed integrated, however with low complexity (Koch et al., 2016). Therefore, *integration* and *differentiation* appear as the two concepts leading to increasing the *complexity* of the information conveyed by brain-scale networks. A more recent theory named algorithmic information theory of consciousness a.k.a.

Kolmogorov Theory (KT; Ruffini, 2017) is also based on the idea that conscious states are associated with higher levels of complexity, and that subjective experience occurs following processes of information compression.

In contrast with the aforementioned theoretical studies of consciousness, only few studies have actually attempted to simulate brain activity associated with consciousness states using neurophysiologically-plausible computational models. Obviously, the *in silico* implementation of neural mechanisms that underlie the emergence and maintenance of consciousness represents a considerable challenge. However, capturing the main features of the most significant common principles from the main theories of consciousness using a computational neuroscience framework appears at reach. For example, a computational model exploring how cortico-cortical connectivity is functionally impaired during sleep (“connectivity breakdown”) has been proposed (Esser et al., 2009). However, most of these models are limited in terms of spatial scale and the represented micro-circuitry. This limitation hinders bridging the micro-circuit scale with the brain-scale, which is of interest in consciousness. The present study proposes to fill this gap, and provides new links between different levels of description (from local neuronal population to whole-brain scale).

Using a bottom-up approach, we developed COALIA, a new computational model of brain-scale electrophysiological activity. This is neither the intention nor objective of the present article to investigate the philosophical question of “qualia” that represents the perception properties. Instead, the model name “COALIA” has been chosen to ground this brain-scale computational model in the field of consciousness research. The model starts from neuronal micro-circuits involving different cellular subtypes that have been reliably identified through neurobiological studies. The basic unit of the model is the neural mass, representing a local population of a few thousands of neurons, which has proven its ability to capture the dynamics of actual neuronal assemblies (Wendling et al., 2002). At the local level, the model includes subsets of pyramidal neurons (glutamatergic), and three different types of interneurons (GABAergic) with appropriate physiologically-based kinetics (fast vs. slow). At the global level, the large-scale model is then constructed on the basis of a standard 66-region brain atlas (Desikan et al., 2006), with one neural mass representing the local field activity of one atlas region. Neural masses are spatially distributed over the cortex, using the template brain morphology (Colin). As they account for distinct cortical regions, neural masses are synaptically connected through long-range glutamatergic projections among pyramidal neurons and GABAergic interneurons. Connectivity is derived from DTI (Diffusion Tensor Imaging) data. Results show that the model captures the large-scale structure of brain connectivity between regions while accounting for the properties of local micro-circuits. It can accurately reproduce EEG activity for different conscious states (e.g., sleep vs. wake), and the breakdown of functional connectivity during sleep as assessed through the replication of TMS-EEG experiments.

In this article, we first describe basic concepts of cortico-cortical and thalamo-cortical networks involved in theories of

consciousness. Then, we present the neural mass modeling approach used to develop the brain-scale model, along with the various microcircuits considered and their functional role. A toy model involving a limited number of neuronal populations is then investigated in order to validate the implemented micro-circuits, before an extension to the whole-brain model. Simulated responses to TMS are generated and quantified in two consciousness states, namely awake and asleep. Results are discussed according to the novelty, performance and limitations of the model, along with its usefulness in consciousness studies. Future extensions are described.

## BACKGROUND: ROLE OF CORTICO- AND THALAMO-CORTICAL NETWORKS IN CONSCIOUSNESS

Consciousness is a global functional state of the brain that is intrinsically linked with neuronal oscillations generated by large-scale cortico-cortical and thalamo-cortical networks (Linás et al., 1998). More specifically, wakefulness is determined by widespread thalamocortical projections (Timofeev and Steriade, 1996; Laureys, 2004) while awareness requires the activation of a wide cortico-cortical network, involving lateral and medial frontal regions parieto-temporal and posterior parietal areas, bilaterally (Laureys et al., 1999). In the model, we included these two key components of consciousness that are briefly reviewed below.

### “Horizontal” Cortico-Cortical Connectivity

Functional connectivity studies have shed light on the functional networks involved in various conscious states (Jin and Chung, 2012). During general anesthesia-induced loss of consciousness, there is a breakdown in cortical effective connectivity (Ferrarelli et al., 2010; Hudetz, 2012; Gómez et al., 2013). As a reminder, effective connectivity is defined as the ability of a neuronal group to causally affect the firing of other neuronal groups (Friston, 2011). In unresponsive patients, impaired consciousness was associated with altered effective connectivity (Varotto et al., 2014; Crone et al., 2015). A protocol of TMS triggered a simple local EEG response indicating a breakdown of effective connectivity at the cortical level, similar to the one previously observed in unconscious sleeping or anesthetized subjects (Casali et al., 2013). Sleep stages have a drastic impact on consciousness and also on functional connectivity. For example, there is a strong reduction of both wakefulness and awareness components of consciousness during NREM sleep, associated with thalamic up-and-down state and cortical slow wave sleep (SWS). However, brain effective connectivity changes significantly (Tononi and Sporns, 2003; Esser et al., 2009). More specifically, cortical activations become more local and stereotypical upon falling into NREM sleep, which indicates an impaired effective cortical connectivity (Massimini et al., 2010). “Horizontal” communication through coherence that involves high frequency oscillations synchronization is a fundamental mechanism in cortical function and perception (Fries, 2005, 2009). Overall, the “awareness” component of consciousness depends on large-scale synchronized communication among

distant neuronal populations distributed over the neocortex (see **Figure 1A**, left panel).

### “Vertical” Thalamo-Cortical Connectivity

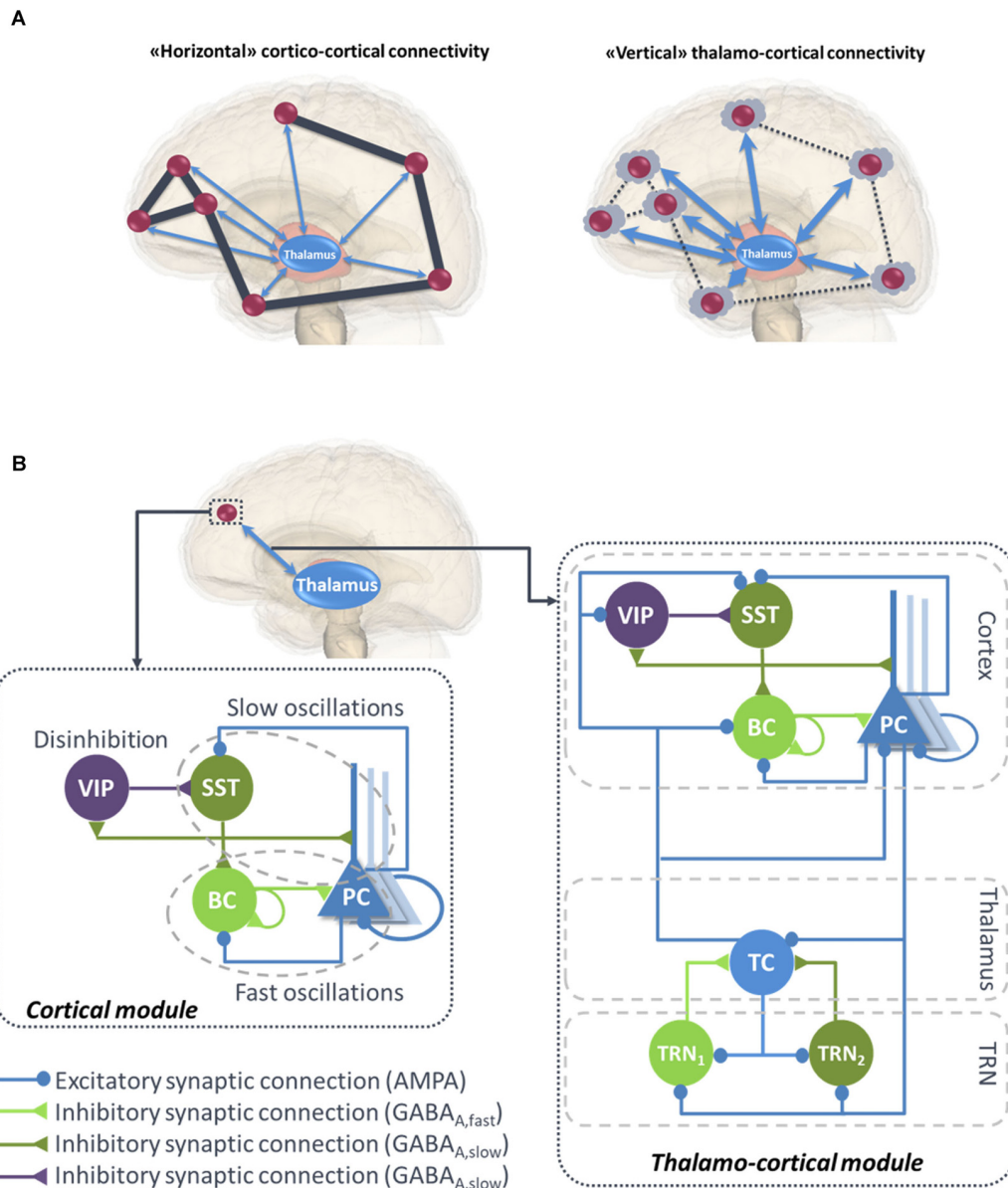
As reported in IIT (Tononi, 2004, 2012), consciousness depends on the brain’s ability to integrate information, which relies on the effective connectivity among functionally specialized regions (or clusters) of the thalamocortical system, and on the segregation of information. One important modulator of cortical connectivity is the activity pattern of thalamocortical cells, tonic vs. up- and down, which is able to modify the excitability level of cortical neuron subpopulations. The thalamic-mediated synchronization of distant cortical areas may coordinate the large-scale integration of information across multiple cortical circuits, consequently influencing the level of arousal and consciousness (Saalmann, 2014). Conversely, during sleep or anesthesia-induced transitions in consciousness, both thalamo-cortical and intra-thalamic functional connectivity are modified (Kim et al., 2012; Hale et al., 2016). In addition, thalamic input to neocortex modifies cortico-cortical connectivity. Upon falling into NREM sleep (when rhythmic thalamo-cortical up-and-down activity develops), cortical activations become more local and stereotypical, indicating a significant decrease of cortico-cortical connectivity (Esser et al., 2009; Massimini et al., 2010; Usami et al., 2015). Recently, it was shown that direct and tonic optogenetic activation of thalamic reticular nuclei (TRN) GABAergic interneurons induces a spatially restricted cortical slow-wave activity (Lewis et al., 2015). This activity was reminiscent of sleep rhythms, and animals exhibited behavioral changes that were consistent with a decrease of arousal.

Overall, this brief literature review suggests that both components of consciousness, namely awareness and wakefulness, are impaired when large-scale cortico-cortical functional connectivity mediated through the binding of synchronized high-frequency oscillations in the beta-gamma band (Schoenberg et al., 2018) and regulated through the thalamus (Nakajima and Halassa, 2017). Meanwhile, during this decrease of awareness and wakefulness, an increase of “vertical” thalamo-cortical connectivity is observed, along with a stronger synchronization of delta oscillations between thalamic cell (TC) assemblies and isolated groups of neocortical neurons (Hill and Tononi, 2005; **Figure 1A**, right panel).

## MATERIALS AND METHODS

### Modeling of Micro- and Macro-Circuits: Neural Mass Model Approach

Neural mass models (NMMs) are a mathematical description of neural dynamics at a mesoscopic scale (from a millimeter to several centimeters of the cortex). This class of models was proposed in the 1970s as an alternative to detailed microscopic models that require a more extensive computational cost (Wilson and Cowan, 1973; Nunez, 1974; Lopes da Silva et al., 1976; Freeman, 1978). NMMs can indeed model the local field potential (LFP) of an entire cortical region using only few state variables (Breakspear, 2017), whereas in detailed



**FIGURE 1 | (A)** Illustration of “horizontal” and “vertical” connectivity. *Left*: Horizontal connectivity refers to the cortico-cortical connections which are functionally effective during wakefulness, with a weak level of thalamo-cortical coupling. *Right*: Vertical connectivity refers to thalamo-cortical projections that functionally impair cortico-cortical connectivity during sleep. **(B)** General architecture of the micro- and macro-circuits implemented in the computational model. The local neural mass model (NMM) of cortical activity is composed of a PC exciting two GABAergic interneurons, namely, the somatic-projecting BC and the dendritic-projecting SST, responsible for the generation of fast and slow oscillations, respectively; the VIP were introduced as they play a crucial role in cortical column communication through disinhibition of SST. The subcortical module consisted in TC sending excitatory glutamatergic projections to the TRN block composed of fast and slow GABAergic interneurons TRN<sub>1</sub> and TRN<sub>2</sub>, respectively. VIP, vasoactive intestinal peptide positive GABAergic interneurons; SST, Somatostatin-positive GABAergic interneurons; BC, Basket-type GABAergic interneurons; PC, Glutamatergic Pyramidal Cells.

models this activity is meticulously described at the level of spatially distributed and interconnected neuron models, each including the properties of ionic channels, axons and dendrites (Wang and Buzsáki, 1996; Whittington et al., 2000; Maex and De Schutter, 2007).

Despite their simplicity, NMMs are neurophysiologically grounded, since they include the connectivity, synaptic kinetics

and firing rates of neuronal sub-types present in the region of interest. The reduced complexity and performance (in term of reproducing actual LFPs) of the NMM approach made it a powerful tool to investigate various cerebral mechanisms, such as the generation of brain rhythms (Jansen and Rit, 1995; David and Friston, 2003; Ursino et al., 2010). NMMs have also been extensively used to study pathological dynamics



such as in epilepsy (Wendling et al., 2002; Traub et al., 2005; Molaee-Ardekani et al., 2010; for a review, see Wendling et al., 2016), Alzheimer's disease (AD; Bhattacharya et al., 2011) and Parkinson's disease (PD; Liu et al., 2016, 2017).

Designing a NMM involves identifying the main subsets of neurons implied in the modeled brain tissue and describing their synaptic interconnections. Based on an exhaustive literature review, we developed a model consisting of coupled NMMs able to simulate both cortical and thalamic activity, as described hereafter.

## A Local Neural Mass Model of Neocortical Activity

Until recently, the classification of GABAergic interneurons was a highly challenging task, regarding the electrophysiological properties, morphology, biochemistry markers and connectivity. However, all recent studies converge towards much simpler functional categories, involving three main classes of GABAergic interneurons in the neocortex (Rudy et al., 2011). As briefly reviewed below, these classes account for: (1) somatic-targeting parvalbumine positive (PV<sup>+</sup>) basket cells (BC); (2) dendritic-targeting somatostatin positive (SST) interneurons; and (3) vasoactive intestinal-peptide (VIP) expressing interneurons (Tremblay et al., 2016; **Figure 1B**, left panel).

### Basket Cells and Fast Oscillations

PV<sup>+</sup> participate to the generation of cortical gamma oscillations through: (1) thalamocortical feedforward inhibition in layer 4; (2) feedback inhibition in layer 2/3; and (3) *via* direct PV<sup>+</sup>/PV<sup>+</sup> coupling through electrical gap-junctions (Povysheva et al., 2008; Buzsáki and Wang, 2012; Lewis et al., 2012; Varga et al., 2014; Womelsdorf et al., 2014; Chen et al., 2017). Recent optogenetic experiments have demonstrated the causal role of somatic-targeting interneurons BC in mediating fast oscillations (>20 Hz; Chen et al., 2017).

### Surround Inhibition by SST GABAergic Interneurons and Slow Oscillations

Neocortical SST neurons can exhibit high levels of spontaneous slow oscillations, and their tonic activity might facilitate fine scale up- and down-regulation of global inhibition levels in the neocortex (Urban-Ciecko and Barth, 2016). Lateral inhibition has been shown to be a fundamental principle in neuronal networks (Harris and Mrsic-Flogel, 2013; Karnani et al., 2014; Harris and Gordon, 2015). Specifically, lateral inhibition between nearby pyramidal cells (PCs) is thought to work through SST interneurons (Kapfer et al., 2007; Silberberg and Markram, 2007; Adesnik et al., 2012). Dendritic inhibition is more effective than perisomatic inhibition in regulating excitatory synaptic integration, therefore SSTs are the key regulator of input-output transformations (Lovett-Barron et al., 2012). Furthermore, the slow kinetics of dendritic-targeting inhibitory postsynaptic membrane potentials (IPSPs) are particularly suited to maximize the localized shunting inhibition effect (Gidon and Segev, 2012; Paulus and Rothwell, 2016). Finally, it was shown that this cell type is especially involved in slow oscillations (Womelsdorf et al., 2014; Urban-Ciecko and Barth, 2016; Funk et al., 2017).

## Cortical Column Communication Through VIP-Controlled Disinhibition

The disinhibition of cortical PCs gates information flow through and between cortical columns (Walker et al., 2016). One of the factors underlying this PC disinhibition is the inhibition of SST cells during active cortical processing, which enhances distal dendritic excitability (Gentet et al., 2012). The activation of VIP neurons strongly inhibits dendritic-targeting SST interneurons mediating surround inhibition, which leads to PCs disinhibition (suppress the inhibition on PCs; Lee et al., 2013; Pi et al., 2013; Fu et al., 2014; Pfeffer, 2014; Yang et al., 2016). This disinaptic mechanism of disinhibition seems to be a generic motif able to suppress the blanket of inhibition mediated by SST neurons (Fino and Yuste, 2011; Karnani et al., 2014). It has been demonstrated indeed in motor, sensory and associative neocortical areas that transient SST activity suppression by VIP activation occurs during visual processing (Lee and Mihalas, 2017), somatosensory integration (Lee et al., 2013; Sohn et al., 2016), locomotion (Dipoppa et al., 2018), top-down modulation (Ayzenshtat et al., 2016) and plasticity during perceptual learning (Williams and Holtmaat, 2019). Since VIP neurons are targeted by long-range cortical glutamatergic projections, they represent a key factor for distal cortico-cortical activation through disinaptic disinhibition.

### Formal Description of Neocortical Model

Based on the above information, we have developed a neocortical module involving PCs and three types of inhibitory subpopulations, namely, BC, SST and VIP (see **Figure 1B**, left panel). BC and SST receive excitatory inputs from PCs that are reciprocally inhibited by both of them. Pyramidal collateral excitation was also implemented *via* an excitatory feedback loop passed by a supplementary excitatory population (PC') analogous to PC, except that it projects only from and to subpopulation PC. The electrical gap-junction mentioned in "Basket Cells and Fast Oscillations" section was implemented through an inhibitory feedback loop characterized by a connectivity constant  $C_{BC}^n$ , where  $n$  is the index of the NMM (e.g.,  $n = 23$ th neural mass). Communication through disinhibition mentioned in "Cortical Column Communication Through VIP-Controlled Disinhibition" section was modeled by inhibitory projections; first from VIP to SST, and second, from the latter to BC. The nonspecific influence from neighboring and distant populations was modeled by a Gaussian input noise corresponding to an excitatory input  $p_c^n(t)$  that globally describes the average density of afferent action potentials. The set of ordinary differential equations (ODEs) modeling the neocortical module is included as a **Supplementary Material** (see Supplementary section "ODEs of the Neocortical Module").

### A Local Neural Mass Model of Thalamic Activity

The thalamus is considered as a complex relay extensively connected with the cortex, as well as most subcortical areas. This central position underlies its key role in several cognitive functions including perception, attention, memory and consciousness. Importantly, even a limited damage in



the thalamus can have major consequences on all the aforementioned functions (Ward, 2013). The thalamo-cortical circuitry as a neural correlate of consciousness has been mentioned by consciousness theories. More evidence for this role was provided by a study analyzing the metabolic activity of posterior midline cortical areas driven by the thalamic nuclei, across different altered consciousness states (Laureys et al., 2006). The authors reported that unresponsive wakefulness syndrome (UWS) patients can be differentiated from minimally conscious state (MCS) patients by a difference in glucose metabolism in these areas. The pivotal role of thalamo-cortical loop in the generation of slow waves and the up-and-down state was appraised in the review of Crunelli et al. (2015) where they enumerated the main lines of evidence supporting this assertion, namely, the strong interconnection between thalamic cells (TCs) and neocortical layers involved in slow waves suggesting that thalamic nuclei can control up-and-down state dynamics in neocortical circuits, the early TC firing in relation to the initiation of cortical UP states, the rhythmic up-and-down state generated by TC neurons and TRN in isolated conditions, and the neocortical UP states readily induced in head-restrained mice by selective optogenetic activation of TC neurons proving the stimuli role of the latter.

The importance of thalamo-cortical connectivity has motivated the development of thalamo-cortical models simulating the interactions between the cortex and thalamus at a mesoscopic level (Suffczynski et al., 2004; Sotero et al., 2007; Bhattacharya et al., 2011; Roberts and Robinson, 2012; Mina et al., 2013; Sen Bhattacharya et al., 2013; Cona et al., 2014). While some models were developed to generate alpha activity (8–12 Hz; Sotero et al., 2007; Bhattacharya et al., 2011; Sen Bhattacharya et al., 2013), others were used to simulate the sleep-wake cycle (Suffczynski et al., 2004; Roberts and Robinson, 2012; Cona et al., 2014).

Since our aim is to develop a computational model able to reproduce brain rhythms corresponding to different consciousness states (e.g., sleep-wake cycle), the inclusion of the thalamus in the model is crucial. A description of the thalamus model is provided in the right panel of **Figure 1B**. The thalamic module includes one population of excitatory glutamatergic neurons TCs, and two inhibitory interneurons from the TRN, TRN<sub>1</sub> and TRN<sub>2</sub> accounting for fast and slow GABAergic IPSPs, respectively. TCs receive GABAergic IPSPs with slow and fast kinetics from the TRNs, whereas the latter receive excitatory inputs from the former. Similarly to the cortical module, a Gaussian input noise corresponding to excitatory input  $p_{Th}^n(t)$  was used to represent nonspecific inputs on TCs. The set of ODEs modeling the thalamic module is provided as a **Supplementary Material** (see Supplementary section “ODEs of the Thalamic Module”).

## Modeling of Large-Scale Cortico-Cortical and Thalamo-Cortical Connectivity

### Cortico-Cortical Connections

PCs originating from a single cortical column target several cell types in distant cortical columns. Glutamatergic PCs target

not only remote PCs by common feedforward excitation but also GABAergic cells by disynaptic cortico-cortical feedforward inhibition (FFI; see **Figure 2**). Even if PV<sup>+</sup> BCs have been shown to receive stronger thalamocortical and intracortical excitatory inputs than SST neurons, it appears that cortico-cortical FFI could be mediated by both types of interneurons (Ma et al., 2010; Tremblay et al., 2016). Importantly, cortico-cortical glutamatergic axons also target VIP neurons (Sohn et al., 2016).

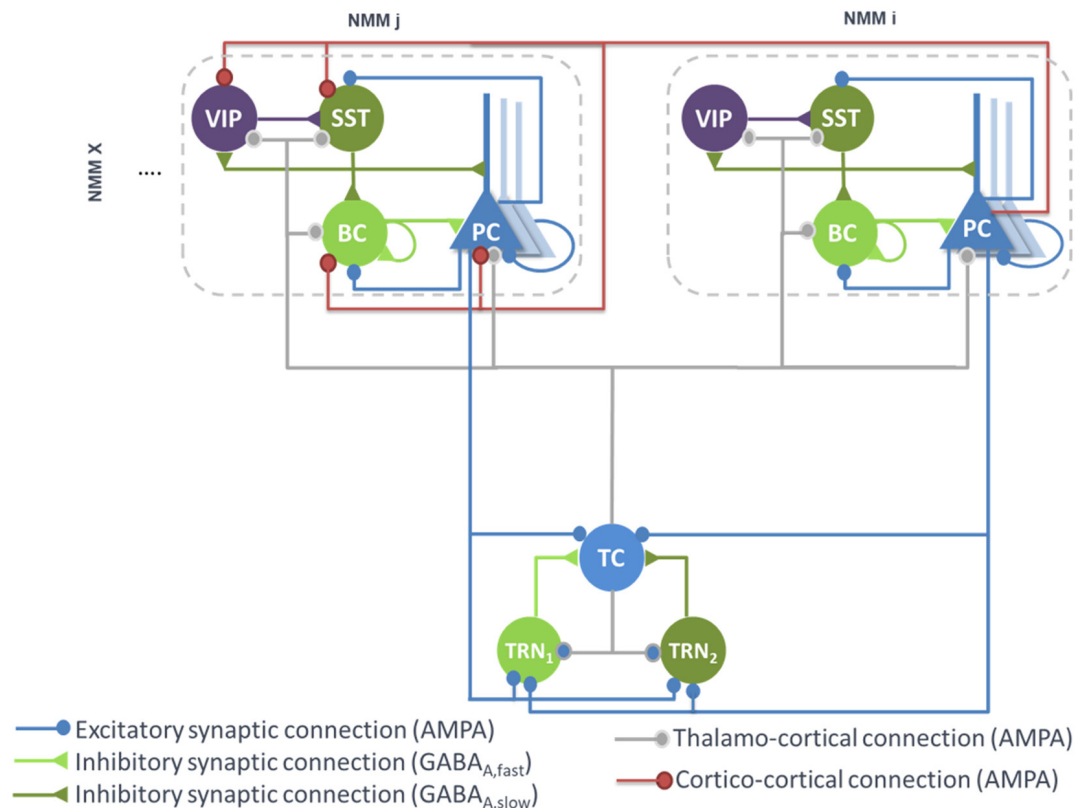
Therefore, feedforward excitation was included in the model by means of a connectivity constant  $K_{P,P}^{ij}$  modeling the average strength of glutamatergic projections from pyramidal subpopulations in NMM “*i*” to their counterpart in NMM “*j*.” Disynaptic cortico-cortical feedforward inhibition was similarly integrated *via* the connectivity constants  $K_{P,BC}^{ij}$ ,  $K_{P,SST}^{ij}$  and  $K_{P,VIP}^{ij}$  denoting glutamatergic projections from PC subpopulations in NMM “*i*” to BC, SST and VIP in NMM “*j*,” respectively. The set of ODEs modeling the cortico-cortical connections is included as a **Supplementary Material** [see Supplementary section “ODEs of Large Scale Cortico- and Thalamo-Cortical Connectivity,” equations (9)–(12)].

In long-range cortico-cortical connections, the time-delay between NMMs “*i*” and “*j*” was controlled by a distance parameter  $D^{ij}$  that reports the Cartesian distance in centimeters (cm) between the two populations. By setting the conduction velocity of action potentials in the brain to  $c$  [ $c \in (10,100)$  cm/s], the time-delay taken by NMM “*j*” to receive a firing rate is straightforwardly deduced as  $D^{ij}/c$ .

### Thalamo-Cortical Connections

The main connections between the thalamus and neocortex were taken into consideration in the model (see **Figure 2**). As in classical thalamocortical models, TCs receive glutamatergic excitatory postsynaptic potentials (EPSPs) from PCs. In turn, PCs receive excitatory input from TCs. Similarly, TRNs also receive excitatory cortical projections. In terms of GABAergic cortical targets, thalamic projections mainly target PV<sup>+</sup> basket cells (Cruikshank et al., 2007; Yang et al., 2013). In the adult brain, thalamic projections onto SST neurons are present but are much weaker as compared to the projections onto PCs and PV<sup>+</sup> cells (Ji et al., 2016). However, robust thalamocortical activation of non-Martinotti dendritic-targeting GABAergic interneurons has been demonstrated (Tan et al., 2008). Long-range connections from cortical areas and/or thalamic nucleus areas can activate VIP neurons, which in turn inhibit SST neurons, and disinhibit PCs dendrites. Such dendritic disinhibitory circuit has been proposed to gate excitatory inputs targeting pyramidal dendrites (Yang et al., 2016; Williams and Holtmaat, 2019). Note that the thalamic compartment implements FFI, since it induces first an EPSP (PCs activation) followed later on by an IPSP (cortical interneurons activation).

Since only one thalamic population was considered in this model, the TC projections to—and from—PCs were included in the model. Connectivity constants  $K_{Th,P}^n$  and  $K_{P,Th}^n$  allow to adjust the strength of efferent TC projections to PCs of NMM “*n*” and efferent pyramidal projections from the NMM “*n*” to TC, respectively. Projections from PCs of NMM “*n*” onto TRN<sub>1</sub> and TRN<sub>2</sub> were also integrated and adjusted with



**FIGURE 2 |** Large-scale architecture of the model. Illustration of the synaptic projections between cortical modules, and between thalamic and cortical modules. Note the presence of long-range thalamocortical and cortico-cortical feedforward inhibition. For the sake of clarity, long-range connections between cortical NMM<sub>i</sub> and NMM<sub>j</sub> are unidirectional, whereas in the model, PCs of NMM<sub>i</sub> also project on neurons of NMM<sub>j</sub>. The strength of the synaptic input of PC onto distant VIP cells is larger than the input onto the other distant interneurons (SST, Basket).

the connectivity constants  $K_{P,TRN1}^n$  and  $K_{P,TRN2}^n$ , respectively. Likewise, projections from TC to GABAergic interneurons in NMM “ $n$ ,” namely BCs, SSTs and VIPs, were included through the connectivity constants  $K_{Th,BC}^n$ ,  $K_{Th,SST}^n$  and  $K_{Th,VIP}^n$ , respectively. The set of ODEs modeling the thalamo-cortical connections is described in Supplementary section “ODEs of Large Scale Cortico- and Thalamo-Cortical Connectivity” [equations (13)–(19)]. It is noteworthy that time-delays between the thalamus and the cortical NMMs were included as in the cortico-cortical long-range connections.

## Implementation and Parameter Tuning

An important step in NMM approaches consists in tuning the model parameters. Some of these parameters, namely time constants in the three modules, were set close to the “standard values” used in neuronal population models, while other parameters such as connectivity constants were adjusted according to the target EEG activities. Two classical well-known examples of conscious and unconscious states are deep sleep—a.k.a. SWS characterized by delta oscillations (0–4 Hz), and wakefulness (background activity). The objective was to reproduce these manifestations of consciousness modulation in the model.

In **Supplementary Table S1**, we provide physiological interpretation and values of model parameters. Regarding the model output, the signal simulated at the level of PCs in the cortical compartment was chosen as the model output since it corresponds to the sum of PSPs, which is the main contribution to LFPs recorded in the neocortex.

When several cortical NMMs are interconnected, a simple way to handle large-scale connectivity is to arrange the connectivity constants in arrays where line and column indices refer to source and target NMMs, respectively. Based on “Modeling of Large-Scale Cortico-Cortical and Thalamo-Cortical Connectivity” section, there are two categories of connectivity array, “excitatory to excitatory” and “excitatory to inhibitory” arrays. In the first category, we consider the matrix  $K_{EXC}$  whose  $(i, j)$ th element represents the glutamatergic projections from NMM “ $i$ ” onto an excitatory subpopulation in NMM “ $j$ .” Hence, if both of them belongs to the cortical module, the  $(i, j)$ th element would correspond to  $K_{P,P}^{i,j}$ . However, if NMM “ $i$ ” coincides with a thalamic population, the  $(i, j)$ th element would be then equal to  $K_{Th,P}^i$ . In the second category, a similar scheme is considered with the difference that the target subpopulation is always an inhibitory interneuron (BC, SST, VIP or TRNs). Consequently, five arrays are considered, namely,  $K_{BC}$ ,

$K_{SST}$ ,  $K_{VIP}$ ,  $K_{TRN1}$  and  $K_{TRN2}$  (see **Supplementary Figure S1**). Note that when only one thalamic NMM is considered, the two last matrices are reduced to vector arrays. Cortico- and thalamo-cortical time-delays were similarly arranged in matrix  $D$  whose  $(i, j)$ th element coincides with the aforementioned  $D^{ij}$ .

The set of second order stochastic nonlinear ODEs obtained for all synaptic interactions present in the model was numerically solved using a fixed step ( $\Delta t = 1$  ms) 4th-order Runge-Kutta method. The model was implemented using an object-oriented language (Objective-C).

To study the model behavior, we followed a two-step approach. We first implemented a “toy model” with a small number of coupled neural masses. This reduced-complexity model allowed to assess the effects of cortico- and thalamo-cortical connectivity matrices on cortical activities associated with different consciousness states (see “A Local Neural Mass Model of Neocortical Activity,” “A Local Neural Mass Model of Thalamic Activity” and “Modeling of Large-Scale Cortico-Cortical and Thalamo-Cortical Connectivity” sections). After validation in the toy model, all mechanisms were implemented in an extended, more realistic model, where the whole brain was considered.

### Toy Model of Cortical Activity

The toy model was composed of one thalamic population connected to four cortical populations ( $N = 4$ ) that were identically tuned. The default values of NMMs intrinsic parameters are listed in **Supplementary Table S1**. Unless explicitly mentioned, these parameters were kept unchanged for all neural activities generated afterward (SWS and background activity). The objective was to verify, in a simplified model, the hypothesis that cortical activity is modulated from deep sleep to wakefulness by tuning only thalamo- and cortico-cortical connectivity, so that, when thalamocortical connectivity increases, the model goes deeper into sleep. This connectivity mechanism was implemented in all connectivity matrices in the model (see **Supplementary Figure S2**).

### Whole Brain Model: From Cortical Activity to EEG

The pipeline used for the simulation of scalp EEG data is described in **Figure 3**. In order to obtain a “realistic” activity during wakefulness and SWS over the entire neocortex, we considered one thalamic population and 66 cortical populations ( $N + 1 = 67$ ). Each time-course at the output of these populations represented the activity of one macro-region of the anatomical parcellation described in Desikan et al. (2006), in which the activity was assumed to be homogenous. In order to generate 67 time-courses from the coupled NMMs (66 cortical regions plus the thalamus), we used a combination of connectivity arrays (**Figure 3B**). First, a matrix of connection weight  $K_{DTI}$  representing a density of fibers between all pairs of 66 cortical regions was used to set the structural connectivity. This matrix, obtained from DTI, is provided in Hagmann et al. (2008). Second, we considered additional functional horizontal (i.e., cortico-cortical) matrices  $K_{Hx}$ , reproducing the coefficients weights used for both wakefulness and sleep in the toy model, taking also into account the new number of populations. In order to apply the connectivity weights defined earlier in the toy model

only to pairs of NMMs that are structurally connected. Structural and horizontal functional matrices were combined using the Hadamard product. Finally the vertical (i.e., thalamo-cortical) connectivity  $K_{Vx}$  was added to this product to obtain a set of anatomo-functional connectivity matrices  $K_x$  such that:

$$K_x = K_{DTI} \odot K_{Hx} + K_{Vx} \quad (1)$$

with  $x = \{EXC, SST, BC, VIP, TRN1, TRN2\}$ . All connectivity matrices are described in **Supplementary Figure S3**.

Using this combination of connectivity weights, and specific parameters defined in **Supplementary Material** for delta and background activity, we built a spatio-temporal source matrix  $S$  containing the time-varying activities of the thalamus and of all cortical macro-regions.

To reconstruct simulated scalp EEG data, we first solved the forward problem using the Boundary Element Method (BEM, OpenMEEG, Gramfort et al., 2010). To this end, a realistic head model was built in Brainstorm (Tadel et al., 2011) from the segmentation of a template T1 magnetic resonance imaging (MRI; Colin 27 template brain, Holmes et al., 1998) previously obtained using the Freesurfer image analysis suite<sup>1</sup> (Dale et al., 1999; Fischl et al., 1999), as illustrated in **Figure 3A**. The head model consisted in three nested homogeneous mesh surfaces shaping the cortical surface (642 vertices), the skull (642 vertices) and the scalp (1,082 vertices) with conductivity values of  $0.33 \text{ Sm}^{-1}$ ,  $0.0082 \text{ Sm}^{-1}$  and  $0.33 \text{ Sm}^{-1}$ , respectively (Gonçalves et al., 2003). The forward problem was then numerically calculated for each vertex of the source mesh obtained from the segmented white matter/gray matter interface of the same template brain. As a result, the leadfield matrix  $A$  represented the contribution of each dipole of the source mesh at the level of 257 scalp electrode positions (high-density EEG), placed over the scalp according to the geodesic convention (EGI®, Eugene, USA). All leadfield vectors of  $A$  belonging to a common region of Desikan Atlas were added to obtain a simplified  $66 \times 257$  matrix  $G$ . The spatio-temporal matrix  $X$  of simulated EEG data was given by the matrix product:

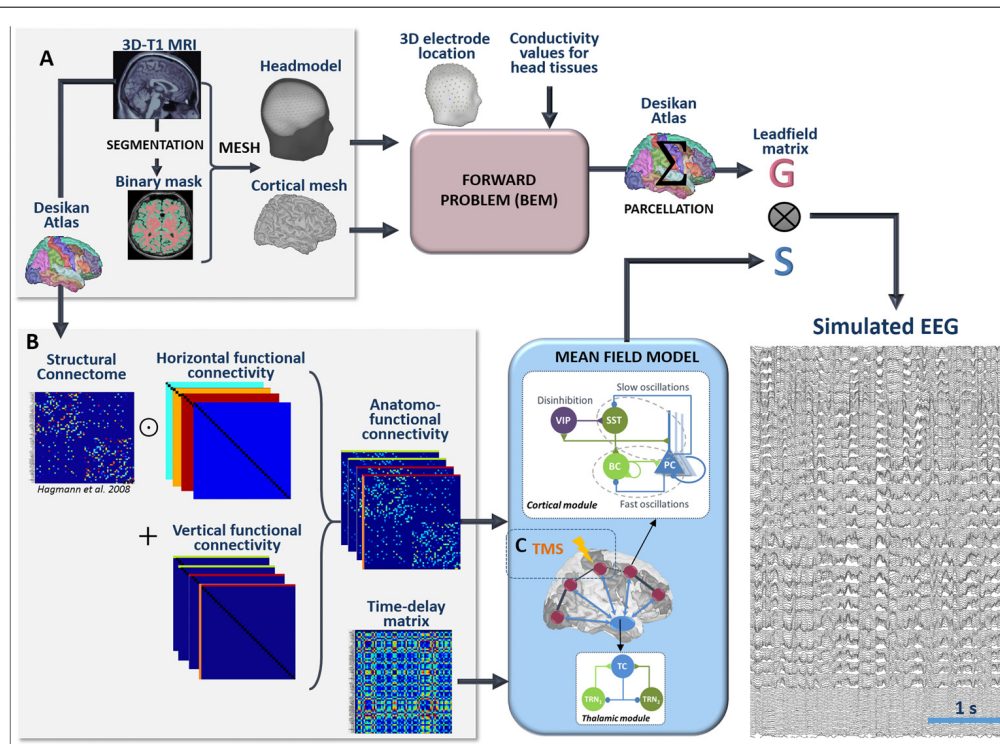
$$X = GS \quad (2)$$

The entire pipeline enabling the simulation of scalp EEG is provided in **Figure 3**. Finally, as detailed in the **Supplementary Material**, a total number of 1,060 ODE's must be solved to run the model. To give an idea of the required computing time, 60 s of simulated EEG could be simulated in 49 s on a 3.5 GHz 6-Core Intel Xeon EG with 64 GB of 1,866 MHz RAM (OsX Mojave) and in 72 s on a standard Intel 2.5 GHz 2-Core Xeon EG with 8 GB RAM (Windows 10). Therefore, the model performed almost real-time simulation on a standard PC.

### Cortical Response to TMS and PCI

In order to simulate TMS-evoked EEG responses that can be compared to those recorded experimentally (Casali et al., 2013), we included the effect of an exogenous, TMS-induced, stimulation in the whole-brain model. The model can simulate

<sup>1</sup><http://surfer.nmr.mgh.harvard.edu/>



**FIGURE 3 |** Full processing pipeline leading to the simulation of scalp EEGs. The pipeline to simulate EEG is a two-step process. **(A)** First, the forward problem is solved at the level of 257 scalp electrodes from a dipole layer constrained to the surface of a cortical mesh (15,000 vertices) midway between the Gray/White matter interface and the cortex surface. The boundary Element Method is used for the calculation within a realistic head model that accounts for the conductivity properties and the geometry of brain, skull and scalp. This step provides a  $257 \times 15,000$  leadfield matrix **A** representing the contribution of each individual cortical dipole at each of the 257 scalp electrodes. In this matrix, leadfield vectors belonging to a common region of the Desikan Atlas are added to obtain a simplified  $66 \times 257$  matrix **G**. Second, the time courses **S** at the whole brain level are obtained in the mean-field model from a set of 66 cortical and one thalamic coupled NMMs. **(B)** Coupling between these 67 NMMs is done using combination of connection weight matrices. Pairs of structurally connected cortical NMMs are first defined from a matrix of connection weights representing a density of fibers between all pairs of 66 cortical regions of the Desikan Atlas. This matrix is provided in Hagmann et al. (2008). Using an element-wise multiplication, this matrix is combined with a set of horizontal (i.e., cortico-cortical) functional connectivity matrices that reproduce the coefficients weights used for wakefulness and sleep in the toy model. Vertical (i.e., thalamo-cortical) connectivity matrices are added to each of these products to obtain connectivity weight matrices that account for anatomical connections as well as cortico-cortical and thalamo-cortical connectivity matrices. Cortico- and thalamo-cortical time-delays were similarly organized in the form of matrices where the elements represent the Cartesian distance between cortical NMMs divided by the mean velocity of traveling for action potentials. **(C)** The mean-field model includes explicitly the contribution of an external stimulus term that represents the effect of TMS. At the output of the pipeline, scalp EEG signals at the level of 257 channels are obtained as the product of leadfield **G** and source time courses **S**.

not only the activity of the 67 pre-defined anatomical regions but also the EEG signals recorded by 257 scalp electrodes. Furthermore, since the model includes, as a connectivity matrix between the 67 regions, a DTI-derived connectivity matrix (Hagmann et al., 2008), it is possible to track the spatio-temporal dynamics of the stimulation-evoked network, i.e., the activated regions, along with the peak latency for each region. In addition, since the model can simulate “wake” and “sleep” states, it provides the opportunity to compare TMS-evoked responses in these two states of consciousness, which have been experimentally recorded in humans. Therefore, the structure of our computational model provides a unique framework to interpret TMS-evoked EEG responses obtained in humans.

In principle, TMS involves a high-intensity current flowing through the stimulation coil, thereby generating a magnetic field penetrating without attenuation through the head. Since

the stimulation pulse is very short ( $\approx 0.1$  ms), the magnetic field gradient  $\text{dB/dt}$  is extremely high ( $> 30,000$  T/s), resulting per Maxwell-Ampere’s law into an electric field at the level of brain tissue. Since the electric field induced in brain tissue is high ( $> \text{V/m}$ , Miranda et al., 2003), this induces neuronal firing, presumably at the bending point of cortical axons, triggering a series of complex activations within the stimulated area (Di Lazzaro and Ziemann, 2013).

We used a simple approach to represent the effect of TMS, consisting in simulating an afferent volley of action potentials (in terms of pulses/s) to the stimulated cortical region. Since the 1 ms time step used to numerically solve the equations of the model was higher than the duration of an actual TMS pulse, the simulated length of this volley of incoming action potential was adapted and fixed to 5 ms. The amplitude of the simulated evoked volley of action potentials was fixed to 1,000 pulses/s and was applied to each cellular subtype of the stimulated region (i.e., PCs



and all types of GABAergic interneurons). For the purpose of this article, we chose to simulate the stimulation of the right motor cortex in both conditions (“wake” and “sleep”), which is known to have a number of anatomical connections that should result in the propagation of the TMS-evoked responses in several cortical structures. The repetition rate of the TMS protocol was fixed to 2 s, and the total simulation duration was a full minute, resulting in a total of 30 TMS-evoked responses available at the source level. The EEG activity at the level of each scalp electrode was then computed from the simulated source activity using our EEG forward problem pipeline, as described previously. From the simulated TMS-evoked responses, the studied outcomes were the anatomical regions activated by TMS, and also their latency from the onset of the TMS pulse.

Finally, simulated EEG responses were used to compute the Perturbational Complexity Index (PCI; Casali et al., 2013). PCI is a measure of TMS-evoked responses complexity, based on the Lempel-Ziv compression algorithm. The basic idea is that, if a sequence of information is complex, then it can be only marginally compressed (low compression rate); and on the opposite, a very simple sequence can be described by a very limited amount of information (high compression rate). To derive PCI, a process similar to the one proposed by Casali et al. (2013) was used:

- The sources activity is stored in a 2D matrix (number of lines equal to the number of channels, number of columns equal to the number of time points).
- A threshold value for the simulated sources activity was set by preserving the highest 20% (proportional threshold) of values once fixing to 0 values below it.
- Each value of the sources activity matrix that is above or equal to the threshold value is set to 1, while all other values are set to 0 (binarization process).
- The Lempel-Ziv algorithm is then applied to the resulting binary matrix.

PCI values were computed in the two different scenarios, corresponding to the “wake” and “sleep” state, respectively.

Overall, our implementation of TMS-evoked responses enables a meaningful comparison with human data, since it results in similar experimentally measurable quantities: impacted anatomical regions, latency of the TMS-evoked response within specific regions, and complexity of the brain-scale response through PCI.

## RESULTS

### Toy Model: The Impact of Cortico- and Thalamo-Cortical Connectivity on the Cortical Rhythms During Sleep and Wakefulness

During the deep sleep state, the thalamocortical connectivity is meant to be strong compared to the cortico-cortical one (see **Supplementary Figure S2**, first column). Our strategy consisted in progressively and simultaneously decreasing and increasing the thalamo- and cortico-cortical connectivity, respectively, in

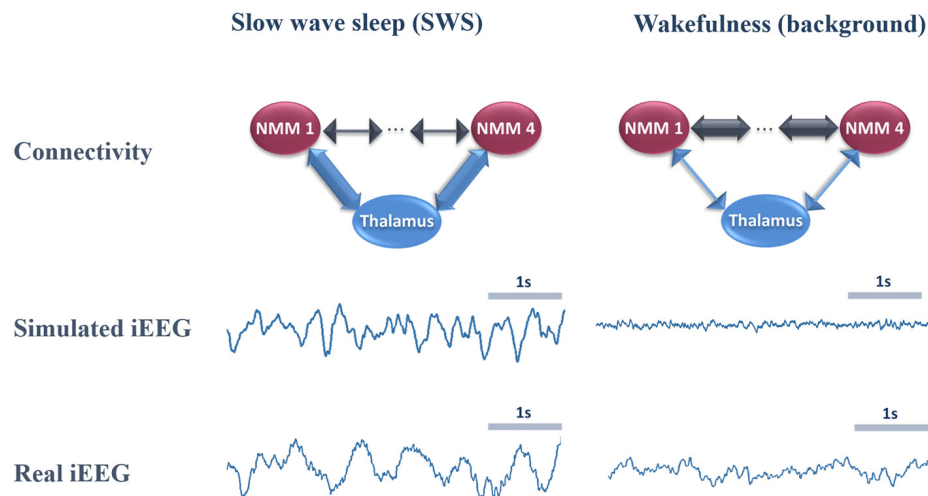
order to switch to wakefulness (see **Supplementary Figure S2**, column 2). Note that this connectivity process was mostly reflected by arrays  $K_{EXC}$  and  $K_{BC}$ , to simulate the strong thalamic projections onto PCs and BCs as reported in “Introduction” section. It is also noteworthy that no time delays were injected in the toy model example.

In **Figure 4**, we provide an example of simulated LFPs a.k.a. intracerebral EEG (iEEG) in comparison with real (human) ones. The left column depicts the scheme of high (and low) thalamo- (and cortico-) cortical connectivity. As depicted, the reinforcement of the thalamocortical loop (TC→PC→TRNs and TC→BC) resulted in the generation of delta oscillations that characterize SWS. In this example, the simulated delta was around 2–3 Hz (slightly faster than real iEEG). Conversely, by reducing the thalamocortical loop (right column), delta waves disappeared and were replaced by background activity, indicative of oscillatory changes observed during the switch from sleep to wake. We emphasize that only the large scale connectivity was tuned while all remaining parameters were kept unchanged, confirming the crucial role of thalamo- and cortico-cortical connectivity in modulating consciousness.

### Whole Brain Model: The Impact of Cortico- and Thalamo-Cortical Connectivity on Scalp EEG Rhythms During Sleep and Wakefulness

The morphology of simulated intracerebral signals was not modified when connectivity matrices were adapted to a larger number of NMMs in order to account for whole-brain activity. In the case of low thalamo-cortical connectivity (wakefulness condition), background activity was similar to signals obtained in the toy model and resembled real intracerebral background activity recorded during wakefulness in humans. Similarly, by increasing thalamo-cortical connectivity, the whole-brain model generated delta activity at a mean frequency of 3.8 Hz that was consistent with the morphology and spectral content of delta activity obtained both from the toy model and from real intracerebral recordings during SWS.

Signals obtained with the whole-brain model in both conditions of wakefulness and SWS were used in the forward calculation to generate simulated scalp EEG data at the level of 257 electrodes (**Figure 5**). In the low-thalamocortical connectivity condition, simulated scalp EEG (**Figure 5A**) resembled scalp EEG background activity as recorded in humans during wakefulness (**Figure 5C**). The spectral analysis disclosed similar sub-band distribution in the simulated vs. real case, although simulated signals contained more beta frequency than real background activity. In the high-thalamo-cortical connectivity condition, simulated scalp EEG (**Figure 5B**) was comparable to scalp EEG signals recorded in humans during SWS (**Figure 5D**). In the simulation case, the peak frequency was 3.8 Hz, thus slightly higher than in the real case (2 Hz). Topographical voltage maps at the peak of delta waves showed analogous distribution over the vertex, the activity in the simulated case being slightly more posterior than on the real case example.



**FIGURE 4 |** Comparison of real and simulated intracerebral EEG (iEEG; toy model,  $N = 4$ ). *Left column:* in the condition of high thalamo-cortical connectivity (i.e., low cortico-cortical connectivity), signals generated by the mean-field model are characterized by delta waves ( $\sim 4$  Hz). These simulated signals are similar (although slightly faster) to delta waves recorded by iEEG during slow wave sleep (SWS) in non-epileptic cortical regions of one patient undergoing an invasive EEG exploration. *Right column:* in the condition of low thalamo-cortical connectivity, signals generated by the mean-field model are similar to background activity recorded by iEEG in real conditions during wakefulness. Note that these iEEG recordings are performed in patients who are candidates to epilepsy surgery. For the sake of this study, only iEEG signals that do not show epileptic activity were retained.

## Bridging Brain Circuits, TMS-Evoked EEG Responses and Complexity Metrics

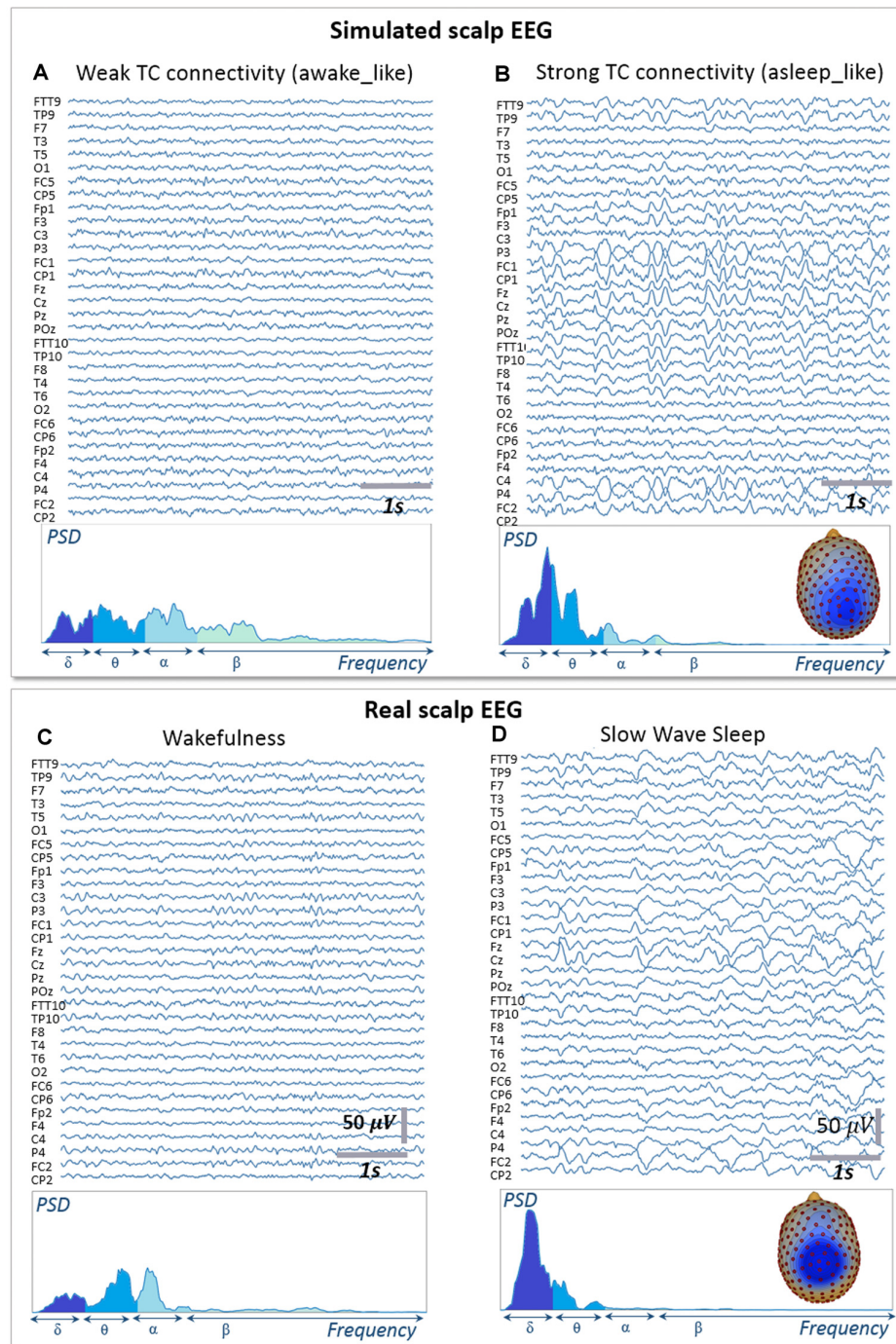
In **Figure 6**, we describe the process used to estimate the complexity of TMS-evoked EEG responses within our brain-scale model and in humans (Casali et al., 2013). As depicted, the simulated TMS-evoked EEG response (**Figure 6B**) was very similar to the human response (**Figure 6A**), not only in terms of length (250 and 350 ms, respectively) but also in terms of regions distant from the right motor cortex (the stimulated area) that are activated post-stimulation. Indeed, in the simulated and experimental data, a first activation of the right pre-central gyrus occurred within 15 ms from the TMS pulse, followed by activity evoked notably in the right precuneus within 40 ms and a common propagation in the contralateral left precuneus at about 60 ms from the stimulus. In both simulated and experimental TMS-evoked responses, activity was evoked in the right frontal lobe within 110–120 ms, with activation of the right superior parietal cortex within 175 ms (human data) and 150 ms (model data). The most important difference is that the human TMS-evoked response was notably longer as compared to the model.

We then compared the simulated and experimental TMS-evoked responses in wakefulness and in sleep, along with their PCI value, which is presented in **Figure 7**. In the two cases, the right motor area is stimulated. In the case of wakefulness, as also illustrated in **Figure 6**, there was a satisfactory agreement in the duration and global shape of the response, and also in terms of the sequence of activated brain regions. More precisely, by comparing **Figure 7A** (upper) and **Figure 7B** (upper), we can conclude that the tracking of the propagation of the TMS-evoked activity revealed that this propagation occurred,

as expected, along documented inter-regional connections from the DTI-derived connectivity matrix used in the model. PCI values obtained were also extremely similar (0.52 and 0.51 for the model and for humans, respectively). Conversely, in the “Sleep” condition, the time course of the TMS-evoked response was significantly shorter (less than 200 ms), which is also comparable to TMS-EEG human recordings. In addition, another striking similarity with human data is that the TMS-evoked activity remained confined to the stimulated area, i.e., the right motor area. The PCI value in this condition was 0.19, which is very similar to the value obtained in humans during sleep (0.23, see **Figure 7A**, lower panel), notably lower than the “Wakefulness” condition, which is also consistent with human data (Casali et al., 2013). Therefore, despite using the exact same TMS pulse characteristics within the two conditions (“Wakefulness” and “Sleep”), the simulated response was drastically different within the model: in wakefulness, the TMS-evoked response resulted in a complex sequence of successive activations within distant, anatomically connected areas in ipsilateral and contralateral regions; while during sleep the TMS-evoked activity remained confined to the stimulation site, even when the anatomical connections were present as in the “sleep” condition.

## DISCUSSION AND PERSPECTIVE

In this article, we have developed the first brain-scale computational model that can reproduce neuronal activity patterns associated with various conscious states, while accounting for key microcircuits at the cellular type scale. A major asset of the model is its strong link with recent neurophysiological and neuroanatomical data: the main cellular

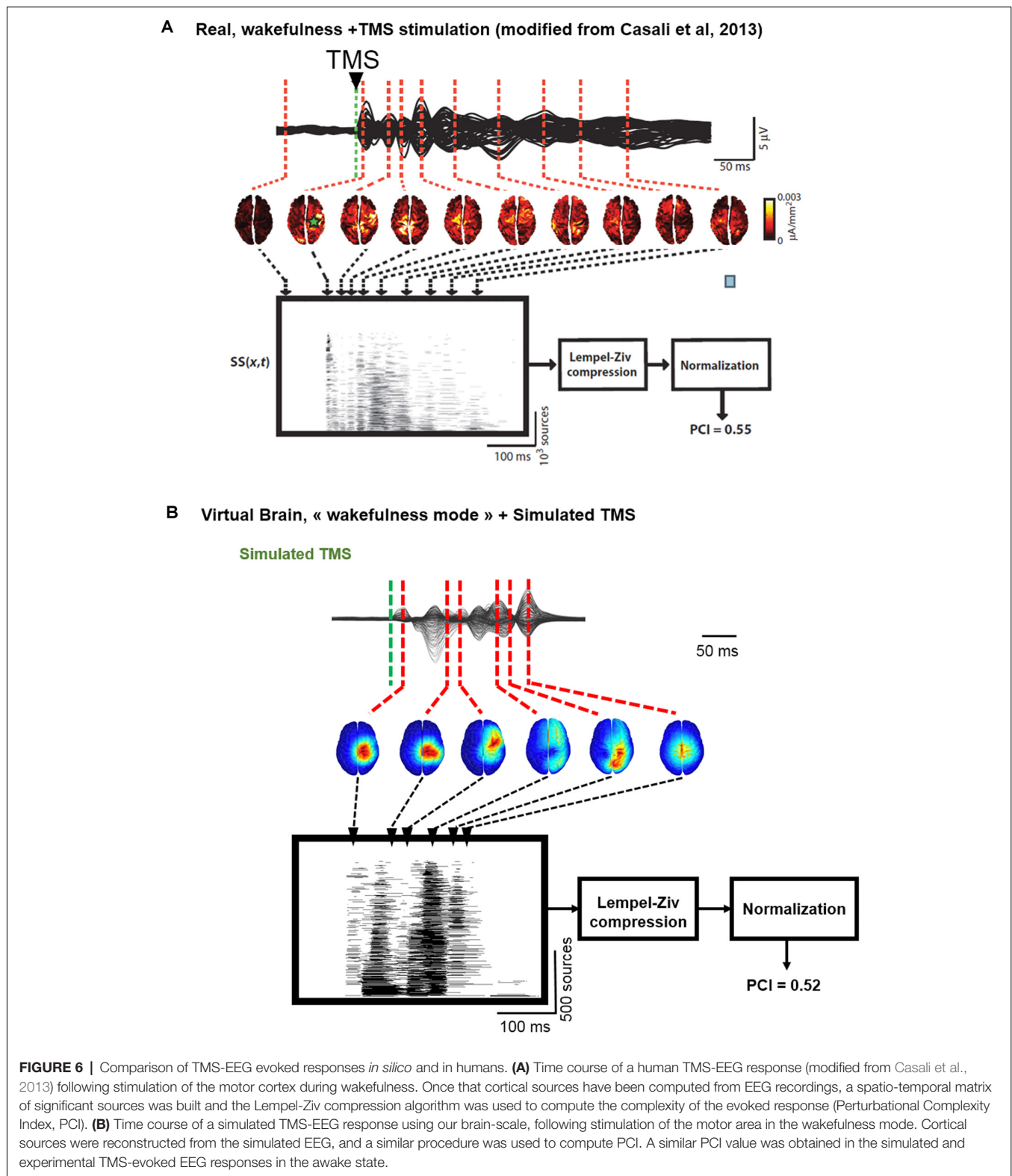


**FIGURE 5 |** Simulated vs. Real EEG during wakefulness and SWS. Signals simulated with the whole-brain model using weak thalamo-cortical connectivity parameters **(A)** display background activity. The morphology and spectral content of these simulated signals are similar to scalp EEG recorded in a human subject during wakefulness in humans **(C)**, except for a higher power spectral density in the beta sub-band. Signals simulated with the whole-brain model using strong thalamo-cortical connectivity parameters **(B)** display delta waves similar to the activity recorded in real condition during SWS in humans **(D)**. The spectral content of signals as well as the topographical voltage distribution at the peak of delta waves were similar in the simulated and real conditions.

types are included (PCs and different types of interneurons), along with their recently elucidated connections that underlie the selective disinhibition of distant neural populations (through

VIP to SST projections), realistic synaptic kinetics, large-scale structural connectivity obtained in humans through DTI and propagation delays between regions based on their spatial

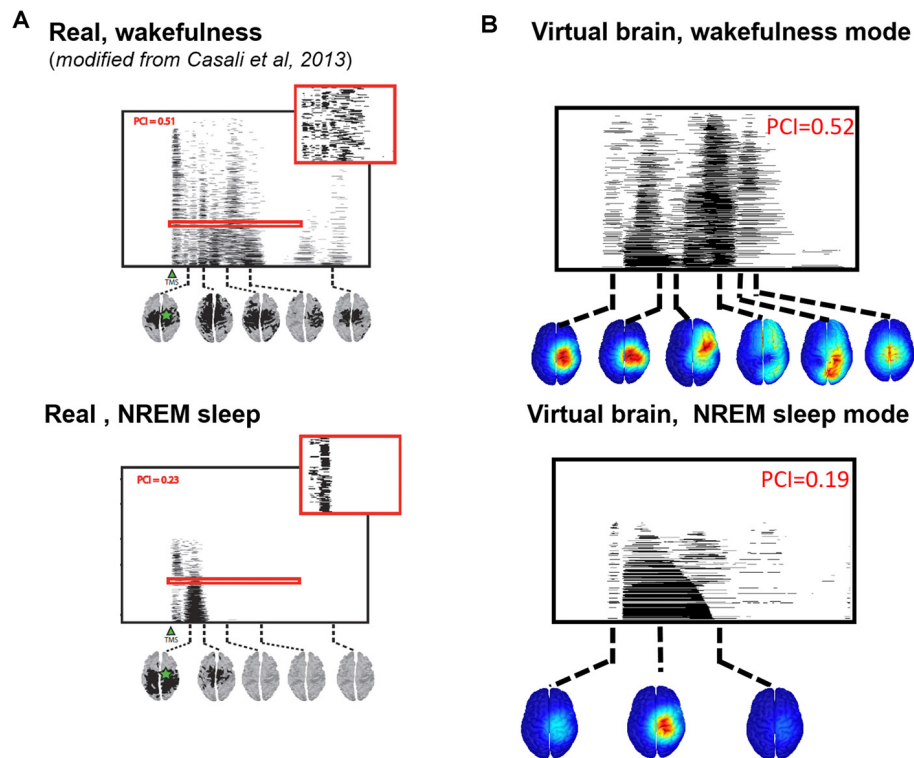




distance. Furthermore, since the model features detailed micro-circuits, DTI-derived brain connectivity matrix and large-scale macro-circuits results also provided a bottom-up description of TMS-evoked responses in humans.

The model offers novel, key insights into the maintenance of the neuronal activity associated with conscious states. Interestingly, this model including a variety of cellular subtypes with accurate synaptic kinetics provides realistic





**FIGURE 7 |** Comparison of TMS-evoked EEG responses in wakefulness and sleep. **(A)** TMS-evoked EEG responses obtained through TMS of the motor cortex in humans (upper panel, during wakefulness; lower panel, during sleep) with associated PCI values. **(B)** TMS-evoked EEG responses obtained through simulated TMS of the motor region (upper panel, in the wakefulness mode; lower panel, during the sleep mode) with associated PCI values.

electrophysiological signals both at the level of cortical sources and of the EEG. First, it explains how thalamo-cortical (vertical) connectivity is critically involved in the gating of cortico-cortical (horizontal) information propagation. If thalamo-cortical activity is indeed rhythmically patterned, the communication between cortical areas is disrupted due to the resulting rhythmic inhibition. This result is therefore in line with the “connectivity breakdown” observed during sleep (Esser et al., 2009; Ferrarelli et al., 2010; Casali et al., 2013).

Crucially, the various conscious states simulated with the model can be tuned by modulating only the thalamo-cortical input, which supports our hypothesis that this represents the crucial control parameter for consciousness, i.e., the dimension of wakefulness. It can be seen essentially as the possibility for information processing to take place between spatially distant cortical areas. Since the propagation of activity is impaired at the cortical level during sleep (absence of wakefulness), no cortical processing can take place, which explains the absence of consciousness. Second, regarding cortico-cortical activity (horizontal connectivity), since the model reproduces with a satisfactory qualitative and quantitative agreement the TMS-evoked EEG responses observed in awake humans (Casali et al., 2013), this suggests that: (1) our modeling hypotheses and choices appear sufficient to capture the essence of TMS-EEG responses; (2) TMS-EEG responses are mainly driven by the underlying structural connectome, which is in

line with recent research pointing at the tight links between structural and functional connectivity (Avena-Koenigsberger et al., 2017); and (3) this modeling approach could be used to assist in the interpretation of TMS-evoked EEG responses in DOC patients, since our model links explicitly the underlying connectivity with the observed TMS-evoked response. Finally, the model validates that, at the brain scale, the disinhibitory disinhibition of distant PCs through the activation of VIP neurons is indeed an effective mechanism enabling the transmission of activity between a few cortical regions. The model therefore confirms the role of a cellular-scale micro-circuit that regulates brain-scale propagation of activity within the cortex.

Among the possibilities to improve the realism and predictive power of our brain-scale model, the most immediate would be the use of structural connectivity matrices averaged among a large number of healthy participants, such as those from the Human Connectome Project<sup>2</sup>. Furthermore, no regional specificities were accounted for between the 66 cortical regions included within the model, since we used standard parameter values for the synaptic gains (e.g., A, B and G) and the same within-population connectivity parameters. By doing so, we have assumed that the large-scale anatomic structure of brain connectivity and the cellular-scale micro-circuits included

<sup>2</sup><http://www.humanconnectomeproject.org>

are the main factors explaining the resulting simulated EEG signals. Another limitation of COALIA under its current form is the simplification of the physiological processes involved in the transition from wakefulness to sleep. The brainstem is indeed involved in the modulation of thalamic activity, and the regulation of cortico-cortical connectivity (Crunelli et al., 2015). In the model, the brainstem input onto the thalamus is not explicitly included. Instead, the modulation of thalamic inhibition between the sleep/wake mode was simulated by changing the excitatory input onto the RT-interneurons or TC cells.

It should be mentioned that in the case of deep sleep simulated signals, the median frequency of delta activity was higher in the model (2–3 Hz) as compared to human data (~1 Hz). This discrepancy is explained by the fact that model does not implement the mechanisms underlying slow oscillations (~1 Hz) generated in cortical and thalamic networks, among which: (i) the sequence of depolarizing periods followed by silent periods (up to 2 s) during up-and-down states as well as; and (ii) GABA<sub>B</sub>-mediated pre-synaptic slow inhibition that also appears to play some role (see review in Neske, 2015). Other limitations likely explain the moderate discrepancies between the simulated and experimental TMS-EEG responses in terms of latencies and localizations, such as the lack of asymmetry in connectivity weights (i.e., all connections were assumed bidirectional and identical).

Let us emphasize that the results presented in the present article are only a first step towards understanding the transition between wake and sleep and that COALIA provides an appropriate framework to achieve that objective. The future prospects regarding our brain-scale EEG model are numerous: in terms of consciousness studies, the model could be used to study the mechanisms underlying the so-called “slow-wave activity saturation,” delta-band activity that appears when the blood concentration of anesthetics is increased (Ni Mhuircheartaigh et al., 2013), and that constitutes a solid marker of the conscious state. Furthermore, this study improves our understanding of active probing paradigms of brain circuits in DOCs, such as the PCI and paves the way toward the design of optimized

stimulation-based metrics to measure consciousness. The model indeed allows to test *in silico* novel neuromodulation protocols based on TMS, transcranial direct current stimulation (tDCS) and transcranial alternating current stimulation (tACS), aiming at quantifying the level of residual consciousness in DOC patients. Beyond applications for consciousness, the model could be exploited to understand the detailed dynamics of TMS-EEG responses and their underlying mechanisms, or to shed light on the mechanisms underlying the generation and propagation of epileptiform activity.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

## AUTHOR CONTRIBUTIONS

SB: model implementation, simulations, data analysis, manuscript write-up. JM: model design, simulations, data analysis, manuscript write-up. IM: simulations, data analysis, manuscript write-up. FW and PB: model design, data analysis, manuscript write-up.

## FUNDING

This work has been fully funded by the LUMINOUS Project. This project has received funding from the European Union's Horizon 2020 research and innovation program H2020-FETOPEN-2014-2015-RIA under agreement No. 686764. This study was supported in part by the Institut National de la Santé et de la Recherche Médicale (INSERM, France).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnsys.2019.00059/full#supplementary-material>.

## REFERENCES

- Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J., and Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. *Nature* 490, 226–231. doi: 10.1038/nature11526
- Avena-Koenigsberger, A., Misisic, B., and Sporns, O. (2017). Communication dynamics in complex brain networks. *Nat. Rev. Neurosci.* 19, 17–33. doi: 10.1038/nrn.2017.149
- Ayzenshtat, I., Karnani, M. M., Jackson, J., and Yuste, R. (2016). Cortical control of spatial resolution by VIP<sup>+</sup> interneurons. *J. Neurosci.* 36, 11498–11509. doi: 10.1523/JNEUROSCI.1920-16.2016
- Bhattacharya, B. S., Coyle, D., and Maguire, L. P. (2011). A thalamo-cortico-thalamic neural mass model to study  $\alpha$  rhythms in Alzheimer's disease. *Neural Netw.* 24, 631–645. doi: 10.1016/j.neunet.2011.02.009
- Breakspear, M. (2017). Dynamic models of large-scale brain activity. *Nat. Neurosci.* 20, 340–352. doi: 10.1038/nn.4497
- Buzsáki, G., and Wang, X.-J. (2012). Mechanisms of  $\gamma$  oscillations. *Annu. Rev. Neurosci.* 35, 203–225. doi: 10.1146/annurev-neuro-062111-150444
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* 5:198ra105. doi: 10.1126/scitranslmed.3006294
- Chen, G., Zhang, Y., Li, X., Zhao, X., Ye, Q., Lin, Y., et al. (2017). Distinct inhibitory circuits orchestrate cortical  $\beta$  and  $\gamma$  band oscillations. *Neuron* 96, 1403.e6–1418.e6. doi: 10.1016/j.neuron.2017.11.033
- Cona, F., Lacanna, M., and Ursino, M. (2014). A thalamo-cortical neural mass model for the simulation of brain rhythms during sleep. *J. Comput. Neurosci.* 37, 125–148. doi: 10.1007/s10827-013-0493-1
- Crone, J. S., Schurz, M., Höller, Y., Bergmann, J., Monti, M., Schmid, E., et al. (2015). Impaired consciousness is linked to changes in effective connectivity of the posterior cingulate cortex within the default mode network. *Neuroimage* 110, 101–109. doi: 10.1016/j.neuroimage.2015.01.037
- Cruikshank, S. J., Lewis, T. J., and Connors, B. W. (2007). Synaptic basis for intense thalamocortical activation of feedforward inhibitory cells in neocortex. *Nat. Neurosci.* 10, 462–468. doi: 10.1038/nn1861

- Crunelli, V., David, F., Lorincz, M. L., and Hughes, S. W. (2015). The thalamocortical network as a single slow wave-generating unit. *Curr. Opin. Neurobiol.* 31, 72–80. doi: 10.1016/j.conb.2014.09.001
- Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194. doi: 10.1006/nimg.1998.0395
- David, O., and Friston, K. J. (2003). A neural mass model for MEG/EEG: coupling and neuronal dynamics. *Neuroimage* 20, 1743–1755. doi: 10.1016/j.neuroimage.2003.07.015
- Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018
- Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.* 4, 752–758. doi: 10.1038/89551
- Dehaene, S., Kerszberg, M., and Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. U S A* 95, 14529–14534. doi: 10.1073/pnas.95.24.14529
- Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Di Lazzaro, V., and Ziemann, U. (2013). The contribution of transcranial magnetic stimulation in the functional evaluation of microcircuits in human motor cortex. *Front. Neural Circuits* 7:18. doi: 10.3389/fncir.2013.00018
- Dipoppa, M., Ranson, A., Krumin, M., Pachitariu, M., Carandini, M., and Harris, K. D. (2018). Vision and locomotion shape the interactions between neuron types in mouse visual cortex. *Neuron* 98, 602.e8–615.e8. doi: 10.1016/j.neuron.2018.03.037
- Esser, S. K., Hill, S., and Tononi, G. (2009). Breakdown of effective connectivity during slow wave sleep: investigating the mechanism underlying a cortical gate using large-scale modeling. *J. Neurophysiol.* 102, 2096–2111. doi: 10.1152/jn.00059.2009
- Ferrarelli, F., Massimini, M., Sarasso, S., Casali, A., Riedner, B. A., Angelini, G., et al. (2010). Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proc. Natl. Acad. Sci. U S A* 107, 2681–2686. doi: 10.1073/pnas.0913008107
- Fino, E., and Yuste, R. (2011). Dense inhibitory connectivity in neocortex. *Neuron* 69, 1188–1203. doi: 10.1016/j.neuron.2011.02.025
- Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207. doi: 10.1006/nimg.1998.0396
- Freeman, W. J. (1978). Models of the dynamics of neural populations. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 34, 9–18.
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011
- Fries, P. (2009). Neuronal  $\gamma$ -band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* 32, 209–224. doi: 10.1146/annurev.neuro.051508.135603
- Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008
- Fu, Y., Tucciarone, J. M., Espinosa, J. S., Sheng, N., Darcy, D. P., Nicoll, R. A., et al. (2014). A cortical circuit for gain control by behavioral state. *Cell* 156, 1139–1152. doi: 10.1016/j.cell.2014.01.050
- Funk, C. M., Peelman, K., Bellesi, M., Marshall, W., Cirelli, C., and Tononi, G. (2017). Role of somatostatin-positive cortical interneurons in the generation of sleep slow waves. *J. Neurosci.* 37, 9132–9148. doi: 10.1523/Jneurosci.1303-17.2017
- Gentet, L. J., Kremer, Y., Taniguchi, H., Huang, Z. J., Staiger, J. F., and Petersen, C. C. H. (2012). Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex. *Nat. Neurosci.* 15, 607–612. doi: 10.1038/nn.3051
- Gidon, A., and Segev, I. (2012). Principles governing the operation of synaptic inhibition in dendrites. *Neuron* 75, 330–341. doi: 10.1016/j.neuron.2012.05.015
- Gómez, F., Phillips, C., Soddu, A., Boly, M., Boveroux, P., Vanhaudenhuyse, A., et al. (2013). Changes in effective connectivity by propofol sedation. *PLoS One* 8:e71370. doi: 10.1371/journal.pone.0071370
- Gonçalves, S. I., de Munck, J. C., Verbunt, J. P., Bijma, F., Heethaar, R. M., and Lopes da Silva, F. (2003). *In vivo* measurement of the brain and skull resistivities using an EIT-based method and realistic models for the head. *IEEE Trans. Biomed. Eng.* 50, 754–767. doi: 10.1109/tbme.2003.812164
- Gramfort, A., Papadopoulos, T., Olivi, E., and Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomed. Eng. Online* 9:45. doi: 10.1186/1475-925x-9-45
- Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6:e159. doi: 10.1371/journal.pbio.0060159
- Hale, J. R., White, T. P., Mayhew, S. D., Wilson, R. S., Rollings, D. T., Khalsa, S., et al. (2016). Altered thalamocortical and intra-thalamic functional connectivity during light sleep compared with wake. *Neuroimage* 125, 657–667. doi: 10.1016/j.neuroimage.2015.10.041
- Harris, A. Z., and Gordon, J. A. (2015). Long-range neural synchrony in behavior. *Annu. Rev. Neurosci.* 38, 171–194. doi: 10.1146/annurev-neuro-071714-034111
- Harris, K. D., and Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature* 503, 51–58. doi: 10.1038/nature12654
- Hill, S., and Tononi, G. (2005). Modeling sleep and wakefulness in the thalamocortical system. *J. Neurophysiol.* 93, 1671–1698. doi: 10.1152/jn.00915.2004
- Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., and Evans, A. C. (1998). Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr.* 22, 324–333. doi: 10.1097/00004728-199803000-00032
- Hudetz, A. G. (2012). General anesthesia and human brain connectivity. *Brain Connect.* 2, 291–302. doi: 10.1089/brain.2012.0107
- Jansen, B. H., and Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol. Cybern.* 73, 357–366. doi: 10.1007/bf00199471
- Ji, X.-Y., Zingg, B., Mesik, L., Xiao, Z., Zhang, L. I., and Tao, H. W. (2016). Thalamocortical innervation pattern in mouse auditory and visual cortex: laminar and cell-type specificity. *Cereb. Cortex* 26, 2612–2625. doi: 10.1093/cercor/bhv099
- Jin, S.-H., and Chung, C. K. (2012). Messages from the brain connectivity regarding neural correlates of consciousness. *Exp. Neurobiol.* 21, 113–122. doi: 10.5607/en.2012.21.3.113
- Kapfer, C., Glickfeld, L. L., Atallah, B. V., and Scanziani, M. (2007). Supralinear increase of recurrent inhibition during sparse activity in the somatosensory cortex. *Nat. Neurosci.* 10, 743–753. doi: 10.1038/nn1909
- Karnani, M. M., Agetsuma, M., and Yuste, R. (2014). A blanket of inhibition: functional inferences from dense inhibitory connectivity. *Curr. Opin. Neurobiol.* 26, 96–102. doi: 10.1016/j.conb.2013.12.015
- Kim, S.-P., Hwang, E., Kang, J.-H., Kim, S., and Choi, J. H. (2012). Changes in the thalamocortical connectivity during anesthesia-induced transitions in consciousness. *Neuroreport* 23, 294–298. doi: 10.1097/wnr.0b013e3283509ba0
- Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.61
- Laureys, S. (2004). Functional neuroimaging in the vegetative state. *NeuroRehabilitation* 19, 335–341.
- Laureys, S., Boly, M., and Maquet, P. (2006). Tracking the recovery of consciousness from coma. *J. Clin. Invest.* 116, 1823–1825. doi: 10.1172/jci29172
- Laureys, S., Goldman, S., Phillips, C., Van Bogaert, P., Aerts, J., Luxen, A., et al. (1999). Impaired effective cortical connectivity in vegetative state: preliminary investigation using PET. *Neuroimage* 9, 377–382. doi: 10.1006/nimg.1998.0414
- Lee, S., Kruglikov, I., Huang, Z. J., Fishell, G., and Rudy, B. (2013). A disinhibitory circuit mediates motor integration in the somatosensory cortex. *Nat. Neurosci.* 16, 1662–1670. doi: 10.1038/nn.3544
- Lee, J. H., and Mihalas, S. (2017). Visual processing mode switching regulated by VIP cells. *Sci. Rep.* 7:1843. doi: 10.1038/s41598-017-01830-0

- Lewis, D. A., Curley, A. A., Glausier, J. R., and Volk, D. W. (2012). Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. *Trends Neurosci.* 35, 57–67. doi: 10.1016/j.tins.2011.10.004
- Lewis, L. D., Voigts, J., Flores, F. J., Schmitt, L. I., Wilson, M. A., Halassa, M. M., et al. (2015). Thalamic reticular nucleus induces fast and local modulation of arousal state. *Elife* 4:e08760. doi: 10.7554/elifelife.08760
- Liu, F., Wang, J., Liu, C., Li, H., Deng, B., Fietkiewicz, C., et al. (2016). A neural mass model of basal ganglia nuclei simulates pathological  $\beta$  rhythm in Parkinson's disease. *Chaos* 26:123113. doi: 10.1063/1.4972200
- Liu, C., Zhu, Y., Liu, F., Wang, J., Li, H., Deng, B., et al. (2017). Neural mass models describing possible origin of the excessive  $\beta$  oscillations correlated with Parkinsonian state. *Neural Netw.* 88, 65–73. doi: 10.1016/j.neunet.2017.01.011
- Llinás, R., Ribary, U., Contreras, D., and Pedroarena, C. (1998). The neuronal basis for consciousness. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1841–1849. doi: 10.1098/rstb.1998.0336
- Lopes da Silva, F. H., van Rotterdam, A., Barts, P., van Heusden, E., and Burr, W. (1976). Models of neuronal populations: the basic mechanisms of rhythmicity. *Prog. Brain Res.* 45, 281–308. doi: 10.1016/s0079-6123(08)60995-4
- Lovett-Barron, M., Turi, G. F., Kaifosh, P., Lee, P. H., Bolze, F., X.-Sun, H., et al. (2012). Regulation of neuronal input transformations by tunable dendritic inhibition. *Nat. Neurosci.* 15, 423–430. doi: 10.1038/nn.3024
- Ma, W.-P., Liu, B.-H., Li, Y.-T., Huang, Z. J., Zhang, L. I., and Tao, H. W. (2010). Visual representations by cortical somatostatin inhibitory neurons—selective but with weak and delayed responses. *J. Neurosci.* 30, 14371–14379. doi: 10.1523/JNEUROSCI.3248-10.2010
- Maex, R., and De Schutter, E. (2007). Mechanism of spontaneous and self-sustained oscillations in networks connected through axo-axonal gap junctions. *Eur. J. Neurosci.* 25, 3347–3358. doi: 10.1111/j.1460-9568.2007.05593.x
- Massimini, M., Ferrarelli, F., Murphy, M. J., Huber, R., Riedner, B. A., Casarotto, S., et al. (2010). Cortical reactivity and effective connectivity during REM sleep in humans. *Cogn. Neurosci.* 1, 176–183. doi: 10.1080/17588921003731578
- Mina, F., Benquet, P., Pasnicu, A., Biraben, A., and Wendling, F. (2013). Modulation of epileptic activity by deep brain stimulation: a model-based study of frequency-dependent effects. *Front. Comput. Neurosci.* 7:94. doi: 10.3389/fncom.2013.00094
- Miranda, P. C., Hallett, M., and Basser, P. J. (2003). The electric field induced in the brain by magnetic stimulation: a 3-D finite-element analysis of the effect of tissue heterogeneity and anisotropy. *IEEE Trans. Biomed. Eng.* 50, 1074–1085. doi: 10.1109/tbme.2003.816079
- Modolo, J., Hassan, M., Ruffini, G., and Legros, A. (2018). Probing the circuits of conscious perception with magnetophosphenes. *Preprint*
- Molae-Ardekani, B., Benquet, P., Bartolomei, F., and Wendling, F. (2010). Computational modeling of high-frequency oscillations at the onset of neocortical partial seizures: from 'altered structure' to 'dysfunction'. *Neuroimage* 52, 1109–1122. doi: 10.1016/j.neuroimage.2009.12.049
- Nakajima, M., and Halassa, M. M. (2017). Thalamic control of functional cortical connectivity. *Curr. Opin. Neurobiol.* 44, 127–131. doi: 10.1016/j.conb.2017.04.001
- Neske, G. T. (2015). The slow oscillation in cortical and thalamic networks: mechanisms and functions. *Front. Neural Circuits* 9:88. doi: 10.3389/fncir.2015.00088
- Ni Mhuircheartaigh, R., Warnaby, C., Rogers, R., Jbabdi, S., and Tracey, I. (2013). Slow-wave activity saturation and thalamocortical isolation during propofol anesthesia in humans. *Sci. Transl. Med.* 5:208ra148. doi: 10.1126/scitranslmed.3006007
- Nunez, P. L. (1974). The brain wave equation: a model for the EEG. *Math. Biosci.* 21, 279–297. doi: 10.1016/0025-5564(74)90020-0
- Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., and Pickard, J. D. (2006). Detecting awareness in the vegetative state. *Science* 313:1402. doi: 10.1126/science.1130197
- Paulus, W., and Rothwell, J. C. (2016). Membrane resistance and shunting inhibition: where biophysics meets state-dependent human neurophysiology. *J. Physiol.* 594, 2719–2728. doi: 10.1113/jp271452
- Pfeffer, C. K. (2014). Inhibitory neurons: vip cells hit the brake on inhibition. *Curr. Biol.* 24, R18–R20. doi: 10.1016/j.cub.2013.11.001
- Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., and Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature* 503, 521–524. doi: 10.1038/nature12676
- Povyshva, N. V., Zaitsev, A. V., Rotaru, D. C., Gonzalez-Burgos, G., Lewis, D. A., and Krimer, L. S. (2008). Parvalbumin-positive basket interneurons in monkey and rat prefrontal cortex. *J. Neurophysiol.* 100, 2348–2360. doi: 10.1152/jn.90396.2008
- Roberts, J. A., and Robinson, P. A. (2012). Corticothalamic dynamics: structure of parameter space, spectra, instabilities, and reduced model. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 85:011910. doi: 10.1103/physreve.85.011910
- Rudy, B., Fishell, G., Lee, S., and Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* 71, 45–61. doi: 10.1002/dneu.20853
- Ruffini, G. (2017). An algorithmic information theory of consciousness. *Neurosci. Conscious.* 2017:nix019. doi: 10.1093/nc/nix019
- Saalmann, Y. B. (2014). Intralaminar and medial thalamic influence on cortical synchrony, information transmission and cognition. *Front. Syst. Neurosci.* 8:83. doi: 10.3389/fnsys.2014.00083
- Schoenberg, P. L. A., Ruf, A., Churchill, J., Brown, D. P., and Brewer, J. A. (2018). Mapping complex mind states: EEG neural substrates of meditative unified compassionate awareness. *Conscious. Cogn.* 57, 41–53. doi: 10.1016/j.concog.2017.11.003
- Sen Bhattacharya, B., Cakir, Y., Serap-Sengor, N., Maguire, L., and Coyle, D. (2013). Model-based bifurcation and power spectral analyses of thalamocortical  $\alpha$  rhythm slowing in Alzheimer's disease. *Neurocomputing* 115, 11–22. doi: 10.1016/j.neucom.2012.10.023
- Silberberg, G., and Markram, H. (2007). Disynaptic inhibition between neocortical pyramidal cells mediated by martinotti cells. *Neuron* 53, 735–746. doi: 10.1016/j.neuron.2007.02.012
- Sohn, J., Okamoto, S., Kataoka, N., Kaneko, T., Nakamura, K., and Hioki, H. (2016). Differential inputs to the perisomatic and distal-dendritic compartments of VIP-positive neurons in layer 2/3 of the mouse barrel cortex. *Front. Neuroanat.* 10:124. doi: 10.3389/fnana.2016.00124
- Sotero, R. C., Trujillo-Barreto, N. J., Iturria-Medina, Y., Carbonell, F., and Jimenez, J. C. (2007). Realistically coupled neural mass models can generate EEG rhythms. *Neural Comput.* 19, 478–512. doi: 10.1162/neco.2007.19.2.478
- Suffczynski, P., Kalitzin, S., and Lopes Da Silva, F. H. (2004). Dynamics of non-convulsive epileptic phenomena modeled by a bistable neuronal network. *Neuroscience* 126, 467–484. doi: 10.1016/j.neuroscience.2004.03.014
- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 2011:879716. doi: 10.1155/2011/879716
- Tan, Z., Hu, H., Huang, Z. J., and Agmon, A. (2008). Robust but delayed thalamocortical activation of dendritic-targeting inhibitory interneurons. *Proc. Natl. Acad. Sci. U S A* 105, 2187–2192. doi: 10.1073/pnas.0710628105
- Timofeev, I., and Steriade, M. (1996). Low-frequency rhythms in the thalamus of intact cortex and decorticated cats. *J. Neurophysiol.* 76, 4152–4168. doi: 10.1152/jn.1996.76.6.4152
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.* 5:42. doi: 10.1186/1471-2202-5-42
- Tononi, G. (2012). Integrated information theory of consciousness: an updated account. *Arch. Ital. Biol.* 150, 56–90. doi: 10.4449/aib.v149i5.1388
- Tononi, G., and Edelman, G. M. (1998). Consciousness and complexity. *Science* 282, 1846–1851. doi: 10.1126/science.282.5395.1846
- Tononi, G., and Sporns, O. (2003). Measuring information integration. *BMC Neurosci.* 4:31. doi: 10.1186/1471-2202-4-31
- Traub, R. D., Contreras, D., Cunningham, M. O., Murray, H., LeBeau, F. E. N., Roopun, A., et al. (2005). Single-column thalamocortical network model exhibiting  $\gamma$  oscillations, sleep spindles, and epileptogenic bursts. *J. Neurophysiol.* 93, 2194–2232. doi: 10.1152/jn.01147.2004
- Tremblay, R., Lee, S., and Rudy, B. (2016). GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* 91, 260–292. doi: 10.1016/j.neuron.2016.06.033
- Urban-Ciecko, J., and Barth, A. L. (2016). Somatostatin-expressing neurons in cortical networks. *Nat. Rev. Neurosci.* 17, 401–409. doi: 10.1038/nrn.2016.53
- Ursino, M., Cona, F., and Zavaglia, M. (2010). The generation of rhythms within a cortical region: analysis of a neural mass model. *Neuroimage* 52, 1080–1094. doi: 10.1016/j.neuroimage.2009.12.084



- Usami, K., Matsumoto, R., Kobayashi, K., Hitomi, T., Shimotake, A., Kikuchi, T., et al. (2015). Sleep modulates cortical connectivity and excitability in humans: direct evidence from neural activity induced by single-pulse electrical stimulation. *Hum. Brain Mapp.* 36, 4714–4729. doi: 10.1002/hbm.22948
- van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., et al. (2018). The threshold for conscious report: signal loss and response bias in visual and frontal cortex. *Science* 360, 537–542. doi: 10.1126/science.aar7186
- Varga, C., Oijala, M., Lish, J., Szabo, G. G., Bezaire, M., Marchionni, I., et al. (2014). Functional fission of parvalbumin interneuron classes during fast network events. *eLife* 3:e04006. doi: 10.7554/elife.04006
- Varotto, G., Fazio, P., Rossi Sebastiano, D., Duran, D., D'Incerti, L., Parati, E., et al. (2014). Altered resting state effective connectivity in long-standing vegetative state patients: an EEG study. *Clin. Neurophysiol.* 125, 63–68. doi: 10.1016/j.clinph.2013.06.016
- Walker, F., Möck, M., Feyerabend, M., Guy, J., Wagener, R. J., Schubert, D., et al. (2016). Parvalbumin- and vasoactive intestinal polypeptide-expressing neocortical interneurons impose differential inhibition on Martinotti cells. *Nat. Commun.* 7:13664. doi: 10.1038/ncomms13664
- Wang, X.-J., and Buzsáki, G. (1996).  $\gamma$  oscillation by synaptic inhibition in a hippocampal interneuronal network model. *J. Neurosci.* 16, 6402–6413. doi: 10.1523/JNEUROSCI.16-20-06402.1996
- Ward, L. M. (2013). The thalamus: gateway to the mind. *Wiley Interdiscip. Rev. Cogn. Sci.* 4, 609–622. doi: 10.1002/wcs.1256
- Wendling, F., Bartolomei, F., Bellanger, J. J., and Chauvel, P. (2002). Epileptic fast activity can be explained by a model of impaired GABAergic dendritic inhibition. *Eur. J. Neurosci.* 15, 1499–1508. doi: 10.1046/j.1460-9568.2002.01985.x
- Wendling, F., Benquet, P., Bartolomei, F., and Jirsa, V. (2016). Computational models of epileptiform activity. *J. Neurosci. Methods* 260, 233–251. doi: 10.1016/j.jneumeth.2015.03.027
- Whittington, M. A., Traub, R. D., Kopell, N., Ermentrout, B., and Buhl, E. H. (2000). Inhibition-based rhythms: experimental and mathematical observations on network dynamics. *Int. J. Psychophysiol.* 38, 315–336. doi: 10.1016/s0167-8760(00)00173-2
- Williams, L. E., and Holtmaat, A. (2019). Higher-order thalamocortical inputs gate synaptic long-term potentiation via disinhibition. *Neuron* 101, 91.e4–102.e4. doi: 10.1016/j.neuron.2018.10.049
- Wilson, H. R., and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* 13, 55–80. doi: 10.1007/bf00288786
- Womelsdorf, T., Valiante, T. A., Sahin, N. T., Miller, K. J., and Tiesinga, P. (2014). Dynamic circuit motifs underlying rhythmic gain control, gating and integration. *Nat. Neurosci.* 17, 1031–1039. doi: 10.1038/nn.3764
- Yang, W., Carrasquillo, Y., Hooks, B. M., Nerbonne, J. M., and Burkhalter, A. (2013). Distinct balance of excitation and inhibition in an interareal feedforward and feedback circuit of mouse visual cortex. *J. Neurosci.* 33, 17373–17384. doi: 10.1523/JNEUROSCI.2515-13.2013
- Yang, G. R., Murray, J. D., and Wang, X.-J. (2016). A dendritic disinhibitory circuit mechanism for pathway-specific gating. *Nat. Commun.* 7:12815. doi: 10.1038/ncomms12815

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a past collaboration with one of the authors FW.

Copyright © 2019 Bensaid, Modolo, Merlet, Wendling and Benquet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Analysis Pipeline for Extracting Features of Cortical Slow Oscillations

Giulia De Bonis<sup>1\*</sup>, Miguel Dasilva<sup>2</sup>, Antonio Pazienti<sup>3</sup>, Maria V. Sanchez-Vives<sup>2,4</sup>, Maurizio Mattia<sup>3</sup> and Pier Stanislaw Paolucci<sup>1</sup>

<sup>1</sup> Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Roma, Rome, Italy, <sup>2</sup> Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain, <sup>3</sup> Istituto Superiore di Sanità (ISS), Rome, Italy, <sup>4</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

## OPEN ACCESS

### Edited by:

Jose Bargas,  
National Autonomous University of  
Mexico, Mexico

### Reviewed by:

Seun Akeju,  
Massachusetts General Hospital,  
Harvard Medical School,  
United States  
Clayton Dickson,  
University of Alberta, Canada

### \*Correspondence:

Giulia De Bonis  
giulia.debonis@roma1.infn.it

**Received:** 08 March 2019

**Accepted:** 05 November 2019

**Published:** 22 November 2019

### Citation:

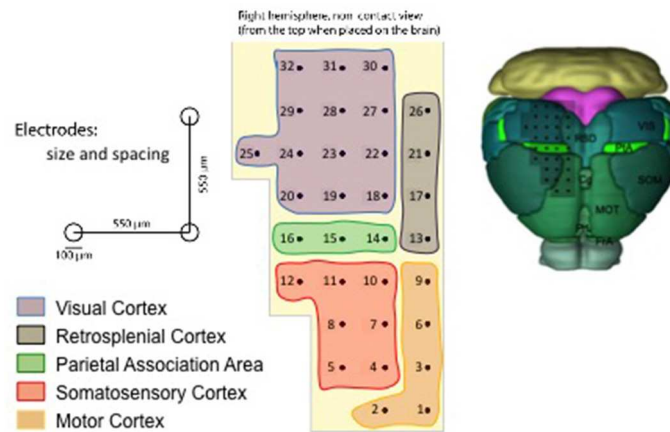
De Bonis G, Dasilva M, Pazienti A, Sanchez-Vives MV, Mattia M and Paolucci PS (2019) Analysis Pipeline for Extracting Features of Cortical Slow Oscillations. *Front. Syst. Neurosci.* 13:70. doi: 10.3389/fnsys.2019.00070

Cortical slow oscillations ( $\lesssim 1$  Hz) are an emergent property of the cortical network that integrate connectivity and physiological features. This rhythm, highly revealing of the characteristics of the underlying dynamics, is a hallmark of low complexity brain states like sleep, and represents a default activity pattern. Here, we present a methodological approach for quantifying the spatial and temporal properties of this emergent activity. We improved and enriched a robust analysis procedure that has already been successfully applied to both *in vitro* and *in vivo* data acquisitions. We tested the new tools of the methodology by analyzing the electrocorticography (ECoG) traces recorded from a custom 32-channel multi-electrode array in wild-type isoflurane-anesthetized mice. The enhanced analysis pipeline, named SWAP (Slow Wave Analysis Pipeline), detects Up and Down states, enables the characterization of the spatial dependency of their statistical properties, and supports the comparison of different subjects. The SWAP is implemented in a data-independent way, allowing its application to other data sets (acquired from different subjects, or with different recording tools), as well as to the outcome of numerical simulations. By using the SWAP, we report statistically significant differences in the observed slow oscillations (SO) across cortical areas and cortical sites. Computing cortical maps by interpolating the features of SO acquired at the electrode positions, we give evidence of gradients at the global scale along an oblique axis directed from fronto-lateral toward occipito-medial regions, further highlighting some heterogeneity within cortical areas. The results obtained using the SWAP will be essential for producing data-driven brain simulations. A spatial characterization of slow oscillations will also trigger a discussion on the role of, and the interplay between, the different regions in the cortex, improving our understanding of the mechanisms of generation and propagation of delta rhythms and, more generally, of cortical properties.

**Keywords:** slow-wave activity, slow oscillations, analysis pipeline, software tools, cortical areas, multi-electrode arrays (MEAs), multi-unit activity (MUA), anesthetized mice

## 1. INTRODUCTION

In a brain manifesting slow-wave activity (SWA), expressed in the cerebral cortex under NREM sleep and deep anesthesia (Steriade et al., 1993), the spiking activity—both single and multi-unit (SUA and MUA), respectively—appears as a regular sequence of Up (high-rate) and Down (almost quiescent) states. Detecting the Up and Down periods and the times of transitions between these states is a long-standing issue for those studies investigating the phenomenon. Depending on the



**FIGURE 1** | Representation of the multi-electrode array (MEA) used for the data acquisition (Pazzini et al., 2018), with the numbering introduced in the SWAP. On the top left, a scheme that indicates the scale of the experiment; the reported dimensions have been adopted to define the reference frame in the pipeline. On the bottom left, the color legend that identifies the cortical areas on the array surface. On the right, an illustration of the grid of electrodes positioned on the mouse cortex (image credit: Allen Institute); the inspected surface is of the order of 10 mm<sup>2</sup>.

available experimental data, different solutions have been developed in the past. At the microscale, threshold-based detection algorithms, leveraging the alternating polarization of the membrane potential in intracellular recordings, have been proposed (Volgushev et al., 2006; Seamari et al., 2007). At the mesoscale, a similar approach has been proposed that focuses on extracellular field potentials by extracting the power of the signal at specific frequency bands (Mukovski et al., 2006). Besides, multi-unit activity sampled from multi-site recordings at the macroscale led to the development of a relatively more sophisticated algorithm detecting Up and Down state transitions by fitting a Hidden Markov Model (Jercog et al., 2017). Here, we improved an approach that relies on the mesoscale signal provided by the MUA extracted as the high-frequency component of the unfiltered field potential. The analysis pipeline, implemented in MATLAB<sup>1</sup>, has already been applied to several experimental data sets, acquired with different setups both *in vitro* and *in vivo* from rodents, ferrets, and monkeys (Ruiz-Mejias et al., 2011, 2016; Mattia et al., 2013; Capone et al., 2019). In this work, we describe an extensive revision of the above referenced procedure, including some new features implemented in Python<sup>2</sup>. The novel components focus on labor-saving automatic adjustment of steering parameters, on automated comparison of data coming from different experiments, and on the evaluation of statistical significance, offering a set of software tools that allow going from raw-data to statistical assessment of results, with consequent advantages in terms of reliability, flexibility, and usability. The current stage of development is the starting point for a cooperative effort aiming at the release to the community of an engineered pipeline<sup>3</sup>.

<sup>1</sup> MATLAB®, The MathWorks, Inc., Natick, Massachusetts, United States.

<sup>2</sup> Python Software Foundation. Python Language Reference, version 2.7.5. Available online at: <http://www.python.org>

<sup>3</sup> A preliminary version of the SWAP, released in the framework of the Human Brain Project, is Available online at: <https://github.com/INM-6/wavescalephant>

The enhanced analysis pipeline has been validated on a new set of data collected *in vivo* using micro-ECoG electrodes (Wang et al., 2009). In particular, the system for brain activity recording employed in this work consists of a 32-electrode array and the test-bench data has been acquired from the cerebral cortex of 11 deeply isoflurane-anesthetized wild-type mice.

The data used in this study was obtained in accordance with the Spanish regulatory laws (BOE-A-2013-6271) which comply with the European Union guidelines on protection of vertebrates used for experimentation (Directive 2010/63/EU of the European Parliament and the Council of 22 September 2010), and the protocol was approved by the Animal Ethics Committee of the University of Barcelona.

The multi-electrode array (MEA) employed for acquiring the data is described in Figure 1 and covers several cortical areas ranging from sensory (V1, S1), motor (M1), and association (PtA) cortices (Pazzini et al., 2018). The unfiltered field potential (UFP) is sampled from each electrode at a frequency of 5 kHz and each acquisition session lasts at least 300 s (see Table 1), thus ensuring a fine inspection of the signal in both space and time. The 11 recording sessions are each from a different animal; that is, 11 independent experiments that collectively represent an extensive data sample covering a wide range of biological and unavoidable physiological variability.

The accurate time-and-space sampling together with the richness of the experimental data have driven the development of new analytical tools that aim to characterize the differences between cortical areas when expressing SWA. In addition, given the unavoidable and physiological variability of the data set, particular attention has been also given to the best strategies to adopt in order to perform a thorough comparison of recordings obtained from the 11 different subjects. The guiding principle when combining data was to keep and use the largest amount of signal, avoiding arbitrary removal of noisy channels or the creation of a subset of “golden” cases. Descriptive statements are accompanied by the assessment of statistical significance of the

**TABLE 1** | Summary of the data acquisition (DAQ) sessions.

Data file	Hemisphere	Duration of the DAQ session [s]	Notes
01	R	304.102	• Excluded: ch. 8 (SD-outlier)
02	R	300.912	• Weak Bimodality: ch. 1, 2, 3, 4, 7, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29 • Excluded: ch. 30, 31, 32 (failure in the DAQ)
03	L	427.223	• Discontinuity in ch. 2, 4, 8, 12, 15, 19, 20, 23, 28, 31, 32 • Excluded: ch. 3, 13 (SD-outlier)
07	L	351.701	–
09	L	321.942	• Discontinuity in ch. 23 • Weak Bimodality: ch. 2, 3, 17, 20 • Excluded: ch. 3, 9 (SD-outlier)
10	L	309.19	• Discontinuity in ch. 12 • Excluded: ch. 12 (failure in the DAQ)
14	R	313.714	–
15	R	329.605	• Excluded: ch. 25 (SD-outlier)
16	R	312.157	• Discontinuity in ch. 14 • Excluded: ch. 4 (SD-outlier); 31 (failure in the DAQ)
17	R	305.967	–
20	R	319.942	• Excluded: ch. 1 (negative asymmetry); 12, 25 (SD-outlier)

The name of the data file reflects the date of acquisition. R for right hemisphere, L for left hemisphere. In general, discontinuities are not critical, since the failure corresponds to an interruption of the DAQ for a limited time interval (usually, a couple of discontinuities per channel, lasting from a few hundreds of millisecond to a few seconds); therefore, discontinuities in the DAQ are managed in the SWAP by identifying, for each problematic channel, the time interval at which the data acquisition fails, and removing it from the time sequence of the signal. By contrast, excluded channels are those for which the signal presents several irregularities, usually resulting in a number of identified upward transitions well below the median computed over the full channel set (tagged as “failure in the DAQ”); in addition, SD-outlier channels are also excluded; the tag “negative asymmetry” corresponds to the case of having a strong negative skew.

results, taking into account the multiplicity of the hypotheses under testing; the obtained claims can give hints as to the mechanisms underlying SWA in mammals.

Furthermore, the outcome of the data analysis can be employed to feed the input of a dedicated spiking neural network simulation (as Paolucci et al., 2013; Pastorelli et al., 2018, 2019), in a data-driven approach of *in silico* studies of the brain, aimed at computing a less stereotypical and a more accurate reproduction of cerebral rhythms, as well as constraining simulations of the cognitive effects of sleep (Capone et al., 2019). The analysis pipeline itself, hereafter named Slow Wave Analysis Pipeline, or SWAP, can be adopted to study the output of the simulation, and to define a set of benchmark observables for confronting models and experiments, with the goal of having a reliable and flexible set of tools available for the characterization of the slow-wave signal in a wide set of cases.

The material in this paper is structured as follows. Section 2 offers an overview of the SWAP, with a collection of results obtained from test-bench data used for illustrating the steps

of the procedure; the focus is on the methods implemented for the identification of the Up/Down state alternation in the ECoG traces, on the techniques for “stacking” and comparing different subjects, and on the statistical treatment of data with hypothesis testing. Section 3 is dedicated to the discussion of methods and results, with concluding remarks and suggestions for future research.

## 2. THE SLOW OSCILLATION ANALYSIS PIPELINE

The study of SWA expressed by the cortex can be tackled starting from a description of the bimodality (i.e., the alternation of Up and Down states), by defining a comprehensive set of observables suited to illustrating the phenomenon of SO. Once this local information is acquired, the second step (not illustrated here) is the characterization of the activity propagation across the brain surface as a wave with delta rhythm. Focusing on the features of Up and Down states, differences between cortical sites can be emphasized.

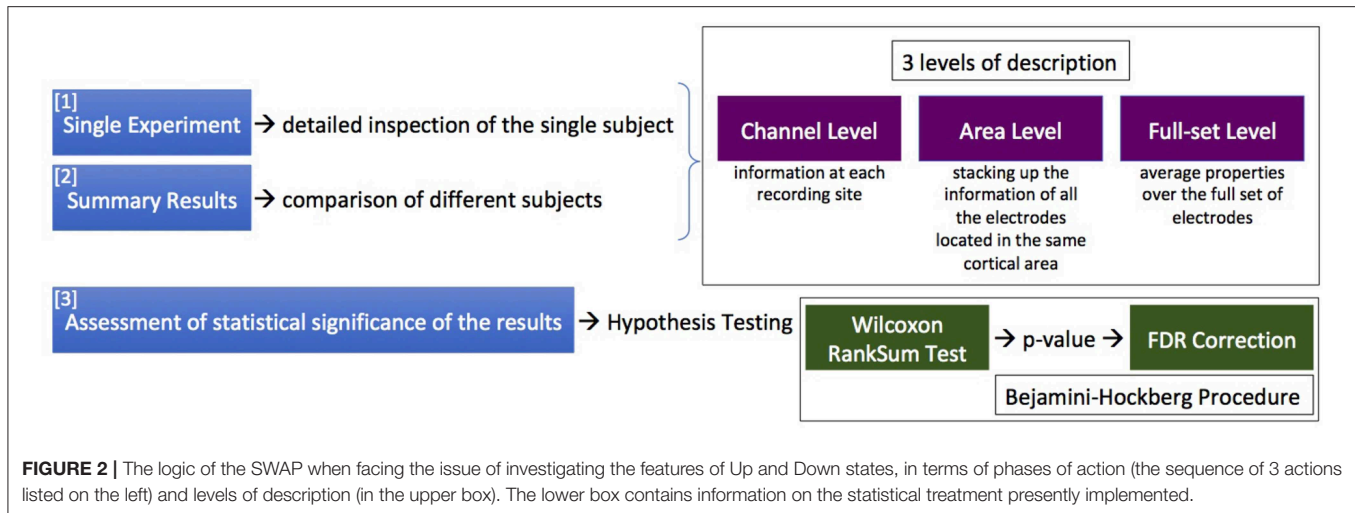
The analysis procedure consists of two phases. First, each data file is examined separately (*Single Experiment*), yielding a detailed inspection of the single subject. The results from the different experiments are then combined (inter-session data) to produce *Summary Results*. Finally, conclusive claims are given with the assessment of statistical significance of the results, taking into account the numerosness of the sample and the multiplicity of the hypotheses under testing.

Three different levels of description can be enabled: (I) the channel level, providing information at each recording site; (II) the area level, stacking up the information of all the electrodes located in the same cortical area; and (III) the full-set level, computing average properties and summary statements for the entire cortex portion under study. The levels of description can be applied at the single experiment, or at the inter-session data; outcomes obtained at the different levels of description are complementary and not strictly separated, and can be superimposed on graphical representations. **Figure 2** details, in terms of phases of action and levels of description, the logic of the SWAP when facing the issue of investigating the features of Up and Down states.

### 2.1. The Channel-Level Description of the Single Experiment: From the Raw Data to the Estimates of MUA and of Transition Times

The channel-level description of the single experiment is obtained with a set of scripts, coded in MATLAB, carrying out a loop over the electrodes in the array to extract the MUA from recordings of the raw signal (UFP). For each channel, the Power Spectral Density (PSD) of the signal is computed and the MUA is used as an estimate of the firing rate of the neurons around the electrode tip (Mattia et al., 2010; Ruiz-Mejias et al., 2011). The logarithm of the MUA is evaluated, and the shape of the  $\log(MUA)$  distribution can be described as a *peak*, at low-MUA values corresponding to Down states, and a *tail*, at high-MUA





values corresponding to the Up states. The  $\log(MUA)$  peak is fitted with a Gaussian function, and the parameters of the fit are used to single out Down-state periods from Up-state periods in the MUA time series. Once the MUA time series is tagged as “Up” and “Down” (binary MUA), a detailed study of the features of such states and of the transitions among them—upward (Down-to-Up) and downward (Up-to-Down)—can be achieved. The workflow is illustrated in **Figure 3**.

In more detail, the initialization phase is carried out with the steering file *setParamsAndOptions*, which takes into account the specificity of the data acquisition (DAQ) sessions, accommodating the frame of reference (the positions of the electrodes in the array) and loading information about the recordings. Some of the settings concern the capability of the algorithm to identify the state transitions, and are fine-tuned following a heuristic approach. The steering file informs what to check in order to perform the analysis pipeline on the given input data. Once the initialization phase is completed, the main loop starts and the flow in the pipeline is as follows (see **Figure 4**):

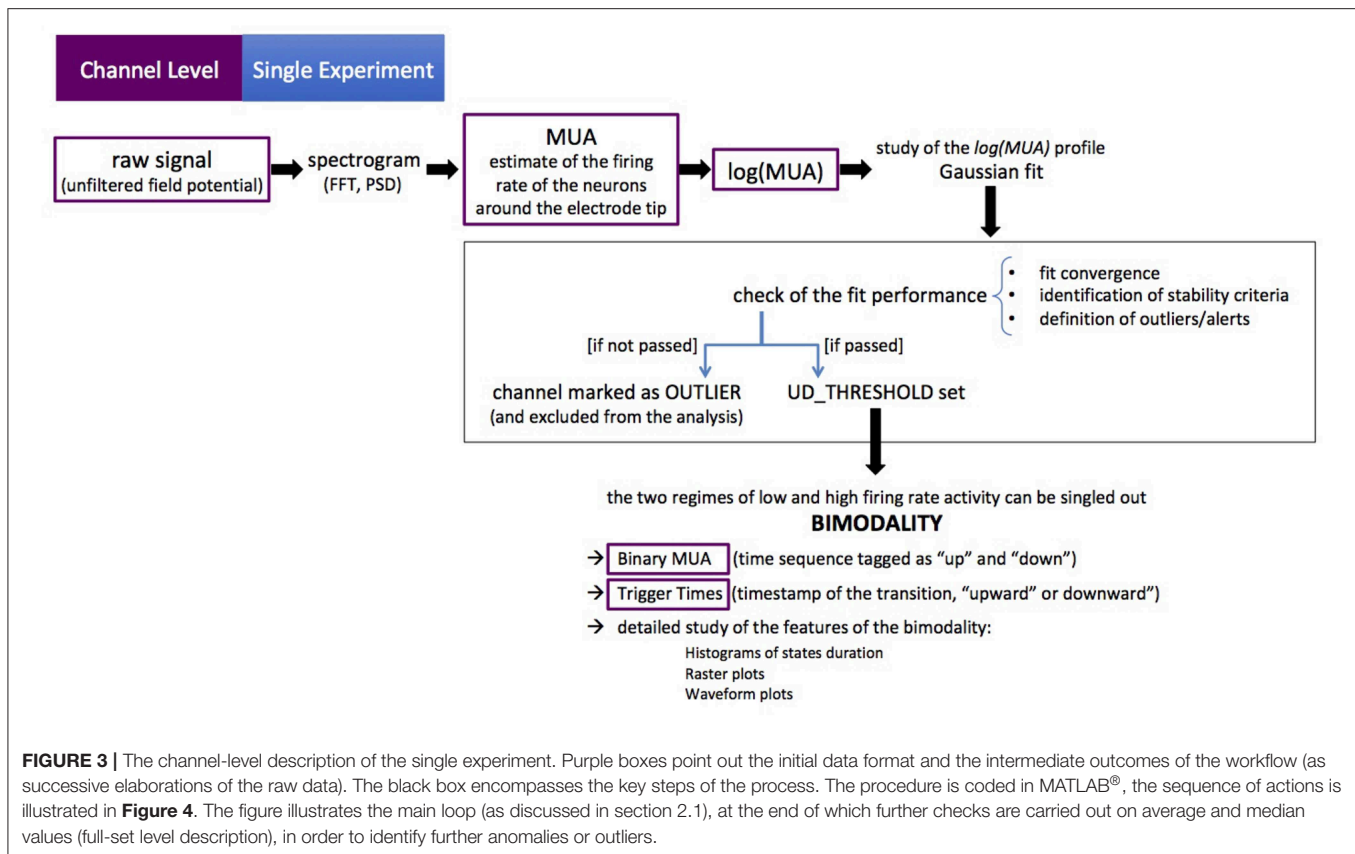
1. Read the analog data (raw signal). For the ECoG data used as a benchmark in this study, input-related actions (open/close the input file and read the input data) are performed by the SONLibrary (Lidierth, 2009), whose functions provide the access to the stored neurophysiological data.
2. Analyze the raw signal in the frequency domain. A moving window of samples is defined for the calculation of the spectrogram; that is, the spectral content of the raw signal as a function of time; the extent of the moving window, together with the sampling frequency, determines the frequency band to be examined. Since the frequency band of interest for the estimate of the MUA is 200–1,500 Hz<sup>4</sup>, a moving window of 5 ms is adopted<sup>5</sup>. FFT (Fast Fourier Transform) and

PSD (Power Spectral Density) are computed using MATLAB functions. Then, for each frequency in the band, the median of the PSD is computed, considering the full set of spectrograms; that is, the collection of moving windows (time steps) that constitute the entire acquisition session. The obtained vector of medians, one value for each frequency, is used as a baseline to normalize the PSD.

3. Evaluate the MUA for each time step as the mean amplitude of the normalized PSD. The MUA is intended as an estimate of the collective firing rate  $r(t)$  of neurons located at the electrode position. The natural logarithm of the MUA is computed, since logarithmic mapping is adopted to emphasize the bimodality of the distribution; because of the normalization to the median of the PSD, negative values and positive values of  $\log(MUA)$  identify a spectral content smaller or larger than the median, respectively (Mattia et al., 2010; Ruiz-Mejias et al., 2011).
4. Fit the distribution of  $\log(MUA)$ , isolating with a Gaussian function the peak corresponding to the (dominant) regime of low-rate states (**Figure 4A**). A simplified description of the bimodality of the SWA in the cortex would require a bimodal fit, and initially the sum of two Gaussian profiles was adopted to describe the shape of the distribution. Conversely, what was discovered after a thorough inspection of a huge number of channels from several different animals (both physiological and pathological subjects at different anesthesia levels) is that, while the Down-state peak is highly stable despite the large variability in the subjects, the content of the  $\log(MUA)$  histogram corresponding to the high-rate regime (obtained from the total distribution after subtracting the Down-state peak) expresses over a large span of “shapes,” and rather than as a “second peak” (with a definite Gaussian profile) it can be generically appointed as a “tail” (**Figure 4A**, inset).
5. Set the *UD\_THRESHOLD*; that is, the level of  $\log(MUA)$  that defines the separation of the two regimes of the bimodality, Up and Down (UD). This is a crucial step in the pipeline, and several options can be adopted to find the optimal criterion,

<sup>4</sup>The experience suggests that frequencies in the raw signal outside this range cannot be associated with the electro-physiological signal of the MUA; in particular, the highest frequency components reflect the presence of electrical noise introduced by the acquisition system.

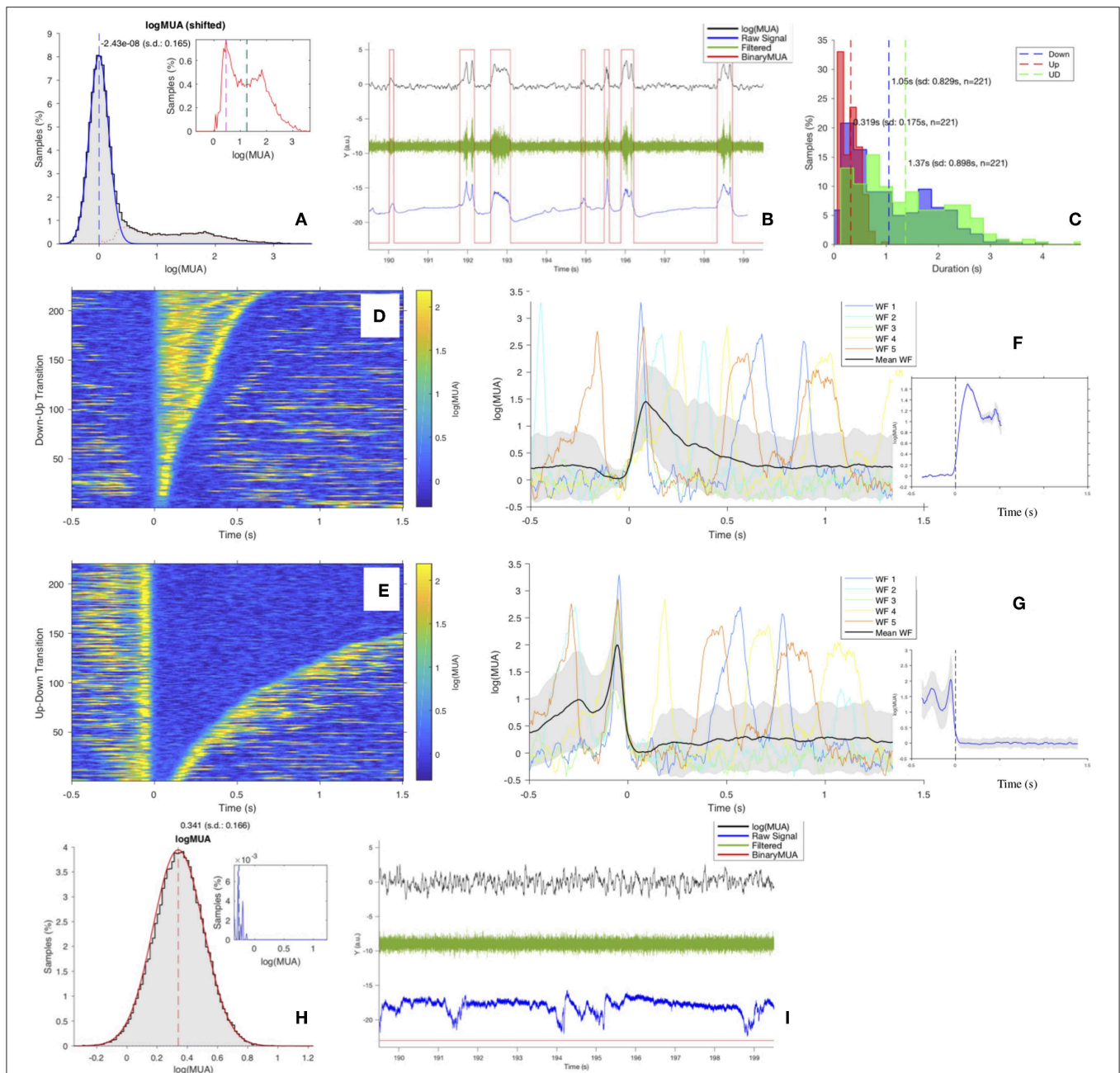
<sup>5</sup>For the benchmark data acquired at 5 kHz, the time window contains 25 samples.



which can depend on the specificity of the data set and on the scope of the analysis. In general, the choice takes into account general settings of the DAQ, the  $\log(MUA)$  distribution, and the results of the Gaussian fit on the Down-state peak (of the given channel, or considering average properties of the recording session). At the channel level in the main loop, checks are introduced to monitor the convergence of the Gaussian fit, the content of the histogram of  $\log(MUA)$  values, and the shape of the distribution (Figure 4A); alerts are activated to signalize: *Weak Bimodality* (if the area of the tail is smaller than a given threshold—currently set at 10% of the total); *Positive Skewness* (if gamma—the coefficient of skewness—of the tail is above 1,  $\gamma > 1$ ); *Negative Skewness* (if  $\gamma < -1$ ); *Right Peak* (if the peak, i.e., the dominant component of  $\log(MUA)$ , is centered at large values, on the right segment of the range, with a tail on the left); *Large Threshold* (if the selected threshold is larger than the mean of the tail); *Few Transitions* (if the number of transitions obtained with the selected threshold is below 3, the minimum requisite to isolate at least an Up state and a Down state). Some of these alerts may help in defining the threshold level, others provide an indication of “problematic” channels to be inspected (that may or may not be tagged as outliers and excluded from further processing at the end of the main loop). Finally, some conditions are “blocking,” in the sense they require a break in the workflow, as is the case with noisy

acquisition channels or spoiled recordings. If no blocking conditions are encountered, meaning that a threshold can be set and Down states and Up states are distinguishable, the following operations are performed:

- Convert the MUA into a binary sequence (*BinaryMUA*) that is 1 or 0 depending on the value of  $\log(MUA)$  above or below the threshold, resulting in the MUA time sequence being tagged as “Up” or “Down” (Figure 4B). Aiming at illustrating the sequence of Up and Down states on different signals, the panel includes  $\log(MUA)$  and *BinaryMUA* together with unfiltered and filtered field potential; the latter is obtained applying the MATLAB *bandpass* function with  $f_{pass} = [200, 1,500]$  Hz and default settings. The horizontal axis corresponds to 10 s of recordings; the vertical axis is in arbitrary units (the raw signal is measured in  $\mu V$ , while MUA is dimensionless since obtained from a ratio of homogeneous quantities), signals values are scaled in order to make all the traces fit on the same plot.
- Label the transitions as upward or downward, and assign the time of transition (*Trigger Time*) with a cubic interpolation of the waveform to locate the time at which the level of MUA crosses the threshold that separates the two regimes of low and high firing rate activity.



**FIGURE 4 |** (A) Distribution of  $\log(MUA)$  as normalized histogram of values shifted at the position of the first peak, The blue curve is the Gaussian fit of the first peak, the  $\mu$  position is marked by the blue vertical dashed line. The red dotted line is the "tail," obtained after subtracting the Gaussian fit from the total distribution (see inset; vertical dashed lines: mode in magenta, median in cyan, mean in black). In this example (Ch. 1, file 01), the tail distribution is far from being Gaussian; the area of the tail is  $\sim 27\%$ , therefore expressing a clear bimodality, coherently with the value of the skewness of the distribution,  $\gamma = 1.97$ . (B) 10 s of recordings of  $\log(MUA)$  (in black),  $BinaryMUA$  (in red), unfiltered raw signal (in blue) and filtered raw signal (in green). (C) Histograms of the duration of Up states (red), of Down states (blue) and of the full cycles (UD cycle, in green); dashed lines, mean values. (D,E) Raster plot of upward (Down-to-Up) and downward (Up-to-Down) transitions, sorted by state duration. (F,G) Waveforms (WFs) of upward and downward transitions, respectively. For each plot, the first 5 transitions are shown, centered around the transition time (Trigger Time); the average transition, computed considering the full set of  $n$  waveforms, is superimposed (black); the shaded area identifies the profile of the waveform at  $\pm 1$  standard deviation (SD) off the mean. The inset shows the average transition front; the shaded area identifies the profile of the waveform at  $\pm 1$  standard error (SEM) of the mean. (H,I) As (A,B), for a channel rejected by the pipeline and excluded from further analysis. The channel presents a monomodal  $\log(MUA)$  with negative asymmetry; as a consequence, no  $UD\_THRESHOLD$  can be set, and no transitions can be identified, evidenced by the filtered field potential (in green) with no signs of state transition and by the  $BinaryMUA$  (in red) that appears constant at low level. Additional details are included in the **Supplementary Material**.



- (c) Study the features of states and transitions, to check the robustness of the algorithm and the stability of the outcome for the different channels. In particular:
  - i. Histograms of the duration of Down states  $d_{DOWN}$ , Up states  $d_{UP}$ , UD-cycles  $d_{UD}$ . The study of these observables is one of the focuses of the analysis; at the channel-level description of the SWAP, the distributions are superimposed for comparison (Figure 4C).
  - ii. Raster plots of states and transitions; each transition (*Trigger Event*) is centered at the trigger time; events can be sorted by their time of occurrence, or by their duration (Figures 4D,E).
  - iii. Waveform plots, for comparing states and transitions, and plots of the average waveform, obtained considering the full set of transitions. A refined algorithm isolates the transition front, to better investigate the transition dynamics (Figures 4F,G).

As anticipated above, once the main loop execution has yielded a full description at the channel level, a key parameter to be monitored for the validation of the procedure is the stability of the conditions used for the identification of the two states (low-MUA and high-MUA). More precisely, since a requirement for the separability of Up and Down states is the successful fit of the Down-state peak of the  $\log(MUA)$  with a Gaussian function (Figure 4A), a similar value for the standard deviation (SD or  $\sigma$ ) of the peak across channels is requested, ensuring comparable SO dynamics of the probed cell assemblies. Indeed, Down states are almost quiescent, and the variability of the MUA is mainly due to the acquisition chain, which has to be the same across channels.

A stable  $\sigma$  allows the application of the same MUA threshold at each recording channel, meaning a unique definition of the Up states, and thus more reliable profiles of traveling wavefronts and a more sound description of the SWA as a collective phenomenon. Therefore, the choice of a fixed  $UD\_THRESHOLD$  is a valid option when the goal of the analysis is to study the dynamics of the propagation of the activity as a slow wave across the cortex. On the other end, this means to decide a key parameter *a priori*, with the burden to be too conservative (increasing the false negative rate) or too tolerant (admitting a larger number of false positives)<sup>6</sup>. Conversely, the option of linking the choice of threshold to “internal agents” (e.g., parameters evaluated during the workflow of the pipeline) can be convenient, for reducing the number of free variables, or for anchoring the false positive rate per channel. A dedicated study of the standard deviation  $\sigma$  has been carried out (section 2.2) on the test-bench data.

Finally, channels not fulfilling the stability requirements are excluded from the analysis and marked as outliers. Here again, the decision on the stability requirements (which parameters to focus on, which threshold levels, which weight to assign

at the different instances, how to define the outliers) is a key element of the pipeline and may largely depend on specific features of the recording sessions and on the goals of the data analysis. As discussed above, the SWAP has set in place alerts, warnings and counters based on parameters considered relevant for the test-bench data, but the elements to be monitored can change if conditions vary. Also configurable are the criteria that define outliers; for results presented here (Table 1), excluded channels are those with the *Right Peak* and with *Few Transitions*, together with the requirement related to the stability of the Gaussian peak, that leaves out channels with  $\sigma$  above  $Q3 + 1.5 \times IQR$  ( $IQR$  is inter-quartile range  $Q3 - Q1$ , with  $Q1$  first quartile and  $Q3$  third quartile). As indicated in the caption of Table 1, the presence of discontinuity in the recording sequence—corresponding to a failure in the data acquisition and a drop in the  $\log(MUA)$ —is not a blocking element, since once the discontinuity is removed, the  $\log(MUA)$  can fulfill the stability requirements. An illustration of a rejected channel is given in Figures 4H,I; further details are included in the Supplementary Material.

## 2.2. Stability of the Data Sample and Channel Selection/Rejection

A key assumption for comparing acquisitions taken at different sites and with different electrodes (or from different animals) is the comparability of the  $\log(MUA)$  profile in the low firing rate regime. A quantitative evaluation of such a requirement is obtained by monitoring the width of the peak fitted with a Gaussian function; that is, the  $\sigma$  parameter of the Gaussian (Figure 4A). The purpose of this analysis step is to evaluate the selection strategy of the  $UD\_THRESHOLD$ , which can be fixed or channel-dependent.

A dedicated routine has been set in place, operating on a given collection of data (inter-session data from different subjects), aiming at stacking the full set of  $\sigma$  parameters estimated from fitting procedures. In Figure 5 we report the statistical distribution of  $\sigma$  values, and we observe large stability across channels and experiments.

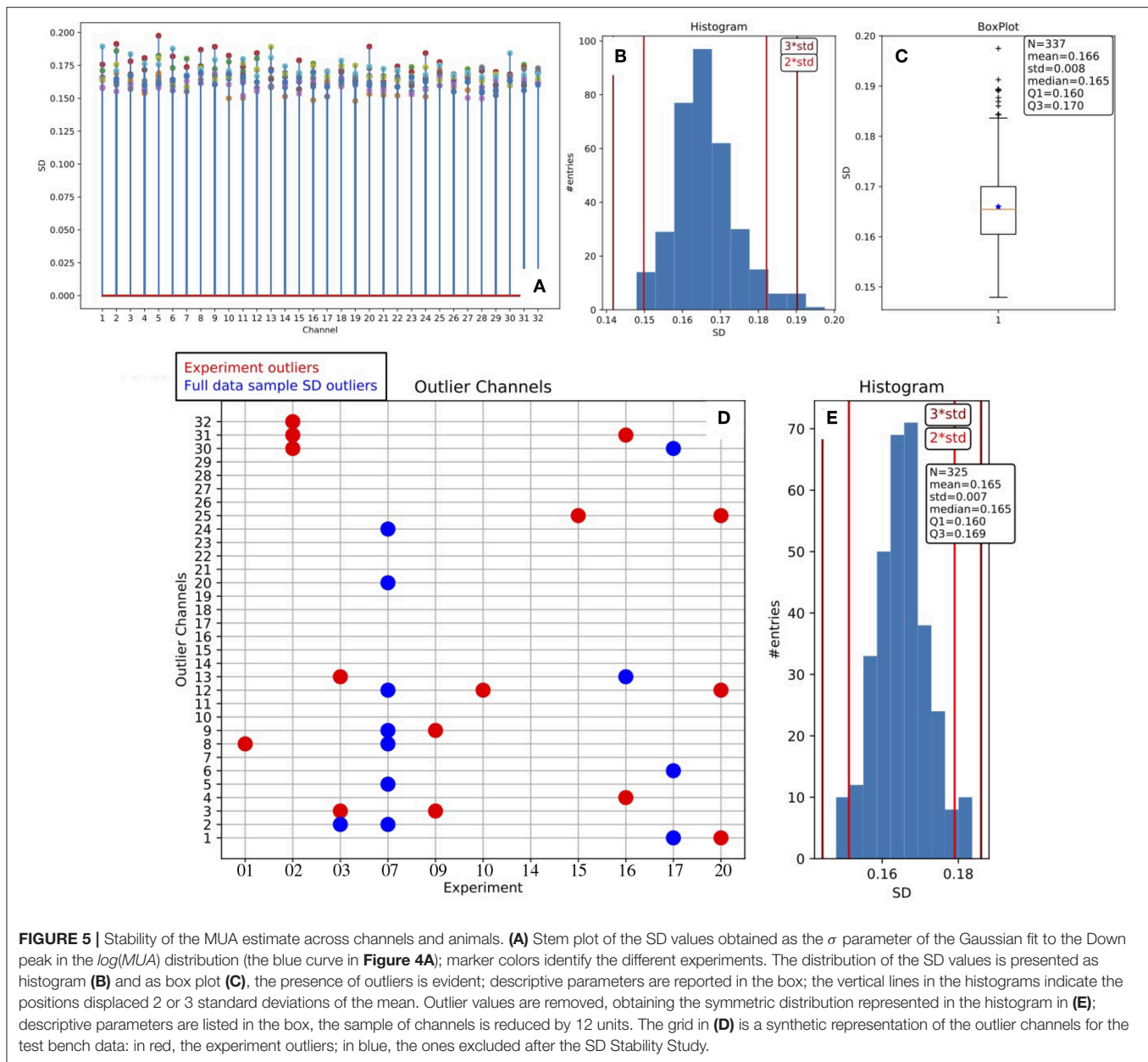
In more detail, the plot offers an overview of the range of variability, showing that the behavior of the parameter is pretty stable inside each experiment, and stable in the entire collection as well<sup>7</sup>. Therefore, we adopted a channel-dependent  $UD\_THRESHOLD$  set at  $2\sigma$ , corresponding to a fixed false positive rate per channel of about 2.25%.

In addition, the results of the *Stability Study* enable channel selection or rejection based on the entire data sample, giving rise to a further list of outliers, to be added to the one already filled out for each DAQ session at the end of the main loop. Considering both lists, a total of 27 outliers were identified and removed from the analysis when inter-session data were taken into account for producing summary results.

<sup>6</sup>In general, less conservative settings are preferable, since the control of false positives can be addressed by checking the spatio-temporal correlations of the propagating signal across the electrodes grid.

<sup>7</sup>Experimental outliers, discussed in section 2.1 and listed in Table 1, are excluded from the representation.





## 2.3. Description at the Cortical Area-Level

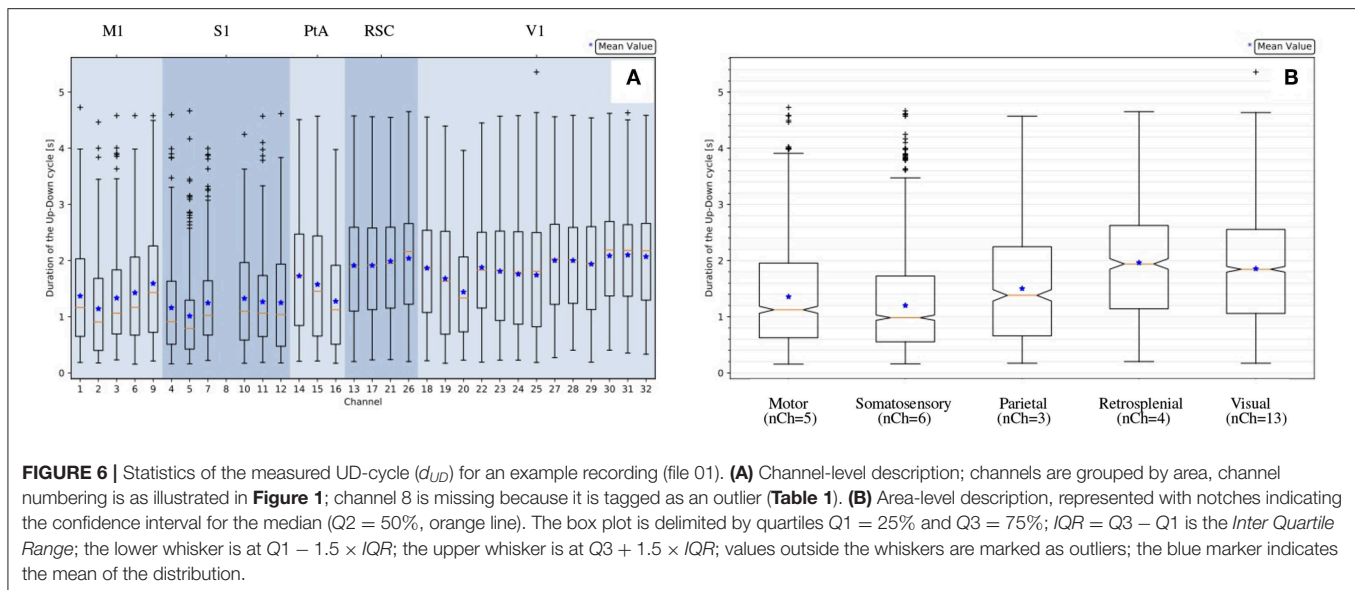
### 2.3.1. A Thumbnail for the Single Experiment

Once the raw signals have been analyzed for each channel, area-related statistics can be obtained as in Ruiz-Mejias et al. (2011). More specifically, the SWAP provides the following observables:

1. Durations of the Down states,  $d_{\text{Down}}[\text{s}]$ ;
2. Durations of the Up states,  $d_{\text{Up}}[\text{s}]$ ;
3. Durations of the UD-cycles (i.e., a pair of consecutive Down and Up states),  $d_{\text{UD}}[\text{s}]$ . The duration of the UD-cycle is an estimate of the SO period;
4. Upward transition slope,  $s_{\text{Up}}[\text{s}^{-1}]$ ;
5. Downward transition slope  $s_{\text{Down}}[\text{s}^{-1}]$ ;
6. Maximum MUA in the Up state ("peak"),  $p[a.u.]$

7. Frequency,  $f[\text{Hz}]$ , defined as  $\frac{1}{d_{\text{UD}}}$ ;

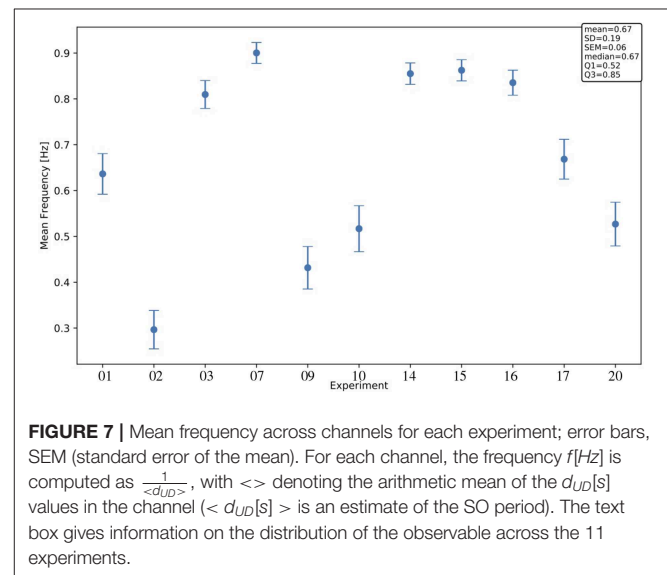
Concerning  $d_{\text{Down}}$ ,  $d_{\text{Up}}$  and  $d_{\text{UD}}$ , as already evident from the histograms (**Figure 4C**) and in agreement with the box plots represented in **Figure 6A**, the distributions at the channel level are skewed and far from being Gaussian. Therefore, for each channel, the median (and not the mean) has been selected as the representative parameter for the observables. The skewness of the distribution is also noticeable at the area level (**Figure 6B**), where the statistics for each distribution are increased since values of electrodes belonging to the same cortical area are grouped together. Therefore, the median is assumed as the representative parameter also at the area level.



Interestingly, significant differences are apparent when comparing median values of different channels and areas. Differences among areas are consistently observed in all animals, despite the large span of values that each observable expresses considering the total of 11 experiments. Indeed, evaluating each observable for each experiment at the full-set level of description, the large variability of the data sample is clearly seen. This is in agreement with the expected biological variability of the subjects, regardless of the identical surgical treatment they have undergone, the comparable drug delivery, and the uniform monitoring conditions during the data taking. In other words, each data session is characterized by its own central values for all the observables of interest, with no clear correlation with the animal's phenotype, or with any other parameter measurable during the data acquisition (Brown et al., 2010; Akeju and Brown, 2017). The spanning of the frequency values across the experiments (**Figure 7**) is in particular revealing, because frequency is a property immediately linkable with the onset of the bimodality and with the propagation of slow waves along the cerebral cortex.

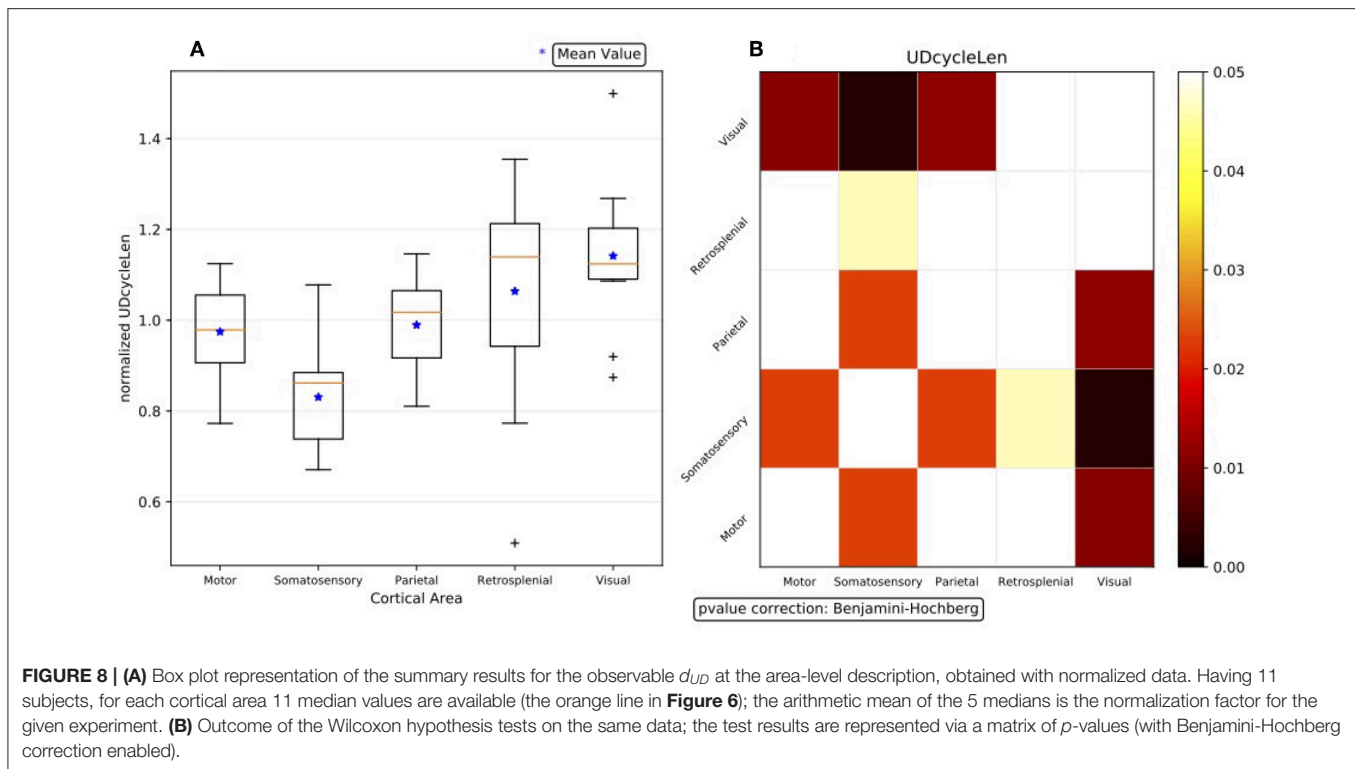
### 2.3.2. Assessment of Statistical Significance at the Area-Level Description for Inter-session Normalized Data

The statistical significance of the effect of differentiation by area, visible for all the observables in the list of interest and for each experiment in the data set, is quantitatively assessed for summary results; that is, when the outcome of the single experiments are combined to drive comprehensive claims on the phenomena. However, the first obstacle met when trying to compare and aggregate results is the large variability exhibited by the different DAQ sessions, which dominates over any other effect and disguises any similarity or common footprint among subjects. Therefore, to confront the area-level descriptions, the median values of the observables for each area are normalized for



each experiment, computing for each observable the arithmetic mean across the cortical areas, and using the obtained mean as a normalization factor for that observable. This procedure highlights any trend expressed at different cortical areas, enabling us to check if different experiments express the same trend. **Figure 8A** shows the summary result at the area-level description obtained with normalized data.

The outcome of the hypothesis tests executed to assess the statistical significance of the differences observed between each couple of cortical areas is represented through a matrix of  $p$ -values, with a graded scale in a given range of confidence (**Figure 8B**). Since no assumptions are made on the model, and because of the already-discussed evidence of non-Gaussianity of the distributions, non-parametric approaches are followed



when analyzing the test bench data, so the Wilcoxon rank-sum test is applied<sup>8</sup>. To take into account multiple comparisons (the so-called “look-elsewhere effect”<sup>9</sup>) and reduce the likelihood of incorrectly rejecting a null hypothesis (type I error) when evaluating a family of simultaneous tests, the robustness of the analysis and the control of the false discovery rate (FDR) is obtained by correcting the  $p$ -values with the Benjamini-Hochberg (BH) procedure (Seabold and Perktold, 2010).

Analogously to what was done for state duration, the study of a possible differentiation at the area-level has also been carried out for slopes and maximum MUA ( $s_{Up}$ ,  $s_{Down}$ ,  $p$ ). Slopes are computed considering the average transition (**Figures 4E,G**), obtained by pooling together the detected Up-to-Down and Down-to-Up transitions. The transition front of the average transition is isolated, considering a 35 ms interval around the transition time  $t_0$  ( $[t_0 - 0.025, t_0 + 0.010]$  for downward transitions;  $[t_0 - 0.010, t_0 + 0.025]$  for upward transitions); the profile is fitted with a cubic and the slope is obtained as the derivative of the polynomial at  $t_0$ . The average upward transition is also used for the estimation of the maximum MUA of the average waveform in the first 250 ms after the transition. Since slopes and maxima are calculated from the average waveform, for each experiment the observables are represented at the channel level by a single value (and not by a distribution). The area-level description for the single experiment

is given by the mean and median of values obtained from all the channels belonging to the specific area. In coherence with the analysis carried out on states duration and in order to apply the same non-parametric Wilcoxon tests, the median is taken as the representative parameter. Summary results are produced after the same normalization adopted for states duration, yielding a similar illustration. A similar approach is followed for frequency  $f$ .

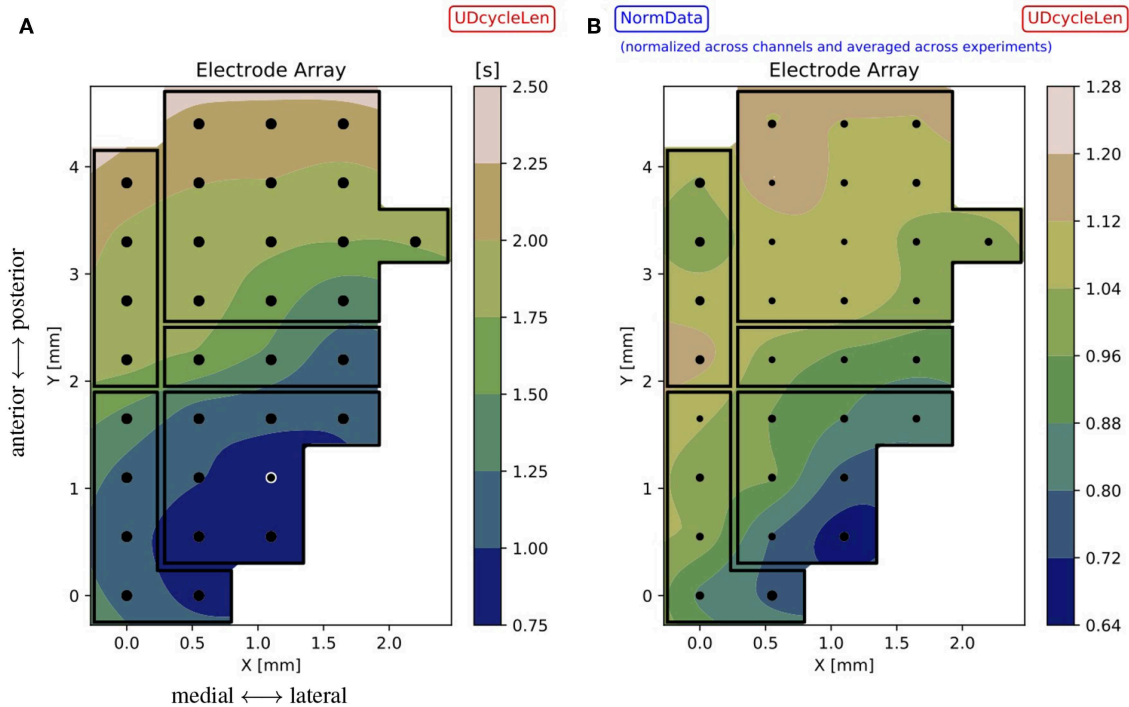
## 2.4. Interpolation of the Array Map

The results obtained with the high spatial resolution probe used for the acquisition of the test bench data can still be improved by interpolating the spaces between the electrodes; the accuracy of the interpolation is assured by the large surface density of sensors offered by the employed MEA. For each experiment and channel, the median ( $d_{Down}$ ,  $d_{Up}$ ,  $d_{UD}$ ) or the mean ( $s_{Up}$ ,  $s_{Down}$ ,  $p$ ,  $f$ ) of the observable is taken. This set of values is used to compute an interpolator<sup>10</sup> that is applied to points ( $xy$  coordinates) on a mesh-grid, made of 50 steps along  $x$  and 90 steps along  $y$ , with a *Grid Step*  $\sim 0.05$  mm, i.e., 1/10 of the *Array Step* = 0.550 mm (**Figure 9A** for the single experiment and for a given observable). The same illustration is applied when representing inter-session normalized data (**Figure 9B**); here, the size of the marker is inversely proportional to the standard deviation of the distribution of values at the given electrode position, thus

<sup>8</sup>The computation of the Wilcoxon rank-sum statistics is carried out with the statistical function `scipy.stats.ranksums` of the `scipy` Python module (Jones et al., 2001)

<sup>9</sup><https://xkcd.com/882/>

<sup>10</sup>The interpolation is carried out with the `scipy` Python module (Jones et al., 2001) using the `scipy.interpolate.Rbf` class for radial basis function (Rbf) interpolation. More on <https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.Rbf.html>



**FIGURE 9 | (A)** Contour plot for the single experiment (here, file 01) illustrating the observable  $d_{UD}$  ( $UDcycleLen$ ) interpolated on the mesh-grid, based on values recorded with the MEA. The array map reflects the DAQ schema of **Figure 1**, with black dots marking the electrode positions and black lines identifying cortical areas; the white circle identifies the outlier channel as resulting from **Figure 5** and **Table 1**. The color bar gives the range of the interpolated variable across the array. **(B)** Summary results for the observable  $d_{UD}$ , represented as a contour plot obtained from normalized data. The interpolation is based on the mean values computed across the experiments, the marker size is inversely proportional to the standard deviation of the distribution of values at each position (the smaller the marker, the less variability measured across the experiments). Since normalized data are used for the plot, the color bar gives indications of trends with respect to the average: for instance, regions colored in blue are characterized by having a duration of the oscillation cycle that is up to 30% less than the average. Comparing **(A)** (single experiment) and **(B)** (inter-session data), a similar tendency is visible for the observable: in particular, the contours evidence gradients along the same dominant antero-posterior direction, from fronto-lateral toward occipito-medial regions. These gradients prevail over the borders of cortical areas, suggesting further differentiation within areas and connections among areas.

measuring the amount of variability registered at that position across the experiments.

## 2.5. From the Array Map to the Assessment of Statistical Significance at the Channel Level for the Normalized Inter-session Data

The contour plots traced on the interpolated array maps give qualitative hints on differentiation among inspected cortical sites, enlightening further details within cortical areas. A more quantitative evaluation can be obtained by performing the statistical analysis of the normalized values resulting from inter-session data, considering each pair of electrodes in the MEA and inspecting if there is a statistically significant difference between the distribution of values measured therein. The idea is to compare the normalized values located at two different electrode positions in the map and to test the validity of the null hypothesis, that is the samples are extracted from the same statistical distribution<sup>11</sup>. We use the Wilcoxon rank-sum

statistics with BH correction on  $p$ -values; in this case, the FDR correction has a larger impact, since for MEAs the number of tests in the family is of the order of hundreds<sup>12</sup>.

Given the large number of hypotheses, electrodes are sorted to emphasize those expressing a significant difference with the others. More specifically, the top-three electrodes in this sorted list are highlighted as “core nodes,” indicating the positions in the cortex displaying the largest significant difference with respect to other positions in the cortex (**Figure 10A**).

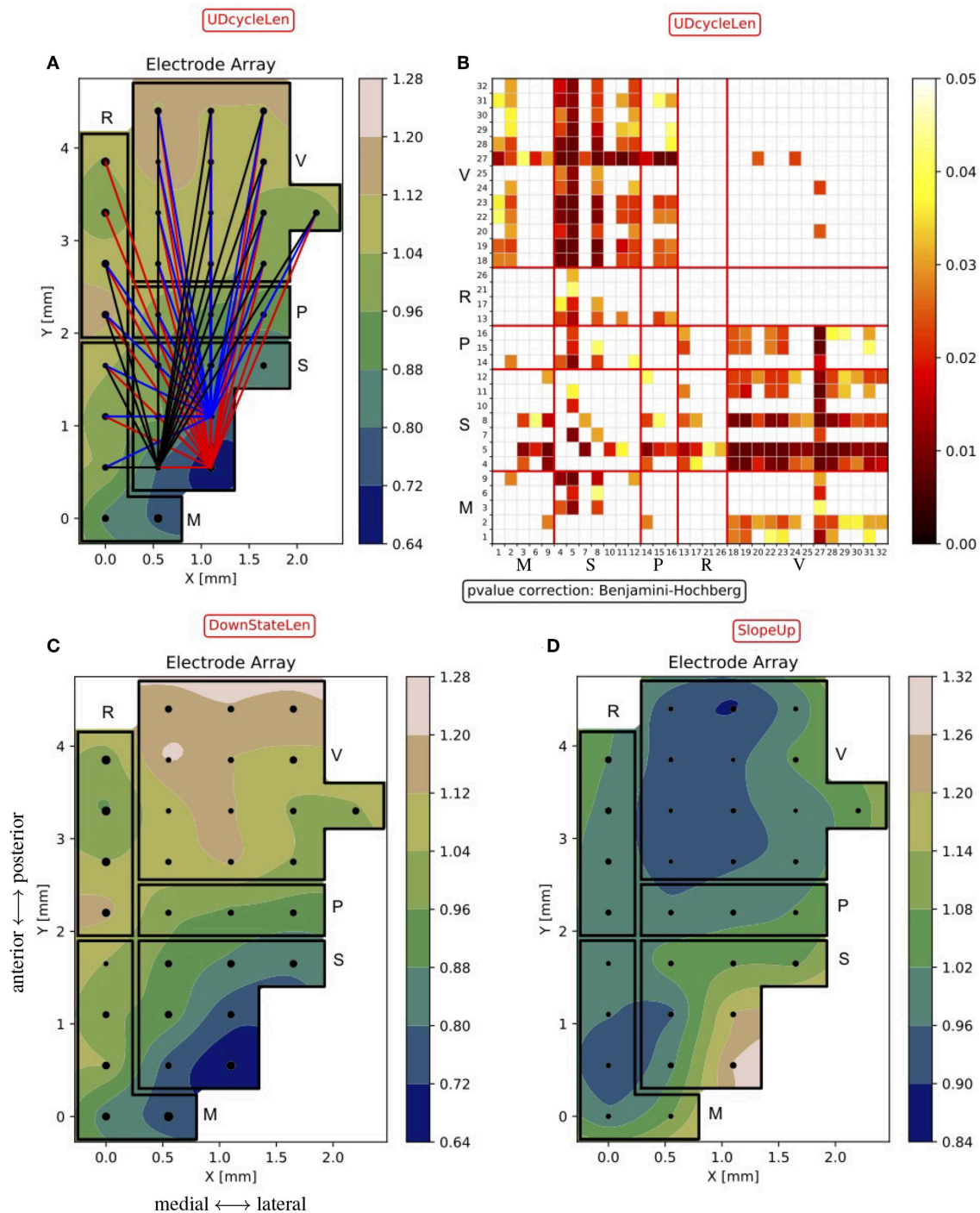
As a result, the somatosensory cortex emerges as the one having the shortest mean Up/Down cycle, in agreement with what is shown in **Figure 8A**. In **Figure 10B**, the differences in the cycle duration at the channel level show a rather homogeneous rhythm within single areas. It is also apparent that the other primary sensory area; that is, the visual cortex, is markedly different from frontal and parietal regions, displaying a significantly longer oscillation period. Inspecting the changes in

maximum of 11 samples can be present at each electrode position, details for each electrode position can be extracted from **Figure 5D**.

<sup>12</sup>If the number of electrodes in the MEA is  $n = 32$ , the number of hypotheses to be checked is  $N = \frac{n(n-1)}{2} = 496$ .

<sup>11</sup>The number of samples at each location depends on how many experiments have channels rejected as outliers at that location. For the test bench data set, a





**FIGURE 10 |** Excitability maps of the probed cortices. **(A)** Normalized  $d_{UD}$  (as in **Figure 9B**) with the “core-nodes” (i.e., the three electrodes with the largest number of significant differences with the other channels). Lines are traced to connect the “core nodes” with their partners in the electrode pair being significantly different ( $p < 0.05$ ). **(B)** Matrix of  $p$ -values of the statistical test on the  $d_{UD}$  differences between channels. The matrix construction is analogous to what is described for **Figure 8**. The numeric sequence of electrodes in the matrix is as in **Figure 6**, i.e., channels are grouped by area (M, motor, S, somatosensory, P, parietal, R, retrosplenial, V, visual). **(C,D)** Normalized Down state duration  $d_{Down}$  (*DownStateLen*) and Down-to-Up transition slope  $s_{Up}$  (*SlopeUp*), averaged across experiments, as in **(A)**.

the Up and Down state duration, we found that the modulation of the oscillation cycle is mainly due to a reduction of the  $d_{Down}$ , as shown in **Figure 10C**. Indeed, only a few channels display a

significant difference in  $d_{Up}$  (not shown) and the  $p$ -value matrix for the  $d_{Down}$  mirrors the one shown in **Figure 10B**. If the local excitability of the cell assemblies underneath the MEA contacts

is well represented by the ratio  $d_{Up}/d_{Down}$  (Ruiz-Mejias et al., 2011; Mattia and Sanchez-Vives, 2012), here the leading role of somatosensory cortex is apparent. Such enhanced excitability is further confirmed by inspecting the slope of the Down-to-Up transition in the MUA (Sanchez-Vives et al., 2010), which is significantly steeper in the same cortical locations (**Figure 10D**).

### 3. DISCUSSION

Biological data are characterized by richness in detail and large variability. The efforts of the data analysis should aim at extracting tendencies and regularities, producing a concise description without hiding or neglecting complexity and details that could convey informative content. This is the guideline followed when developing the SWAP, starting from a solid backbone that has been deeply revised and enriched with new features. In particular, the opportunity of using the MEA data as a test bench has allowed us to focus on the spatial differentiation of the observables, with the aim of uncovering hints as to the local excitability of the cortical assemblies. The developed methodology is robust and easily re-configurable, flexible and adaptable at different acquisition conditions, and also suitable to be applied to the output of simulations. In this framework, the SWAP can be employed in bridge theory, simulations and experiments, providing a set of general tools that allow an effective comparison between heterogeneous data sets. The adoption of a unique analysis procedure is also useful for comparing different simulation engines; the SWAP can be applied to define benchmarks and evaluate the performance of numerical models and implementations. Several studies are ongoing for the application of the SWAP to a large variety of data sets (knock-out mice, subjects in different brain states, data collected with optical techniques), and the stability and reliability of the analysis procedure has so far been confirmed. Moreover, several improvements are ongoing, focusing on the reliability and robustness of the analysis algorithms when increasing the number of channels (electrodes or pixels), as in recently developed multi-electrode arrays and in optical imaging data. The latter in particular is a case where we have already successfully applied our protocol (see Celotto et al., 2018, calcium-imaging data, GCaMP6f model, submitted for publication). In general, the approach we follow can be extended provided that single-channel records allow the reliable disentanglement of Up and Down states of the underlying cortical assemblies. The introduced new features in the analysis pipeline have been coded in Python with the aim of realizing open software tools for the scientific community. The complete transition toward open software is in the list of objectives to fulfill in the near future.

Concerning the interpretation of the results, the analysis of the large set of data collected from 11 wild-type anesthetized mice with the 32-channel MEA allows us to claim a statistically significant differentiation of cortical areas for several parameters that characterize the onset of SWA along the cerebral cortex. Starting from these observables, the excitability of the cortical tissue expressing SWA can be investigated. Indeed, larger

excitability is expected to be associated with faster transition fronts (in particular, upward slopes), shorter duty-cycles (i.e., smaller  $d_{UD}$ , dominated by the duration of the Down states) and accordingly larger frequencies (Sanchez-Vives et al., 2010). For instance, a smaller  $d_{Down}$  reveals the case in which excitability translates to faster Down-to-Up transitions. These features are particularly apparent in the somatosensory area, likely the most excitable cortical region we observed; conversely, the occipital lobe (retrosplenial and visual areas) acts as the least excitable. Activation waves traveling across the cortex during SWA are expected to be sensitive to the diverse cortical excitability, as more reactive (i.e., more excitable) areas are expected to display a smaller “inertia” in recruiting neurons involved in the collective phenomena associated with the high-firing Up states. As a consequence, waves might be originating from highly excitable regions such that preferential propagation pathways are expected, as previously highlighted both in humans (Massimini et al., 2004) and rodents (Ruiz-Mejias et al., 2011; Mohajerani et al., 2013; Stroh et al., 2013). In turn, heterogeneous excitability might also underlie the non-global nature of the SWA phenomenon observed when wakefulness is approaching (Nir et al., 2011; Vyazovskiy et al., 2011) or under pathological conditions (Sanchez-Vives et al., 2017).

It must be pointed out that the study here presented is “static,” in the sense that the absolute time sequence of the events is not taken into account: states and transitions are time-squashed and stacked regardless of their time of occurrence in the DAQ session, and thus any time-dependent effect, like fatigue and recovery intervals of the neurons, that could affect the excitability and alter the responsiveness of cortical regions, is excluded from this analysis. In other words, the figures presented in this paper can be seen as the average photography of the phenomena investing the monitored cortex during the acquisition period. Related to this, excitability and responsiveness can also be altered by wave propagation dynamics, in particular if different propagation patterns coexist and overlap in the same time interval, for instance when two slow waves originating at distinct sites travel along the cortex at different speeds and each one follows its own path (Sancristóbal et al., 2016; Capone et al., 2019). Again, not using the information on the absolute timing of the events at the electrode positions and intending the results presented here to be the average photography of the SWA in the cortex, the included results reflect the dominant propagation pattern, namely the antero-posterior direction (Massimini et al., 2004; Ruiz-Mejias et al., 2011; Stroh et al., 2013; Chan et al., 2015), in particular along an oblique axis directed from fronto-lateral toward occipito-medial regions, as suggested by values depicted in the contour plots. As in this step of the SWAP we have not focused on time-dependent effects, and their impact on the area-differentiation of the bistability modes will be evaluated in further studies.

Finally, although in all the animals involved in this study the level of administered anesthesia was the same, the observed inter-subject variability (**Figure 7**) could in principle be explained by animals being in different brain states. This suggests the possibility of exploiting the characterization of slow oscillations in the cerebral cortex as a new tool for effective classification

of the brain states. Several techniques are currently under study, based for instance on using the principal components analysis (PCA) to identify and single out the different sources of variability in the experimental data set. Indeed, a more reliable classification of brain states (i.e., of the DAQ sessions that constitute the statistical sample) would provide a more robust comparison of the experiments, allowing us to overcome the limits derived from the need to use normalized data.

## DATA AVAILABILITY STATEMENT

Data will be available on reasonable request.

## ETHICS STATEMENT

This study was carried out in accordance with the Spanish regulatory laws (BOE-A-2013-6271) which comply with the European Union guidelines on protection of vertebrates used for experimentation (Directive 2010/63/EU of the European Parliament and the Council of 22 September 2010), and the protocol was approved by the Animal Ethics Committee of the University of Barcelona.

## AUTHOR CONTRIBUTIONS

MS-V, MM, MD, PP, and GD conceived and designed the study. MD performed the data acquisition under the supervision of MS-V. MM provided the backbone of the analysis pipeline. GD revised the analysis pipeline, improved it with new functions,

applied it to the test bench data, performed the statistical analysis, wrote the first draft of the manuscript. PP and MM contributed to the data analysis and to the interpretation of the results. AP revised sections of the manuscript. All authors contributed to manuscript revision, and read and approved the submitted version.

## FUNDING

This work has received funding from the European Union Horizon 2020 Research and Innovation Programme under Specific Grant Agreements No. 785907 (HBP SGA2) and No. 720270 (HBP SGA1).

## ACKNOWLEDGMENTS

This study was carried out in the framework of the Human Brain Project (HBP<sup>13</sup>), funded under Specific Grant Agreements No. 785907 (HBP SGA2) and No. 720270 (HBP SGA1), in particular within activities of sub-project SP3 (Systems and Cognitive Neuroscience). This manuscript has been released as a Pre-Print at De Bonis et al. (2019).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnsys.2019.00070/full#supplementary-material>

<sup>13</sup><https://www.humanbrainproject.eu/en/>

## REFERENCES

- Akeju, O., and Brown, E. N. (2017). Neural oscillations demonstrate that general anesthesia and sedative states are neurophysiologically distinct from sleep. *Curr. Opin. Neurobiol.* 44, 178–185. doi: 10.1016/j.conb.2017.04.011
- Brown, E. N., Lydic, R., and Schiff, R. (2010). General anesthesia, sleep, and coma. *N. Engl. J. Med.* 363, 2638–2650. doi: 10.1056/NEJMra0808281
- Capone, C., Pastorelli, E., Golosio, B., and Paolucci, P. S. (2019). Sleep-like slow oscillations improve visual classification through synaptic homeostasis and memory association in a thalamo-cortical model. *Sci. Rep.* 9:8990. doi: 10.1038/s41598-019-45525-0
- Capone, C., Rebollo, B., Muñoz, A., Illa, X., Del Giudice, P., Sanchez-Vives, M. V., et al. (2019). Slow waves in cortical slices: how spontaneous activity is shaped by laminar structure. *Cereb. Cortex* 29, 319–335. doi: 10.1093/cercor/bhx326
- Celotto, M., De Luca, C., Muratore, P., Resta, F., Allegra Mascaro, A. L., Pavone, F. S., et al. (2018). Analysis and model of cortical slow waves acquired with optical techniques. *arXiv:1811.11687*.
- Chan, A. W., Mohajerani, M. H., LeDue, J. M., Wang, Y. T., and Murphy, T. H. (2015). Mesoscale infraslow spontaneous membrane potential fluctuations recapitulate high-frequency activity cortical motifs. *Nat. Comm.* 6, 1–12. doi: 10.1038/ncomms8738
- De Bonis, G., Dasilva, M., Pazienti, A., Sanchez-Vives, M. V., Mattia, M., and Paolucci, P. S. (2019). Slow waves analysis pipeline for extracting the features of the bi-modality from the cerebral cortex of anesthetized mice. *arXiv:1902.08599*.
- Jercog, D., Roxin, A., Barthó, P., Luczak, A., Compte, A., and de la Rocha, J. (2017). Up-down cortical dynamics reflect state transitions in a bistable network. *eLife* 6:e22425. doi: 10.7554/eLife.22425
- Jones, E., Oliphant, T., and Peterson, P. (2001). *SciPy: Open Source Scientific Tools for Python*. Available online at: <http://www.scipy.org/> (accessed November 13, 2019).
- Lidierth, M. (2009). sigTOOL: a MATLAB-based environment for sharing laboratory-developed software to analyze biological signals. *J. Neurosci. Methods* 178, 188–196. doi: 10.1016/j.jneumeth.2008.11.004
- Massimini, M., Huber, R., Ferrarelli, F., Hill, S. L., and Tononi, G. (2004). The sleep slow oscillation as a traveling wave. *J. Neurosci.* 24, 6862–6870. doi: 10.1523/JNEUROSCI.1318-04.2004
- Mattia, M., Ferraina, S., and Del Giudice, P. (2010). Dissociated multi-unit activity and local field potentials: a theory inspired analysis of a motor decision task. *Neuroimage* 52, 812–823. doi: 10.1016/j.neuroimage.2010.01.063
- Mattia, M., Pani, P., Mirabella, G., Costa, S., Del Giudice, P., and Ferraina, S. (2013). Heterogeneous attractor cell assemblies for motor planning in premotor cortex. *J. Neurosci.* 33, 11155–11168. doi: 10.1523/JNEUROSCI.4664-12.2013
- Mattia, M., and Sanchez-Vives, M. V. (2012). Exploring the spectrum of dynamical regimes and timescales in spontaneous cortical activity. *Cogn. Neurodyn.* 6, 239–250. doi: 10.1007/s11571-011-9179-4
- Mohajerani, M. H., Chan, A. W., Mohsenvand, M., LeDue, J., Liu, R., McVea, D. A., et al. (2013). Spontaneous cortical activity alternates between motifs defined by regional axonal projections. *Nat. Neurosci.* 16, 1426–1435. doi: 10.1038/nn.3499
- Mukovski, M., Chauvette, S., Timofeev, I., and Volgushev, M. (2006). Detection of active and silent states in neocortical neurons from the field potential signal during slow-wave sleep. *Cereb. Cortex* 17, 400–414. doi: 10.1093/cercor/bhj157
- Nir, Y., Staba, R. J., Andrillon, T., Vyazovskiy, V. V., Cirelli, C., Fried, I., et al. (2011). Regional slow waves and spindles in human sleep. *Neuron* 70, 153–169. doi: 10.1016/j.neuron.2011.02.043

- Paolucci, P. S., Ammendola, R., Biagioni, A., Frezza, O., Lo Cicero, F., Lonardo, A., et al. (2013). Distributed simulation of polychronous and plastic spiking neural networks: strong and weak scaling of a representative mini-application benchmark executed on a small-scale commodity cluster. *arXiv:1310.8478*.
- Pastorelli, E., Capone, C., Simula, F., Sanchez-Vives, M. V., Del Giudice, P., Mattia, M., et al. (2019). Scaling of a large-scale simulation of synchronous slow-wave and asynchronous awake-like activity of a cortical model with long-range interconnections. *Front. Syst. Neurosci.* 13:33. doi: 10.3389/fnsys.2019.00033
- Pastorelli, E., Paolucci, P. S., Simula, F., Biagioni, A., Capuani, F., Cretaro, P., et al. (2018). "Gaussian and exponential lateral connectivity on distributed spiking neural network simulation," in *2018 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)* (Cambridge, UK), 658–665. doi: 10.1109/PDP2018.2018.00110
- Pazzini, L., Polese, D., Weinert, J. F., Maiolo, L., Maita, F., Marrani, M., et al. (2018). An ultra-compact integrated system for brain activity recording and stimulation validated over cortical slow oscillations *in vivo* and *in vitro*. *Sci. Rep.* 8:16717. doi: 10.1038/s41598-018-34560-y
- Ruiz-Mejias, M., Ciria-Suarez, L., Mattia, M., and Sanchez-Vives, M. V. (2011). Slow and fast rhythms generated in the cerebral cortex of the anesthetized mouse. *J. Neurophysiol.* 106, 2910–2921. doi: 10.1152/jn.00440.2011
- Ruiz-Mejias, M., Perez-Mendez, L., Ciria-Suarez, L., Martinez de Lagran, M., Gener, T., Sancristobal, B., et al. (2016). Over-expression of Dyrk1A, a Down syndrome candidate, decreases excitability and impairs gamma oscillations in the pre-frontal Cortex. *J. Neurosci.* 36, 3648–3659. doi: 10.1523/JNEUROSCI.2517-15.2016
- Sanchez-Vives, M. V., Massimini, M., and Mattia, M. (2017). Shaping the default activity pattern of the cortical network. *Neuron* 94, 993–1001. doi: 10.1016/j.neuron.2017.05.015
- Sanchez-Vives, M. V., Mattia, M., Compte, A., Perez-Zabalza, M., Winograd, M., Descalzo, V. F., et al. (2010). Inhibitory modulation of cortical up states. *J. Neurophysiol.* 104, 1314–1324. doi: 10.1152/jn.00178.2010
- Sancristóbal, B., Rebollo, B., Boada, P., Sanchez-Vives, M. V., and Garcia-Ojalvo, J. (2016). Collective stochastic coherence in recurrent neuronal networks. *Nat. Phys.* 12, 1–8. doi: 10.1038/nphys3739
- Seabold, S., and Perktold, J. (2010). "Statsmodels: econometric and statistical modeling with python," in *9th Python in Science Conference* (Austin, TX).
- Seamari, Y., Narváez, J. A., Vico, F. J., Lobo, D., and Sanchez-Vives, M. V. (2007). Robust off- and online separation of intracellularly recorded up and down cortical states. *PLoS ONE* 2:e888. doi: 10.1371/journal.pone.0000888
- Steriade, M., Nunez, A., and Amzica, F. (1993). A novel slow (< 1 Hz) oscillation of neocortical neurons *in vivo*: depolarizing and hyperpolarizing components. *J. Neurosci.* 13, 3252–3265.
- Stroh, A., Adelsberger, H., Groh, A., Rühlmann, C., Fischer, S., Schierloh, A., et al. (2013). Making waves: initiation and propagation of corticothalamic Ca<sup>2+</sup> waves *in vivo*. *Neuron* 77:1136–1150. doi: 10.1016/j.neuron.2013.01.031
- Volgushev, M., Chauvette, S., Mukovski, M., and Timofeev, I. (2006). Precise long-range synchronization of activity and silence in neocortical neurons during slow-wave sleep. *J. Neurosci.* 26, 5665–5672. doi: 10.1523/JNEUROSCI.0279-06.2006
- Vyazovskiy, V. V., Olcese, U., Hanlon, E. C., Nir, Y., Cirelli, C., and Tononi, G. (2011). Local sleep in awake rats. *Nature* 472, 443–447. doi: 10.1038/nature10009
- Wang, W., Degenhart, A. D., Collinger, J. L., Vinjamuri, R., Sudre, G. P., Adelson, P. D., et al. (2009). "Human motor cortical activity recorded with micro-ecog electrodes, during individual finger movements," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Minneapolis, MN), 586–589. doi: 10.1109/IEMBS.2009.5333704

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 De Bonis, Dasilva, Pazienti, Sanchez-Vives, Mattia and Paolucci. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Bridging Single Neuron Dynamics to Global Brain States

Jennifer S. Goldman<sup>1\*</sup>, Núria Tort-Colet<sup>1</sup>, Matteo di Volo<sup>1†</sup>, Eduarda Susin<sup>1</sup>, Jules Bouté<sup>1</sup>, Melissa Dali<sup>1</sup>, Mallory Carlu<sup>1</sup>, Trang-Anh Nghiem<sup>2</sup>, Tomasz Górski<sup>1</sup> and Alain Destexhe<sup>1</sup>

<sup>1</sup> Department of Integrative and Computational Neuroscience (ICN), Centre National de la Recherche Scientifique (CNRS), Paris-Saclay Institute of Neuroscience (NeuroPSI), Gif-sur-Yvette, France, <sup>2</sup> Department of Physics, Ecole Normale Supérieure, Paris, France

## OPEN ACCESS

### Edited by:

Per E. Roland,  
University of Copenhagen, Denmark

### Reviewed by:

Mario Rosanova,  
University of Milan, Italy  
Sacha Jennifer van Albada,  
Jülich Research Centre, Germany

### \*Correspondence:

Jennifer S. Goldman  
jennifer.goldman@mail.mcgill.ca

### †Present address:

Matteo di Volo,  
Laboratoire de Physique Théorique et  
Modélisation, Université de  
Cergy-Pontoise, Cergy-Pontoise,  
France

**Received:** 28 June 2019

**Accepted:** 19 November 2019

**Published:** 06 December 2019

### Citation:

Goldman JS, Tort-Colet N, di Volo M, Susin E, Bouté J, Dali M, Carlu M, Nghiem T-A, Górski T and Destexhe A (2019) Bridging Single Neuron Dynamics to Global Brain States. *Front. Syst. Neurosci.* 13:75. doi: 10.3389/fnsys.2019.00075

Biological neural networks produce information backgrounds of multi-scale spontaneous activity that become more complex in brain states displaying higher capacities for cognition, for instance, attentive awake versus asleep or anesthetized states. Here, we review brain state-dependent mechanisms spanning ion channel currents (microscale) to the dynamics of brain-wide, distributed, transient functional assemblies (macroscale). Not unlike how microscopic interactions between molecules underlie structures formed in macroscopic states of matter, using statistical physics, the dynamics of microscopic neural phenomena can be linked to macroscopic brain dynamics through mesoscopic scales. Beyond spontaneous dynamics, it is observed that stimuli evoke collapses of complexity, most remarkable over high dimensional, asynchronous, irregular background dynamics during consciousness. In contrast, complexity may not be further collapsed beyond synchrony and regularity characteristic of unconscious spontaneous activity. We propose that increased dimensionality of spontaneous dynamics during conscious states supports responsiveness, enhancing neural networks' emergent capacity to robustly encode information over multiple scales.

**Keywords:** computational neuroscience, neural network models, mean-field models, membrane biophysics, low-dimensional manifold, cerebral cortex, coupling, desynchronized

## INTRODUCTION

Brain activity transitions between healthy states, including stages of sleep, restful and aroused waking, as well as pathological states such as epilepsy, coma, and unresponsive wakefulness syndrome. From such a diversity of brain states, phenomenological categories encompassing similar spatio-temporal activity patterns can roughly, but usefully, be defined: unconscious (e.g., sleep and anesthesia) and conscious (e.g., waking and dreaming) brain states. At the macroscopic, global scale, unconscious brain states are dominated by high voltage, low frequency oscillatory brain activity related to the microscopic alternation of synchronous neuronal spiking and near silence (Steriade et al., 1993; Brown et al., 2010). Conversely, conscious states are macroscopically characterized by low voltage, high frequency, complex “disorganized” dynamics resulting from more asynchronous irregular (AI) microscopic network activity (Tsodyks and Sejnowski, 1995; Van Vreeswijk and Sompolinsky, 1996; Brunel, 2000), thought to be important for neural coding (Skarda and Freeman, 1987; Van Vreeswijk and Sompolinsky, 1996; Tononi and Edelman, 1998; Zerlaut and Destexhe, 2017).

Much as different states of matter like solids, liquids, and gases emerge from interactions between populations of molecules, different brain states may emerge from the interactions between populations of neurons. Statistical physics provides a mathematical framework to

uncover structures of microscopic interactions underlying macroscopic properties. In this sense, macroscopically observed high synchrony, low complexity brain signals recorded from unconscious states may be accounted for by an increased coupling in the system's components, behaving more like a solid (Peyrache et al., 2012; Le Van Quyen et al., 2016; Olcese et al., 2016; Nghiem et al., 2018a). In contrast, conscious brain states may be described as higher complexity (Sitt et al., 2014; Engemann et al., 2018; Nghiem et al., 2018a), perhaps liquid-like.

Though quantitative expressions directly linking order and complexity are not straightforward, various definitions and metrics of complexity have been described to vary between brain states. Reports of enhanced complexity in conscious compared to unconscious states may be understood as increased dimensionality (El Boustani and Destexhe, 2010), namely the number of degrees of freedom needed to capture a system's dynamics. Intuitively, dimensionality relates, though is not reducible to, algorithmic complexity which quantifies the length of a deterministic algorithm required to reproduce an exact signal. For a random signal resulting from purely stochastic dynamics (similar to neural activity observed during conscious states), the length of the algorithm would be as long as the signal itself. In contrast, a purely oscillatory signal (reminiscent of unconscious brain dynamics) can be recapitulated by a shorter algorithm, easily described by a periodic trajectory in few dimensions.

Here, we aim to connect spatial scales from microscopic (nanometers to micrometers—molecules to whole neurons) to macroscopic brain activity (centimeters to meters—brain areas to individual subjects' brains), describing both spontaneous and evoked dynamics. Toward linking interpretations of studies between scales, mesoscopic data (micrometers to millimeters—populations of thousands to tens of thousands of neurons) have been useful to inform models of neuronal assemblies. The perspective concludes by discussing a hypothesis best tested with a multi-scale understanding of brain function: the global complexity of neural activity increases in conscious brain states so as to enhance responsiveness to stimuli. We suggest responsiveness may depend on the capacity of neural networks to transiently collapse the dimensionality of collective dynamics—in particular neural assemblies sensitive to stimulus features—into evoked low-dimensional trajectories supporting neural codes (**Figure 1A**).

## MACROSCOPIC SIGNALS VARY ROBUSTLY BETWEEN BRAIN STATES

Both spontaneous and evoked (**Figures 1A,B**) neural signals vary macroscopically across brain states, as demonstrated in electroencephalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI). In unconscious states, neural activity is dominated by low-frequency, high-amplitude signals (Niedermeyer and Lopes da Silva, 2005). Accordingly, analyses of entropy (Sitt et al., 2014; Engemann et al., 2018), complexity (Tononi and Edelman, 1998), and dimensionality (El Boustani and Destexhe, 2010)

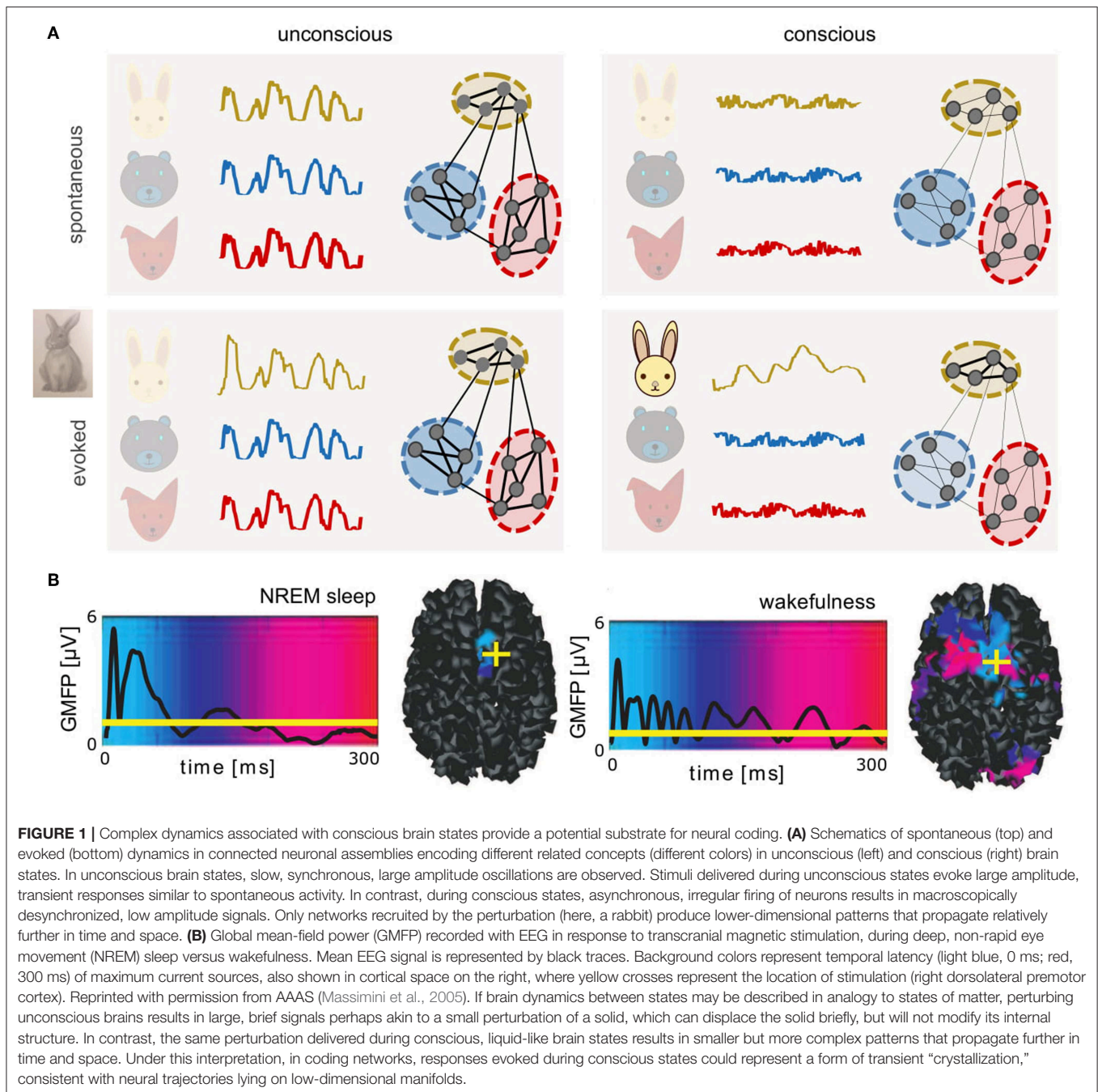
during unconscious states indicate a relative simplicity of signals compared to conscious states. In unconscious states, synchronous activity slowly sweeps across the cortex (Massimini et al., 2004) along paths formed by cortical tracts (Capone et al., 2017). In both conscious resting and unconscious states, the default mode network (Raichle et al., 2001; Boly et al., 2008) establishes a pattern of synchronization between brain areas, producing correlations in ultra-slow ( $< 1$  Hz) dynamics (Brookes et al., 2011). Sustained, slow oscillations were initially reported in the thalamocortical system (Steriade, 2003), but are also observed experimentally in isolated cortex, without thalamus (Sanchez-Vives and McCormick, 2000; Timofeev et al., 2000). Thalamocortical connections shape slow wave dynamics (Destexhe et al., 2007; Poulet et al., 2012; David et al., 2013; Crunelli et al., 2015; Zucca et al., 2019) although slow oscillations appear to be the default state of cortical networks (Sanchez-Vives and McCormick, 2000; Sanchez-Vives et al., 2017).

Patterns of neocortical regions activated in resting state networks have been successfully retrieved using eigenmodes of the structural connectivity matrix, i.e., the possible oscillatory patterns at frequencies allowed by white matter tract lengths (Atasoy et al., 2016). In active states, the executive control network replaces the default mode (Fox et al., 2005), and the co-activation of different cortical regions is more strongly controlled by correlations in external stimuli than by white matter structural connectivity (Gilson et al., 2018), with patterns of activity propagating recurrently between low-level, sensory areas and high-level, associative areas.

During conscious states, on the background of globally disorganized neural activity, transient patterns emerge (Duncan-Johnson and Donchin, 1982; Goodin and Aminoff, 1984; Sur and Sinha, 2009; Uhlhaas et al., 2009; Luck and Kappenman, 2011; Churchland et al., 2012; Sato et al., 2012; Singer, 2013; Chemla et al., 2019). Under an interpretation of brain states in analogy to states of matter, microscopic changes in the interactions between neurons could permit the emergence of larger-scale structures in brain activity.

## MICROSCOPIC MECHANISMS; BIOPHYSICS OF BRAIN STATES

Experiments have demonstrated that during unconscious brain states, the membrane potential ( $V_m$ ) of single cells slowly oscillates between hyperpolarized and depolarized potentials associated with alternating periods of silence (Down states, also termed "OFF periods") and AI-like firing (Up states, also termed "ON periods") (Steriade et al., 1993) (**Figure 2A**). During conscious brain states, neurons show sustained but sparse and irregular AI firing patterns (Vreeswijk and Sompolinsky, 1998; Destexhe et al., 1999; Brunel, 2000; Steriade, 2000; Renart et al., 2010; Dehghani et al., 2016; di Volo and Torcini, 2018). It was found that, during AI states, excitatory (E) and inhibitory (I) synaptic inputs are near-balanced (Dehghani et al., 2016), as predicted theoretically (Van Vreeswijk and Sompolinsky, 1996). In AI states, voltage fluctuations drive neurons over the threshold for firing action potentials, resulting in irregular

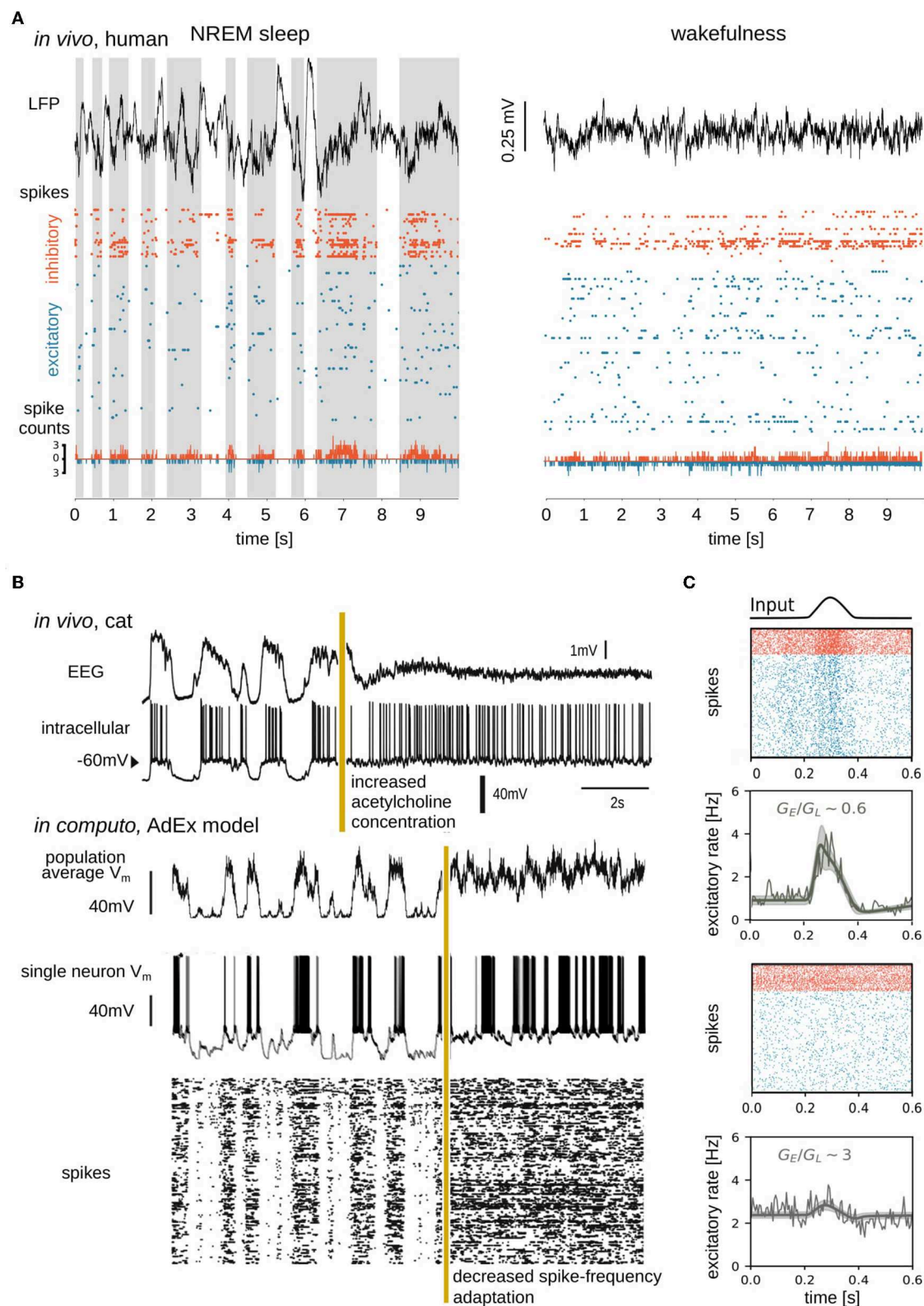


spiking dynamics, also known as fluctuation-driven regimes (Kuhn et al., 2004; Destexhe, 2007; Destexhe and Rudolph-Lilith, 2012). To understand mechanisms at work during fluctuation-driven dynamics, computational models have further shown that three parameters are important to capture neuronal responses in this regime, the average membrane voltage  $V_m$ , the amplitude of  $V_m$  fluctuations, and the conductance state of the membrane (Reig et al., 2015; Zerlaut et al., 2016).

Neuromodulators, including acetylcholine, play important biological roles in modulating the membrane properties of

neurons (McCormick, 1992) and thus transitions between AI and slow oscillatory dynamics through the regulation of membrane currents (Hill and Tononi, 2005). Neuromodulators are present at higher concentrations during conscious states (McCormick, 1992; Jones, 2003) and, most generally, inhibit potassium (activity-dependent and leak  $K^+$ ) channels, which leads to depolarization of cells and suppression of *spike-frequency adaptation*. At low neuromodulatory concentrations, during unconscious states,  $K^+$  leak channels are constitutively open and activity-dependent  $K^+$  channels open when neurons





**FIGURE 2 |** Simple, high-amplitude signals in unconscious brain states are associated with synchronous regular neuronal firing, whereas complex, low-amplitude signals in conscious brain states emerge from asynchronous irregular firing. **(A)** Data sample from Peyrache et al. (2012), Dehghani et al. (2016),

(Continued)



**FIGURE 2** | Le Van Quyen et al. (2016), Telericzuk et al. (2017), and Nghiem et al. (2018b), containing local field potential (LFP; top), spike times (action potentials; middle), and spike counts (bottom) recorded from a human subject during NREM sleep (left) and wakefulness (right). Spikes from inhibitory (orange) and excitatory (blue) neurons were separated and spike counts were calculated in bins of 5ms. Up states shaded in the left panel. **(B)** Transition between slow-wave (unconscious) and activated (conscious) state dynamics *in vivo* (top) and *in silico* (bottom). Experimentally the transition is generated by electrical stimulation of acetylcholine neurons in the pedunculopontine tegmentum (PPT) of anesthetized cat (Volgushev et al., 2011), triggering awake-like, desynchronized dynamics in cortex (Rudolph et al., 2005). A prominent consequence of enhancing cholinergic signaling in cortex is a reduction of spike-frequency adaptation (McCormick, 1992). *In silico*, a similarly desynchronizing effect can be generated by reducing the parameter responsible for spike-frequency adaptation. Simulated traces shown in the bottom were modified from Destexhe (2009), which used a network of adaptive exponential integrate-and-fire neurons. The average  $V_m$  of the network, the  $V_m$  of a randomly chosen neuron, and the raster plot of the network are shown. Reproduced with permission from Destexhe (2009). **(C)** State dependence of network responsiveness. The responsiveness of two spiking networks to a Gaussian pulse is shown. Raster plots display spike times of excitatory (blue) and inhibitory (orange) neurons connected by conductance-based synapses. Population activity (spike counts, thin line), as well as mean-field model (thick lines), and standard deviation (shaded area) of population firing rate generated by a mean-field model developed in di Volo et al. (2019). Responsiveness is found to vary between different network states, obtained by changing the ratio of the time-averaged global excitatory conductance ( $G_E$ ) (Destexhe et al., 2003) to membrane leakage conductance ( $G_L$ ).

spike, allowing  $K^+$  ions to exit cells thus hyperpolarizing the membrane. Accumulating self-inhibition in the form of spike-frequency adaptation during Up periods results in the transition to Down states. Conversely, spike-frequency adaptation wears off during Down states, allowing noise fluctuations (present ubiquitously; Destexhe and Rudolph-Lilith, 2012) to trigger transitions to Up states (Destexhe, 2009; Jercog et al., 2017; Nghiem T.-A. E. et al., 2018; di Volo et al., 2019) (**Figure 2B**). Computationally speaking, for high values of spike-frequency adaptation, bistability can be observed, with solutions at firing rate zero (Down state) and non-zero (Up state) values (Holcman and Tsodyks, 2006; di Volo et al., 2019). The more chaotic dynamics of AI states associated with consciousness allows for more reliable stimulus encoding (D'Andola et al., 2017), more reliable propagation (Zerlaut and Destexhe, 2017), and more sustained responses (Nghiem T.-A. E. et al., 2018) to stimuli over time. In contrast, during unconscious states, neuronal responses are more unreliable and vary greatly depending on the stimulus amplitude and whether cells receive inputs in Up or Down periods (Rosanova and Timofeev, 2005; Reig et al., 2015).

The Ising model for spin glasses (Jaynes, 1982) fitted to neural data (Schneidman et al., 2006) has revealed divergent types of emergent neuronal dynamics in conscious and unconscious states. While neuronal interactions are pairwise in wakefulness (Nghiem et al., 2017), coupling becomes population-wide in deep sleep (Tavoni et al., 2017; Nghiem et al., 2018b). In particular, inhibitory neurons organize synchronous activity across populations (Nghiem et al., 2018b; Zanoci et al., 2019), especially during deep sleep (Peyrache et al., 2012; Le Van Quyen et al., 2016; Olcese et al., 2016) where inhibitory neurons regulate rhythms of slow wave dynamics (Compte et al., 2008; Funk et al., 2017; Zucca et al., 2017, 2019).

To summarize, between unconscious and conscious brain states, microscopic data appear intuitively related to macroscopic data: synchronous microscopic Up and Down states resulting from constitutive and activity-dependent, hyperpolarizing currents due to reduced neuromodulation correspond to relatively simple, high-amplitude macroscopic dynamics observed in unconscious states. Active, disorganized, desynchronized, AI, low adaptation, high neuromodulation conditions correspond to low amplitude, complex, conscious brain signals. On backgrounds of differing spontaneous dynamics, generalizable patterns of activity (a.k.a. neural

graphoelements) are observed. Cash et al. have elegantly shown that K-complexes (graphoelements characteristic of sleep stage 2) are complementarily observed both at microscopic and macroscopic scales (Cash et al., 2009). Other identifiable patterns also begin to emerge in empirical and theoretical data, including phase cones (Freeman and Barrie, 2000) and interacting traveling waves (Sato et al., 2012; Chemla et al., 2019). Since statistical physics has successfully described neuronal interactions for different brain states, we ask next whether mesoscale methods from statistical physics can help represent spontaneous and evoked dynamics of neuronal populations, thus formally linking knowledge between micro- and macroscopic scales.

## MESOSCALE BRIDGES; POPULATIONS OF NEURONS

Brain dynamics at mesoscopic scales, describing thousands of neurons, are investigated empirically by electrophysiology and more recently, voltage-sensitive dyes (Arieli et al., 1996; Chemla and Chavane, 2010). At mesoscales, brain activity follows the trend of increasing complexity of spontaneous activity with consciousness (**Figure 2A**). Studying the effects of inputs at the mesoscale, studies have shown that perturbations during deep sleep states induce slow waves, but, during waking states, perturbations can result in chains of phase-locked activity (Pigorini et al., 2015) leading to causal global interactions (Rosanova et al., 2018).

*Mean-field* models offer a formalism for scaling up microscopic detail to collective macroscopic dynamics using few equations, offering a computational advantage for simulations. In describing states of matter, mean-field models simplify the probabilistic behavior of molecules to the relatively more predictable behavior of macroscopic states (Kadanoff, 2009). A rich literature has begun to develop mean-field models of neuronal populations, showing that global variables describing population activity can be usefully derived from the biophysics of neurons and their interactions (Ohira and Cowan, 1993; Ginzburg and Sompolinsky, 1994; El Boustani and Destexhe, 2009; Buice et al., 2010; Dahmen et al., 2016). Mean-field models have qualitatively reproduced temporal features of spontaneous dynamics including AI (El Boustani and Destexhe, 2009), Up and Down dynamics (Compte et al., 2003; Jercog et al., 2017;

Tartaglia and Brunel, 2017; di Volo et al., 2019), and transitions between these states (di Volo et al., 2019; Tort-Colet et al., 2019). In addition, connecting mean-fields provides a tool for simulating the propagation of patterns through time and space, across mesoscale structures. For example, recent work deriving mean-field models of networks with conductance-based synapses has reproduced the suppressive interaction between traveling waves observed in visual cortex during conscious states, a biological phenomenon that could not be captured by current-based networks (Chemla et al., 2019).

Mean-field models have highlighted that, while complicated to apply mathematically in the framework of conductance-based models (di Volo et al., 2019), voltage-dependent interactions constitute a significant non-linearity in the membrane evolution equations. Voltage-dependent interactions appear to be important for explaining non-trivial responses of biological neurons, through the mean and fluctuations of the cells' membrane voltage (Reig et al., 2015). In fact, while these results do not imply that differences in responsiveness are due only to conductances, they show that voltage dependent synapses play a role in the nonlinear state-dependent response of a neural network. As shown in **Figure 2C**, various levels of membrane conductance, regulated by voltage-dependent synapses, are shown to differently shape population responses.

Finally, renormalization group theory, a method of coarse-graining microscopic detail to obtain macroscopic laws helping to understand how order can emerge from apparent disorder (Wilson, 1979; Cardy, 1996; Goldenfeld, 2018) has recently begun to be applied to neural assemblies (Meshulam et al., 2019), laying further foundation for the formal connection of our understanding of brain function across scales.

## DISCUSSION

In this paper, we briefly reviewed work on the measurement and modeling of brain states at different scales, from single neurons to cell assemblies and global brain activity, considering both spontaneous and evoked dynamics. In particular, we highlighted that increased complexity in the dynamics of conscious brain states relates to changes in single-neuron biophysics, tuned by neuromodulation. In unconscious states, reduced neuromodulation promotes activity-dependent self-inhibition of excitatory neurons as they spike, leading to alternating, synchronous transients of silence and firing, that produce high-amplitude, low-complexity, synchronous signals, on resonant frequencies of the structural connectome. During conscious states, neuronal discharges are asynchronous, irregular and fluctuation-driven, resulting from sustained membrane depolarization in cortical neurons, promoting effective neural communication.

Beyond conscious and unconscious categories proposed here for the sake of brevity, important differences exist within categories of unconscious and conscious states (Brown et al., 2010; El Boustani and Destexhe, 2010; Nghiem et al., 2018a). Unlike healthy wakefulness and sleep, epileptic

networks display both excessively high conductance and strongly synchronized, regular dynamics (El Boustani and Destexhe, 2010). Further, brain signals in coma are both low-amplitude and low-complexity, in contrast to high-amplitude signals observed in other unconscious states, but also to complex signals observed in conscious states (El Boustani and Destexhe, 2010). Such anomalous deviations from the overall trend of coordinated changes in complexity and amplitude may illuminate mechanisms underlying disease-causing deviations from healthy brain states (Mackey and Glass, 1977).

To characterize brain states, it has been useful to consider not only spontaneous dynamics but also patterns evoked by perturbations. It was found that macroscopic responsiveness highly depends on brain state and different patterns of responses are evoked in conscious versus unconscious states (Massimini et al., 2005). Such state-dependent responsiveness can also be seen at the level of local networks *in vivo* and *in silico*, for example in the different reliability of responses to perturbations given during Up and Down periods of slow waves (Reig et al., 2015; Zerlaut and Destexhe, 2017). In simulations, different responsiveness could be accounted for by three parameters: membrane voltage, voltage fluctuation amplitude, and membrane conductance (Reig et al., 2015). These parameters could be well described by mean-field models (di Volo et al., 2019), able to capture fundamental properties of spontaneous dynamics and also state-dependent responses at mesoscales. As such, the data-driven coupling of such mean-field models may serve as natural candidates for modeling the emergence of mesoscopic and macroscopic-scale patterns.

Transient collapses of dimensionality found in encoding networks were also discussed as substrates potentially supporting neural codes. Such collapses in complexity have been observed in active ensembles at scales spanning microscopic (Churchland et al., 2010; Fairhall, 2019) to macroscopic (Quiroga et al., 2001; Zang et al., 2004) activity. This echoes recent work studying recordings of neural populations which highlighted that neural representations of stimuli may lie on low-dimensional manifolds (Churchland et al., 2012; Sadtler et al., 2014; Gallego et al., 2017; Zhao and Park, 2017; Golub et al., 2018; Chaudhuri et al., 2019; Recanatesi et al., 2019; Stringer et al., 2019). Indeed neurons do not fire independently, which would yield dynamics of dimensionality as high as the number of neurons, but instead follow constrained trajectories of activity that can be captured by descriptions of much lower dimensionality that depend on spontaneous and evoked dynamics. For example, a neural population firing in synchrony could be fully described by a periodic orbit trajectory constrained to a low-dimensional space (Churchland et al., 2012). Since spontaneous global network activity increases in dimensionality during conscious states, we ask whether the transient collapse of complexity in specific networks, translating the emergence of simpler dynamical structures from disorder, may be associated to neural codes.

As an analogy, windmills facing all in one direction display low complexity, but can only be synchronously active or inactive. Windmills facing in random directions,

in contrast, are a higher complexity configuration able to represent wind from any direction through the activity of a subset. The activity of an ensemble of windmills tuned to a particular direction of wind could represent a collapse of complexity and the generation of information by that subset (in this case, about the direction of wind). Similarly, enhanced dimensionality associated with conscious states could subserve neural information through the collapse of complexity in neural assemblies tuned to encode particular representations.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## REFERENCES

- Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* 273, 1868–1871. doi: 10.1126/science.273.5283.1868
- Atasoy, S., Donnelly, I., and Pearson, J. (2016). Human brain networks function in connectome-specific harmonic waves. *Nat. Commun.* 7:10340. doi: 10.1038/ncomms10340
- Boly, M., Phillips, C., Tshibanda, L., Vanhaudenhuyse, A., Schabus, M., Dang-Vu, T. T., et al. (2008). Intrinsic brain activity in altered states of consciousness: how conscious is the default mode of brain function? *Ann. N. Y. Acad. Sci.* 1129:119. doi: 10.1196/annals.1417.015
- Brookes, M. J., Woolrich, M., Luckhoo, H., Price, D., Hale, J. R., Stephenson, M. C., et al. (2011). Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16783–16788. doi: 10.1073/pnas.1112685108
- Brown, E. N., Lydic, R., and Schiff, N. D. (2010). General anesthesia, sleep, and coma. *New Engl. J. Med.* 363, 2638–2650. doi: 10.1056/NEJMra0808281
- Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* 8, 183–208. doi: 10.1023/A:1008925309027
- Buice, M. A., Cowan, J. D., and Chow, C. C. (2010). Systematic fluctuation expansion for neural network activity equations. *Neural Comput.* 22, 377–426. doi: 10.1162/neco.2009.02-09-960
- Capone, C., Rebollo, B., Muñoz, A., Illa, X., Del Giudice, P., Sanchez-Vives, M. V., et al. (2017). Slow waves in cortical slices: how spontaneous activity is shaped by laminar structure. *Cereb. Cortex* 29, 319–335. doi: 10.1093/cercor/bhx326
- Cardy, J. (1996). *Scaling and Renormalization in Statistical Physics. Vol. 5*. Cambridge: Cambridge University Press.
- Cash, S. S., Halgren, E., Dehghani, N., Rossetti, A. O., Thesen, T., Wang, C., et al. (2009). The human k-complex represents an isolated cortical down-state. *Science* 324, 1084–1087. doi: 10.1126/science.1169626
- Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A., and Fiete, I. (2019). The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat. Neurosci.* 22, 1512–1520. doi: 10.1038/s41593-019-0460-x
- Chemla, S., and Chavane, F. (2010). Voltage-sensitive dye imaging: technique review and models. *J. Physiol.* 104, 40–50. doi: 10.1016/j.jphysparis.2009.11.009
- Chemla, S., Reynaud, A., di Volo, M., Zerlaut, Y., Perrinet, L., Destexhe, A., et al. (2019). Suppressive traveling waves shape representations of illusory motion in primary visual cortex of awake primate. *J. Neurosci.* 39, 4282–4298. doi: 10.1523/JNEUROSCI.2792-18.2019
- Churchland, M. M., Byron, M. Y., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* 13, 369–378. doi: 10.1038/nn.2501
- Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., et al. (2012). Neural population dynamics during reaching. *Nature* 487, 51–56. doi: 10.1038/nature11129
- Compte, A., Reig, R., Descalzo, V. F., Harvey, M. A., Puccini, G. D., and Sanchez-Vives, M. V. (2008). Spontaneous high-frequency (10–80 Hz) oscillations during up states in the cerebral cortex *in vitro*. *J. Neurosci.* 28, 13828–13844. doi: 10.1523/JNEUROSCI.2684-08.2008
- Compte, A., Sanchez-Vives, M. V., McCormick, D. A., and Wang, X.-J. (2003). Cellular and network mechanisms of slow oscillatory activity (< 1 Hz) and wave propagations in a cortical network model. *J. Neurophysiol.* 89, 2707–2725. doi: 10.1152/jn.00845.2002
- Crunelli, V., David, F., Lőrincz, M. L., and Hughes, S. W. (2015). The thalamocortical network as a single slow wave-generating unit. *Curr. Opin. Neurobiol.* 31, 72–80. doi: 10.1016/j.conb.2014.09.001
- Dahmen, D., Bos, H., and Helias, M. (2016). Correlated fluctuations in strongly coupled binary networks beyond equilibrium. *Phys. Rev. X* 6:031024. doi: 10.1103/PhysRevX.6.031024
- D'Andola, M., Rebollo, B., Casali, A. G., Weinert, J. F., Pigorini, A., Villa, R., et al. (2017). Bistability, causality, and complexity in cortical networks: an *in vitro* perturbational study. *Cereb. Cortex* 28, 2233–2242. doi: 10.1093/cercor/bhx122
- David, F., Schmiedt, J. T., Taylor, H. L., Orban, G., Di Giovanni, G., Uebele, V. N., et al. (2013). Essential thalamic contribution to slow waves of natural sleep. *J. Neurosci.* 33, 19599–19610. doi: 10.1523/JNEUROSCI.3169-13.2013
- Dehghani, N., Peyrache, A., Telenczuk, B., Le Van Quyen, M., Halgren, E., Cash, S. S., et al. (2016). Dynamic balance of excitation and inhibition in human and monkey neocortex. *Sci. Rep.* 6:23176. doi: 10.1038/srep23176
- Destexhe, A. (2007). High-conductance state. *Scholarpedia* 2:1341. doi: 10.4249/scholarpedia.1341
- Destexhe, A. (2009). Self-sustained asynchronous irregular states and up-down states in thalamic, cortical and thalamocortical networks of nonlinear integrate-and-fire neurons. *J. Comput. Neurosci.* 27, 493–506. doi: 10.1007/s10827-009-0164-4
- Destexhe, A., Contreras, D., and Steriade, M. (1999). Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. *J. Neurosci.* 19, 4595–4608. doi: 10.1523/JNEUROSCI.19-11-04595.1999
- Destexhe, A., Hughes, S. W., Rudolph, M., and Crunelli, V. (2007). Are corticothalamic 'up' states fragments of wakefulness? *Trends Neurosci.* 30, 334–342. doi: 10.1016/j.tins.2007.04.006
- Destexhe, A., Rudolph, M., and Paré, D. (2003). The high-conductance state of neocortical neurons *in vivo*. *Nat. Rev. Neurosci.* 4, 739–751. doi: 10.1038/nrn1198
- Destexhe, A., and Rudolph-Lilith, M. (2012). *Neuronal Noise*. New York, NY: Springer.
- di Volo, M., Romagnoni, A., Capone, C., and Destexhe, A. (2019). Biologically realistic mean-field models of conductance-based networks of spiking neurons with adaptation. *Neural Comput.* 31, 653–680. doi: 10.1162/neco\_a\_01173

## FUNDING

This research was supported by the Centre National de la Recherche Scientifique (CNRS), the European Community (Human Brain Project, H2020-785907), and by École des Neurosciences de Paris (ENP).

## ACKNOWLEDGMENTS

The authors would like to thank Simone Blanco Malerba, Cristiano Capone, Stephen E. Clarke, Damien Depannemaeker, Anton Filipchuk, Enrique Hansen, J. N. K. Jaynes, Vicente Medel, Archibald Parsons, Mattias Peuvrier, Wolf Singer, Kat Uesat, Bahar Hazal Yalçınkaya, and Yann Zerlaut for useful discussion of the manuscript.



- di Volo, M., and Torcini, A. (2018). Transition from asynchronous to oscillatory dynamics in balanced spiking networks with instantaneous synapses. *Phys. Rev. Lett.* 121:128301. doi: 10.1103/PhysRevLett.121.128301
- Duncan-Johnson, C. C., and Donchin, E. (1982). The p300 component of the event-related brain potential as an index of information processing. *Biol. Psychol.* 14, 1–52. doi: 10.1016/0301-0511(82)90016-3
- El Boustani, S., and Destexhe, A. (2009). A master equation formalism for macroscopic modeling of asynchronous irregular activity states. *Neural Comput.* 21, 46–100. doi: 10.1162/neco.2009.02-08-710
- El Boustani, S., and Destexhe, A. (2010). Brain dynamics at multiple scales: can one reconcile the apparent low-dimensional chaos of macroscopic variables with the seemingly stochastic behavior of single neurons? *Int. J. Bifur. Chaos* 20, 1687–1702. doi: 10.1142/S0218127410026769
- Engemann, D. A., Raimondo, F., King, J.-R., Rohaut, B., Louppe, G., Faugeras, F., et al. (2018). Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain* 141, 3179–3192. doi: 10.1093/brain/awy251
- Fairhall, A. L. (2019). Whither variability? *Nat. Neurosci.* 22, 329–330. doi: 10.1038/s41593-019-0344-0
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102
- Freeman, W. J., and Barrie, J. M. (2000). Analysis of spatial patterns of phase in neocortical gamma EEGs in rabbit. *J. Neurophysiol.* 84, 1266–1278. doi: 10.1152/jn.2000.84.3.1266
- Funk, C. M., Peelman, K., Bellesi, M., Marshall, W., Cirelli, C., and Tononi, G. (2017). Role of somatostatin-positive cortical interneurons in the generation of sleep slow waves. *J. Neurosci.* 37, 9132–9148. doi: 10.1523/JNEUROSCI.1303-17.2017
- Gallego, J. A., Perich, M. G., Miller, L. E., and Solla, S. A. (2017). Neural manifolds for the control of movement. *Neuron* 94, 978–984. doi: 10.1016/j.neuron.2017.05.025
- Gilson, M., Deco, G., Friston, K. J., Hagmann, P., Mantini, D., Betti, V., et al. (2018). Effective connectivity inferred from fMRI transition dynamics during movie viewing points to a balanced reconfiguration of cortical interactions. *Neuroimage* 180, 534–546. doi: 10.1016/j.neuroimage.2017.09.061
- Ginzburg, I., and Sompolinsky, H. (1994). Theory of correlations in stochastic neural networks. *Phys. Rev. E* 50, 3171–3191. doi: 10.1103/PhysRevE.50.3171
- Goldenfeld, N. (2018). *Lectures on Phase Transitions and the Renormalization Group*. Boca Raton, FL: CRC Press.
- Golub, M. D., Sadtler, P. T., Oby, E. R., Quick, K. M., Ryu, S. I., Tyler-Kabara, E. C., et al. (2018). Learning by neural reassociation. *Nat. Neurosci.* 21, 607–616. doi: 10.1038/s41593-018-0095-3
- Goodin, D. S., and Aminoff, M. J. (1984). The relationship between the evoked potential and brain events in sensory discrimination and motor response. *Brain* 107, 241–251. doi: 10.1093/brain/107.1.241
- Hill, S., and Tononi, G. (2005). Modeling sleep and wakefulness in the thalamocortical system. *J. Neurophysiol.* 93, 1671–1698. doi: 10.1152/jn.00915.2004
- Holcman, D., and Tsodyks, M. (2006). The emergence of up and down states in cortical networks. *PLoS Comput. Biol.* 2:e23. doi: 10.1371/journal.pcbi.0020023
- Jaynes, E. T. (1982). On the rationale of maximum-entropy methods. *Proc. IEEE* 70, 939–952. doi: 10.1109/PROC.1982.12425
- Jercog, D., Roxin, A., Bartho, P., Luczak, A., Compte, A., and de la Rocha, J. (2017). Up-down cortical dynamics reflect state transitions in a bistable network. *Elife* 6:e22425. doi: 10.7554/eLife.22425
- Jones, B. E. (2003). Arousal systems. *Front. Biosci.* 8:438–451. doi: 10.2741/1074
- Kadanoff, L. P. (2009). More is the same; phase transitions and mean field theories. *J. Stat. Phys.* 137:777. doi: 10.1007/s10955-009-9814-1
- Kuhn, A., Aertsen, A., and Rotter, S. (2004). Neuronal integration of synaptic input in the fluctuation-driven regime. *J. Neurosci.* 24, 2345–2356. doi: 10.1523/JNEUROSCI.3349-03.2004
- Le Van Quyen, M., Muller, L. E., Telenczuk, B., Halgren, E., Cash, S., Hatsopoulos, N. G., et al. (2016). High-frequency oscillations in human and monkey neocortex during the wake-sleep cycle. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9363–9368. doi: 10.1073/pnas.1523583113
- Luck, S. J., and Kappenman, E. S. (2011). *The Oxford Handbook of Event-Related Potential Components*. Oxford: Oxford University Press.
- Mackey, M. C., and Glass, L. (1977). Oscillation and chaos in physiological control systems. *Science* 197, 287–289. doi: 10.1126/science.267326
- Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. doi: 10.1126/science.1117256
- Massimini, M., Huber, R., Ferrarelli, F., Hill, S., and Tononi, G. (2004). The sleep slow oscillation as a traveling wave. *J. Neurosci.* 24, 6862–6870. doi: 10.1523/JNEUROSCI.1318-04.2004
- McCormick, D. A. (1992). Neurotransmitter actions in the thalamus and cerebral cortex and their role in neuromodulation of thalamocortical activity. *Prog. Neurobiol.* 39, 337–388. doi: 10.1016/0301-0082(92)90012-4
- Meshulam, L., Gauthier, J. L., Brody, C. D., Tank, D. W., and Bialek, W. (2019). Coarse graining, fixed points, and scaling in a large population of neurons. *Phys. Rev. Lett.* 123:178103. doi: 10.1103/PhysRevLett.123.178103
- Nghiem, T.-A., Lina, J.-M., di Volo, M., Capone, C., Evans, A. C., Destexhe, A., et al. (2018a). State equation from the spectral structure of human brain activity. *arXiv [preprint]*. arXiv:1806.07365.
- Nghiem, T.-A., Marre, O., Destexhe, A., and Ferrari, U. (2017). “Pairwise ising model analysis of human cortical neuron recordings,” in *International Conference on Geometric Science of Information* (New York, NY: Springer), 257–264. doi: 10.1007/978-3-319-68445-1\_30
- Nghiem, T.-A., Telenczuk, B., Marre, O., Destexhe, A., and Ferrari, U. (2018b). Maximum-entropy models reveal the excitatory and inhibitory correlation structures in cortical neuronal activity. *Phys. Rev. E* 98:012402. doi: 10.1103/PhysRevE.98.012402
- Nghiem, T.-A. E., Tort-Colet, N., Gorski, T., Ferrari, U., Moghimyifiroozabad, S., Goldman, J. S., et al. (2018). Cholinergic switch between two different types of slow waves in cerebral cortex. *bioRxiv [Preprint]*. doi: 10.1101/430405
- Niedermeyer, E., and Lopes da Silva, F. H. (2005). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Philadelphia, PA: Lippincott Williams & Wilkins.
- Ohira, T., and Cowan, J. D. (1993). Master-equation approach to stochastic neurodynamics. *Phys. Rev. E* 48:2259. doi: 10.1103/physreve.48.2259
- Olcese, U., Bos, J. J., Vinck, M., Lankelma, J. V., van Mourik-Donga, L. B., Schlumm, F., et al. (2016). Spike-based functional connectivity in cerebral cortex and hippocampus: loss of global connectivity is coupled to preservation of local connectivity during non-REM sleep. *J. Neurosci.* 36, 7676–7692. doi: 10.1523/JNEUROSCI.4201-15.2016
- Peyrache, A., Dehghani, N., Eskandar, E. N., Madsen, J. R., Anderson, W. S., Donoghue, J. A., et al. (2012). Spatiotemporal dynamics of neocortical excitation and inhibition during human sleep. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1731–1736. doi: 10.1073/pnas.1109895109
- Pigorini, A., Sarasso, S., Proserpio, P., Szymanski, C., Arnulfo, G., Casarotto, S., et al. (2015). Bistability breaks-off deterministic responses to intracortical stimulation during non-rem sleep. *Neuroimage* 112, 105–113. doi: 10.1016/j.neuroimage.2015.02.056
- Poulet, J. F., Fernandez, L. M., Crochet, S., and Petersen, C. C. (2012). Thalamic control of cortical states. *Nat. Neurosci.* 15, 370–372. doi: 10.1038/nn.3035
- Quiroga, R. Q., Rosso, O. A., Başar, E., and Schürmann, M. (2001). Wavelet entropy in event-related potentials: a new method shows ordering of EEG oscillations. *Biol. Cybern.* 84, 291–299. doi: 10.1007/s004220000212
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682. doi: 10.1073/pnas.98.2.676
- Recanatesi, S., Ocker, G. K., Buice, M. A., and Shea-Brown, E. (2019). Dimensionality in recurrent spiking networks: global trends in activity and local origins in connectivity. *PLoS Comput. Biol.* 15:e1006446. doi: 10.1371/journal.pcbi.1006446
- Reig, R., Zerlaut, Y., Vergara, R., Destexhe, A., and Sanchez-Vives, M. V. (2015). Gain modulation of synaptic inputs by network state in auditory cortex *in vivo*. *J. Neurosci.* 35, 2689–2702. doi: 10.1523/JNEUROSCI.2004-14.2015
- Renart, A., De La Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., et al. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590. doi: 10.1126/science.1179850
- Rosanova, M., Fecchio, M., Casarotto, S., Sarasso, S., Casali, A., Pigorini, A., et al. (2018). Sleep-like cortical off-periods disrupt causality and complexity in the brain of unresponsive wakefulness syndrome patients. *Nat. Commun.* 9:4427. doi: 10.1038/s41467-018-06871-1



- Rosanov, M., and Timofeev, I. (2005). Neuronal mechanisms mediating the variability of somatosensory evoked potentials during sleep oscillations in cats. *J. Physiol.* 562, 569–582. doi: 10.1113/jphysiol.2004.071381
- Rudolph, M., Pelletier, J.-G., Paré, D., and Destexhe, A. (2005). Characterization of synaptic conductances and integrative properties during electrically-induced EEG-activated states in neocortical neurons *in vivo*. *J. Neurophysiol.* 94, 2805–2821. doi: 10.1152/jn.01313.2004
- Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., et al. (2014). Neural constraints on learning. *Nature* 512, 423–426. doi: 10.1038/nature13665
- Sanchez-Vives, M. V., Massimini, M., and Mattia, M. (2017). Shaping the default activity pattern of the cortical network. *Neuron* 94, 993–1001. doi: 10.1016/j.neuron.2017.05.015
- Sanchez-Vives, M. V., and McCormick, D. A. (2000). Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nat. Neurosci.* 3, 1027–1034. doi: 10.1038/79848
- Sato, T. K., Nauhaus, I., and Carandini, M. (2012). Traveling waves in visual cortex. *Neuron* 75, 218–229. doi: 10.1016/j.neuron.2012.06.029
- Schneidman, E., Berry II, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012. doi: 10.1038/nature04701
- Singer, W. (2013). Cortical dynamics revisited. *Trends Cogn. Sci.* 17, 616–626. doi: 10.1016/j.tics.2013.09.006
- Sitt, J. D., King, J.-R., El Karoui, I., Rohaut, B., Faugeras, F., Gramfort, A., et al. (2014). Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* 137, 2258–2270. doi: 10.1093/brain/awu141
- Skarda, C. A., and Freeman, W. J. (1987). How brains make chaos in order to make sense of the world. *Behav. Brain Sci.* 10, 161–173. doi: 10.1017/S0140525X00047336
- Steriade, M. (2000). Corticothalamic resonance, states of vigilance and mentation. *Neuroscience* 101, 243–276. doi: 10.1016/S0306-4522(00)00353-5
- Steriade, M. (2003). *Neuronal Substrates of Sleep and Epilepsy*. Cambridge: Cambridge University Press.
- Steriade, M., Nunez, A., and Amzica, F. (1993). A novel slow (< 1 Hz) oscillation of neocortical neurons *in vivo*: depolarizing and hyperpolarizing components. *J. Neurosci.* 13, 3252–3265. doi: 10.1523/JNEUROSCI.13-08-03252.1993
- Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature* 571, 361–365. doi: 10.1038/s41586-019-1346-5
- Sur, S., and Sinha, V. (2009). Event-related potential: an overview. *Ind. Psychiatry J.* 18, 70–73. doi: 10.4103/0972-6748.57865
- Tartaglia, E. M., and Brunel, N. (2017). Bistability and up/down state alternations in inhibition-dominated randomly connected networks of lif neurons. *Sci. Rep.* 7:11916. doi: 10.1038/s41598-017-12033-y
- Tavoni, G., Ferrari, U., Battaglia, F. P., Cocco, S., and Monasson, R. (2017). Functional coupling networks inferred from prefrontal cortex activity show experience-related effective plasticity. *Netw. Neurosci.* 1, 275–301. doi: 10.1162/NETN\_a\_00014
- Teleńczuk, B., Dehghani, N., Le Van Quyen, M., Cash, S. S., Halgren, E., Hatsopoulos, N. G., et al. (2017). Local field potentials primarily reflect inhibitory neuron activity in human and monkey cortex. *Sci. Rep.* 7:40211. doi: 10.1038/srep40211
- Timofeev, I., Grenier, F., Bazhenov, M., Sejnowski, T., and Steriade, M. (2000). Origin of slow cortical oscillations in deafferented cortical slabs. *Cereb. Cortex* 10, 1185–1199. doi: 10.1093/cercor/10.12.1185
- Tononi, G., and Edelman, G. M. (1998). Consciousness and complexity. *Science* 282, 1846–1851. doi: 10.1126/science.282.5395.1846
- Tort-Colet, N., Capone, C., Sanchez-Vives, M. V., and Mattia, M. (2019). Attractor competition enriches cortical dynamics during awakening from anesthesia. *bioRxiv [Preprint]*. doi: 10.1101/517102
- Tsodyks, M. V., and Sejnowski, T. (1995). Rapid state switching in balanced cortical network models. *Network* 6, 111–124. doi: 10.1088/0954-898X\_6\_2\_001
- Uhlhaas, P., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikolić, D., et al. (2009). Neural synchrony in cortical networks: history, concept and current status. *Front. Integr. Neurosci.* 3:17. doi: 10.3389/neuro.07.017.2009
- Van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726. doi: 10.1126/science.274.5293.1724
- Volgushev, M., Chauvette, S., and Timofeev, I. (2011). “Long-range correlation of the membrane potential in neocortical neurons during slow oscillation,” in *Progress in Brain Research, Vol. 193* (Amsterdam: Elsevier), 181–199.
- Vreeswijk, C. V., and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput.* 10, 1321–1371. doi: 10.1162/089976698300017214
- Wilson, K. G. (1979). Problems in physics with many scales of length. *Sci. Am.* 241, 158–179. doi: 10.1038/scientificamerican0879-158
- Zang, Y., Jiang, T., Lu, Y., He, Y., and Tian, L. (2004). Regional homogeneity approach to fMRI data analysis. *Neuroimage* 22, 394–400. doi: 10.1016/j.neuroimage.2003.12.030
- Zanoci, C., Dehghani, N., and Tegmark, M. (2019). Ensemble inhibition and excitation in the human cortex: an ising-model analysis with uncertainties. *Phys. Rev. E* 99:032408. doi: 10.1103/PhysRevE.99.032408
- Zerlaut, Y., and Destexhe, A. (2017). Enhanced responsiveness and low-level awareness in stochastic network states. *Neuron* 94, 1002–1009. doi: 10.1016/j.neuron.2017.04.001
- Zerlaut, Y., Teleńczuk, B., Deleuze, C., Bal, T., Ouanounou, G., and Destexhe, A. (2016). Heterogeneous firing rate response of mouse layer V pyramidal neurons in the fluctuation-driven regime. *J. Physiol.* 594, 3791–3808. doi: 10.1113/JP272317
- Zhao, Y., and Park, I. M. (2017). Variational latent gaussian process for recovering single-trial dynamics from population spike trains. *Neural Comput.* 29, 1293–1316. doi: 10.1162/NECO\_a\_00953
- Zucca, S., D’Urso, G., Pasquale, V., Vecchia, D., Pica, G., Bovetti, S., et al. (2017). An inhibitory gate for state transition in cortex. *Elife* 6:e26177. doi: 10.7554/eLife.26177
- Zucca, S., Pasquale, V., de Leon Roig, P. L., Panzeri, S., and Fellin, T. (2019). Thalamic drive of cortical parvalbumin-positive interneurons during down states in anesthetized mice. *Curr. Biol.* 29, 1481–1490. doi: 10.1016/j.cub.2019.04.007

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Goldman, Tort-Colet, di Volo, Susin, Bouté, Dali, Carlu, Nghiem, Górski and Destexhe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Principles of Neurorehabilitation After Stroke Based on Motor Learning and Brain Plasticity Mechanisms

Martina Maier<sup>1</sup>, Belén Rubio Ballester<sup>1</sup> and Paul F. M. J. Verschure<sup>1,2\*</sup>

<sup>1</sup> Laboratory of Synthetic, Perceptive, Emotive and Cognitive Systems, Institute for Bioengineering of Catalonia, The Barcelona Institute of Science and Technology, Barcelona, Spain, <sup>2</sup> Institutio Catalana de Recerca i Estudis Avançats, Barcelona, Spain

## OPEN ACCESS

### Edited by:

Olivia Gosseries,  
University of Liège, Belgium

### Reviewed by:

Michelle Hook,  
Texas A&M University, United States  
Stephan Patrick Swinnen,  
KU Leuven, Belgium

### \*Correspondence:

Paul F. M. J. Verschure  
pverschure@ibecbarcelona.eu

**Received:** 31 March 2019

**Accepted:** 19 November 2019

**Published:** 17 December 2019

### Citation:

Maier M, Ballester BR and  
Verschure PFMJ (2019) Principles  
of Neurorehabilitation After Stroke  
Based on Motor Learning and Brain  
Plasticity Mechanisms.  
*Front. Syst. Neurosci.* 13:74.  
doi: 10.3389/fnsys.2019.00074

What are the principles underlying effective neurorehabilitation? The aim of neurorehabilitation is to exploit interventions based on human and animal studies about learning and adaptation, as well as to show that the activation of experience-dependent neuronal plasticity augments functional recovery after stroke. Instead of teaching compensatory strategies that do not reduce impairment but allow the patient to return home as soon as possible, functional recovery might be more sustainable as it ensures a long-term reduction in impairment and an improvement in quality of life. At the same time, neurorehabilitation permits the scientific community to collect valuable data, which allows inferring about the principles of brain organization. Hence neuroscience sheds light on the mechanisms of learning new functions or relearning lost ones. However, current rehabilitation methods lack the exact operationalization of evidence gained from skill learning literature, leading to an urgent need to bridge motor learning theory and present clinical work in order to identify a set of ingredients and practical applications that could guide future interventions. This work aims to unify the neuroscientific literature relevant to the recovery process and rehabilitation practice in order to provide a synthesis of the principles that constitute an effective neurorehabilitation approach. Previous attempts to achieve this goal either focused on a subset of principles or did not link clinical application to the principles of motor learning and recovery. We identified 15 principles of motor learning based on existing literature: massed practice, spaced practice, dosage, task-specific practice, goal-oriented practice, variable practice, increasing difficulty, multisensory stimulation, rhythmic cueing, explicit feedback/knowledge of results, implicit feedback/knowledge of performance, modulate effector selection, action observation/embodied practice, motor imagery, and social interaction. We comment on trials that successfully implemented these principles and report evidence from experiments with healthy individuals as well as clinical work.

**Keywords:** neurorehabilitation, motor learning, plasticity, stroke, principles

**Abbreviations:** ADLs, activities of daily living; CIMT, constraint-induced movement therapy; fMRI, functional magnetic resonance imaging; KR, knowledge of results; KP, knowledge of performance; LTP, long-term potentiation; PD, Parkinson's disease; TMS, transcranial magnetic stimulation; VR, virtual reality.

## INTRODUCTION

So far there is no clear understanding of the principles underlying effective neurorehabilitation approaches. Therapeutic protocols can be readily described by the following aspects: the body part trained (e.g., the legs), the tools or machines used for the training (e.g., a treadmill), the activity performed (e.g., walking), and when the therapy commences (e.g., during the acute phase after a stroke). However, an intervention typically includes more elements. For instance, the use of the less affected limb can be restricted, and the therapist can encourage the patient to spend more time exercising or give feedback about task performance. While some interventions, like CIMT, clearly define their active ingredients (Carter et al., 2010; Proffitt and Lange, 2015) that should lead to effective recovery (Kwakkel et al., 2015), most others do not. Neurorehabilitation research aims to find interventions that promote recovery and to establish whether the presence or absence of improvement can be explained by any neuronal changes that occur in the post-stroke brain (Dobkin, 2005). Neuroscience can help us to create interventions that lead to changes in the brain; however, with no clear understanding of what an intervention does, attributing causality remains difficult. One way to formalize an intervention is by breaking it into parts, studying the behavioral and neural effects of these parts, and deriving principles from them—in the case of stroke neurorehabilitation, these would be principles that optimize acquisition, retention, and generalization of skills.

While there are plenty of meta-analyses that look at training effectiveness in terms of individual body parts/functions, tools, or machines and activities (Langhorne et al., 2009; Veerbeek et al., 2014), the effect of experience remains much less clear in spite of attempts to formalize and identify the principles of neurorehabilitation. A review of principles of experience-dependent neural plasticity by Kleim and Jones (2008) explains why training is crucial for recovery. According to their work, neurorehabilitation presumes that exposure to specific training experiences leads to improvement of impairment by activating neural plasticity mechanisms. Consequently most of the work in the field focuses on the identification of scientifically grounded principles that should guide the design of these training experiences. In this vein, Kleim and Jones (2008) elaborated on five main principles of effective training experience — specificity, repetition, intensity, time, and salience — but offered little concrete applicability. Another synthesis addressed further principles (forced use, massed practice, spaced practice, task-oriented functional training, randomized training); however, the main focus of the review was on individual body functions, methods, or tools, providing a global view on rehabilitation strategies (Dobkin, 2004). Two meta-analyses investigated specific principles. One looked only at the principle of intensity and found that more therapy time did enhance functional recovery (Kwakkel, 2009). Another determined that repetition does improve upper and lower limb function (Thomas et al., 2017). However, both studies did not investigate the mechanisms that would lead to the effects observed. Similarly, a review that analyzed CIMT, which combines several principles in one method, gained interesting insights in its efficacy but did not

explain the results from a neuroscientific, mechanistic point of view (Kwakkel et al., 2015). The work by Levin et al. (2015), on the other hand, tried to link the principles of motor learning to the application of these principles in novel rehabilitation methods while offering some neuroscientific reasoning for doing so. Their review addresses the difficulty of the task, the organization of movement, movements to the contralateral workspace, visual cues and objects and the interaction with them, sensory feedback, feedback about performance and results, repetitions, variability, and motivation. However, the included motor control and motor learning principles were not well defined and therefore leave room for interpretation (Levin et al., 2015).

In a previous meta-analysis (Maier et al., 2019), we compiled a list of principles for neurorehabilitation based on literature on motor learning and recovery: massed practice, dosage, structured practice, task-specific practice, variable practice, multisensory stimulation, increasing difficulty, explicit feedback/knowledge of results, implicit feedback/knowledge of performance, movement representation, and promotion of the use of the affected limb. We then performed a content analysis to determine whether these principles were present in the clinical studies included in the review, but we did not provide an analysis of the principles identified. In this work, we aim to extend the number of principles found and, for each of them, unify the neuroscientific literature from human or animal studies on motor learning and comment on the observed neuronal effects. We also include evidence from clinical studies to show its effect in recovering functionality after stroke. Some principles already serve as building blocks of effective rehabilitation programs, e.g., CIMT (Kwakkel et al., 2015), Bobath (Kollen et al., 2009), enriched rehabilitation (Livingston-Thomas et al., 2016), VR-based rehabilitation (Laver et al., 2017), and exogenous or robotic interventions (Langhorne et al., 2011). However, transferring these principles into clinical practice faces the challenge of operationalizing them. We comment on these difficulties and the gaps between theory, evidence, and operationalization that we encountered. Consequently, this work can serve clinicians and researchers as a practical guide of principles to investigate further effective neurorehabilitation approaches.

## MATERIALS AND METHODS

In this conceptual analysis, the rehabilitation experience is broken down into individual parts that are termed principles of neurorehabilitation. They are principles because they are evidenced by experimental data, and together, they could form the foundation of a higher-order theoretical framework. As a first attempt, a list of 11 principles was compiled based on existing literature in a recent meta-analysis on the effectiveness of VR-based rehabilitation systems for stroke recovery (Maier et al., 2019). For the current work, the list has been revised, and additional principles have been identified through a computerized search in PubMed Central using the keywords “principles of motor learning,” “principles of recovery,” “principles of experience-dependent learning” and “principles of neurorehabilitation.” We restricted the search to the last

5 years to obtain currently used principles. We focused on reviews, perspectives, and debates around rehabilitation methods and interventions for stroke recovery and excluded articles that explained study protocols or clinical trials, prevention methods, pharmaceutical or medical interventions, or stroke taxonomies. The principles mentioned in each paper were compared with the original list and added if they were not present. Afterward, we summarized for each principle the historical background based on motor learning literature and its contribution to learning based on human or animal studies. Further, where available, neurological effects and clinical outcomes were included as well.

## IDENTIFICATION OF PRINCIPLES OF NEUROREHABILITATION

Our computerized search yielded 548 records, of which 74 were deemed adequate for further screening after we examined if their titles either contained any of the search terms or appeared to discuss post-stroke rehabilitation strategies. After analysis of their abstracts and full-texts, the principles mentioned in 17 articles were extracted. We excluded papers if their title or abstract reported or compared surgical or pharmaceutical interventions as well as if they discussed stroke taxonomies, proposed study protocols or clinical trials, covered principles unrelated to stroke and/or stroke rehabilitation itself (e.g., principles for disease prevention, pre- and post-operative care, care facilities, patient management, therapist education, nursing practice, dietary recommendation, veterinary etc.), or looked into patient or caregiver perception. The articles and reviews selected spawned various research fields in neurorehabilitation: Motor learning (Winstein et al., 2014), therapies [physical therapy (Veerbeek et al., 2014), upper limb immobilization (Furlan et al., 2016), environmental enrichment (Livingston-Thomas et al., 2016), aerobic training (Billinger, 2015; Hasan et al., 2016), CIMT (Kwakkel et al., 2015; Zhang et al., 2017), cognitive rehabilitation (Middleton and Schwartz, 2012), music therapy (Zhang et al., 2016)], tools and methods [hand robotics (Yue et al., 2017), VR (Darekar et al., 2015; Fu et al., 2015), neurofeedback (Renton et al., 2017)], and principles [dose and timing (Basso and Lang, 2017)]. Together with previously collated literature, we identified 15 principles.

The identified principles from the meta-analysis are as follows:

- **Massed practice/repetitive practice** (Middleton and Schwartz, 2012; Veerbeek et al., 2014; Fu et al., 2015; Kwakkel et al., 2015; Furlan et al., 2016; Zhang et al., 2016).
- **Spaced practice** (Middleton and Schwartz, 2012; Billinger, 2015; Hasan et al., 2016; Livingston-Thomas et al., 2016).
- **Dosage/duration** (Veerbeek et al., 2014; Winstein et al., 2014; Billinger, 2015; Darekar et al., 2015; Kwakkel et al., 2015; Hasan et al., 2016; Livingston-Thomas et al., 2016; Basso and Lang, 2017; Zhang et al., 2017).
- **Task-specific practice** (Veerbeek et al., 2014; Winstein et al., 2014; Fu et al., 2015; Kwakkel et al., 2015; Furlan et al., 2016; Livingston-Thomas et al., 2016; Yue et al., 2017).

- **Variable practice** (Darekar et al., 2015; Fu et al., 2015; Livingston-Thomas et al., 2016).
- **Increasing difficulty** (Winstein et al., 2014; Fu et al., 2015; Kwakkel et al., 2015; Furlan et al., 2016; Hasan et al., 2016; Livingston-Thomas et al., 2016; Zhang et al., 2017).
- **Multisensory stimulation** (Veerbeek et al., 2014; Livingston-Thomas et al., 2016; Yue et al., 2017).
- **Explicit feedback/knowledge of results** (Middleton and Schwartz, 2012; Veerbeek et al., 2014; Darekar et al., 2015; Fu et al., 2015; Renton et al., 2017).
- **Implicit feedback/knowledge of performance** (Veerbeek et al., 2014; Darekar et al., 2015; Fu et al., 2015; Zhang et al., 2016; Renton et al., 2017; Yue et al., 2017).
- **Modulate effector selection** (Veerbeek et al., 2014; Winstein et al., 2014; Kwakkel et al., 2015; Furlan et al., 2016; Zhang et al., 2017).
- **Action observation/embodied practice** (Veerbeek et al., 2014; Fu et al., 2015; Yue et al., 2017).

Additional principles encountered through the search:

- **Goal-oriented practice** (Winstein et al., 2014; Fu et al., 2015; Yue et al., 2017).
- **Rhythmic cueing** (Middleton and Schwartz, 2012; Veerbeek et al., 2014; Zhang et al., 2016).
- **Motor imagery/mental practice** (Veerbeek et al., 2014).
- **Social interaction** (Winstein et al., 2014; Fu et al., 2015; Livingston-Thomas et al., 2016; Zhang et al., 2017).

In the following sections, we summarize for each principle the theoretical background, the evidence for motor learning, and the clinical effectiveness. We also added studies that comment on the neurological changes observed after applying the principles in motor learning tasks. The detailed neurological changes reported by these studies can be found in **Table 1**.

## Massed Practice/Repetitive Practice

Massed practice was defined as work episodes with very brief to no rest periods (Schmidt and Lee, 2011). Within a work episode, a skill can be trained repeatedly in a *constant* or *blocked* fashion (Ammons, 1947; Mulligan et al., 1980). In the field of rehabilitation, the term describes the prolonged and repeated use of the more affected limb (Taub et al., 1999). Theoretically, learning through repetitions can speed-up the shaping of priors, which, together with likelihoods based on sensory input, aid in making an optimal estimate for action selection (Körding and Wolpert, 2006). Animal studies have shown that repeating skilled movements leads to localized changes in the area responsible for the movement, whereas the pure repetition of unskilled movement does not (Plautz et al., 2000). In humans, early studies have shown that blocked practice leads to faster acquisition, but poorer retention and less transfer than variable practice (Shea and Morgan, 1979) and that massed practice without breaks seems less effective for motor performance (Ammons, 1947; Ammons and Willig, 1956).

In standard therapies or clinical studies, the amount of repetition is typically not quantified but was observed to be an order of magnitude lower than in studies investigating recovery



**TABLE 1** | Overview of the neuronal changes due to exposure to principles of neurorehabilitation included in this manuscript.

Experience-dependent changes	Principles	Brain areas	References
<b>Cellular/neuronal level</b>			
Increased neuronal activity	Spaced practice	Task/stimulus-dependent	Gerbier and Toppino, 2015
Increased cell survival and improved LTP	Spaced practice	Hippocampus	Scharf et al., 2002; Sisti et al., 2007
Upregulation of growth factors (protein 43, synaptophysin)	Dosage	Intact corticospinal tract	Zhao et al., 2013
Inhibition of upregulation of growth-inhibiting factors (NogoA, Nogo receptors and RhoA)	Dosage	Peri-infarct cortex	Zhao et al., 2013
Dopamine-dependent synaptic plasticity	Explicit feedback	Striatum	Kawagoe et al., 1998
Complex spikes in Purkinje cells	Implicit feedback	Cerebellum	Kitazawa et al., 1998
<b>Cortical motor areas</b>			
Expansion or change of effector representation/cortical map, dependent on effector trained	Massed practice	Motor cortex	Plautz et al., 2000
Increased excitability	- Dosage	- Motor cortex	- Liepert et al., 2000; Veerbeek et al., 2014
	- Variable practice	- Motor cortex	- Lage et al., 2015; Lin et al., 2011
Normalization of activation in ipsilesional cortex	Dosage	Motor cortex	Schaechter, 2004
Change in sensorimotor organization	Multisensory stimulation	Motor cortex	Rosenkranz and Rothwell, 2006
Increased neuronal recruitment during acquisition, decreased activity during retention	Variable practice	Prefrontal areas, PMA, inferior frontal areas	Lage et al., 2015; Lin et al., 2011
Increased cortical activation in lesioned hemisphere during paretic movement	- Task-specific practice	- SMC, PMC - SMC	- Jang et al., 2003 - Wilkins et al., 2017
	- Modulate effector selection	- SSC/SMA, dorsal PMC	- Johansen-Berg et al., 2002
Increased cortical activation in contralesional hemisphere during paretic movement	Rhythmic cueing	SMC	Luft et al., 2004
Decreased activation in contralesional hemisphere during paretic movement	Task-specific practice	- SMC, PMC, SMA - Motor cortex - SMA, PMA	- Jang et al., 2003 - Boyd et al., 2010 - Wilkins et al., 2017
Increased laterality index during paretic movement	Task-specific practice	- SMC - Motor cortex - SMC, SMA, PMA	- Jang et al., 2003 - Boyd et al., 2010 - Wilkins et al., 2017
Increased power spectra	Multisensory stimulation	SMC, SSC	Gomez-Rodriguez et al., 2011
<b>Fronto-parietal network</b>			
Increased activation of contralateral fronto-parietal network	Goal-oriented practice	Motor cortex, SMA, SSC, parietal areas	Nathan et al., 2012
Increased activation of bilateral parietal areas, together with lateralized pre-motor areas and sensorimotor areas	Increasing difficulty	PMC, SMA, SMC, SPA, IPA	Wexler et al., 1997; Winstein et al., 1997
Increased activation of bilateral parietal, premotor and visual areas	Action observation	Dorsal and ventral PMC, pre-SMA, SPA, IPA, visual cortex	Hardwick et al., 2018
Increased activation of lateralized parietal areas, together with pre-motor areas	Motor imagery	Bilateral dorsal PMC, left ventral PMC, Bilateral pre-SMA, left IPA, left SPA,	Hardwick et al., 2018
Increased activation and functional connectivity	Mirror therapy	- Ipsilateral motor cortex, visual processing areas	- Arya, 2016

(Continued)

TABLE 1 | Continued

Experience-dependent changes	Principles	Brain areas	References
		- Bilateral PMA, contralateral SMA and SMC, parietal cortex	- Hardwick et al., 2018
<b>Cerebellum</b>			
Increased activation	- Rhythmic cueing - Modulate effector selection	- Cerebellum (ipsilesional) - Cerebellum (bilateral)	- Luft et al., 2004 - Johansen-Berg et al., 2002
<b>Somatosensory Cortex</b>			
Reversal of SEP to pre-infarct	Dosage	Somatosensory cortex	Joo et al., 2012
<b>Extended networks</b>			
Auditory feedback lead to reduced activity during acquisition	Implicit feedback	SMC, SMA, opercular, temporal and parietal areas	Ronsse et al., 2011
Visual feedback lead to increased activity during acquisition	Implicit feedback	Occipital gyri, cerebellar lobules and vermis	Ronsse et al., 2011
Visual feedback preserved activation, when no feedback was given during testing	Implicit feedback	Occipitotemporal cortex	Ronsse et al., 2011
Auditory feedback suppressed activity, when no feedback was given during testing	Implicit feedback	Auditory cortex	Ronsse et al., 2011
Increased fractional anisotropy	Rhythmic cueing	Arcuate fasciculus (white matter tract connecting auditory and motor regions)	Moore et al., 2017
Activity in social cue network	Social interaction	Right posterior STS, right anterior STS, right TPJ	Redcay et al., 2010

LTP, long-term potentiation; PMC, premotor cortex; SPA, superior parietal area; IPA, inferior parietal area; SEP, somatosensory-evoked potentials; SMC, sensorimotor cortex; SMA, supplementary motor area; SSC, somatosensory cortex; STS superior temporal sulcus; TPJ, temporoparietal junction.

in rats and monkeys (Lang et al., 2007). Instead, the evidence for massed practice relies typically on the number of sessions or duration (French et al., 2016). A study looking into the feasibility of translating repetition amounts of animals to humans found improved motor functioning after training with high-repetition doses. However, no “pure” repetition training was provided, as the protocol included a variety of tasks that increased in difficulty (Birkenmeier et al., 2010). On the contrary, a study comparing four groups with different repetition amounts did not find significant differences based on the number of repetitions (Basso and Lang, 2017). This intervention included other principles as well. Meta-analyses confirm the mixed effects of repetitive training on improvement (Langhorne et al., 2011; Veerbeek et al., 2014; French et al., 2016; Thomas et al., 2017). Hence, massed practice appears to be a commonly used ingredient, but its clinical operationalization is often confounded with other principles. In order to investigate its true effects on recovery and compare across studies, the repetitions within a training session and across therapy duration should be measured and quantified.

## Spaced Practice

Spaced practice implies that training should be structured in time to include rest periods between repetitions or sessions (Lee and Genovese, 1988; Schmidt and Lee, 2011). Instead of spaced practice, the term *distributed practice* is often used in literature. However, some authors use the term distributed practice as a combination of spaced and massed practice (Cepeda et al., 2006). Research on human skill acquisition suggests that increasing the time spacing between learning periods improves final test performance (Cepeda et al., 2006). However, when these learning periods are too long, learning and retention rates drop (Savion-Lemieux and Penhune, 2005). The mechanisms behind the effects of distributed practice remain unclear. It has been hypothesized that the first exposure to a stimulus pre-activates its representation in memory, requiring no further activation in a subsequent repetition trial, leading to a poorer internal representation of that stimulus, which has been termed as the repetition suppression effect (Gerbier and Toppino, 2015). Animal and fMRI studies support this hypothesis, showing that

neuronal activation decreases after stimulus repetition where the magnitude is modulated by the delay between the first and second presentation, with larger delays leading to greater decreases (Brown et al., 1987; Henson et al., 2000, 2004; Henson, 2003). Spaced practice might counteract the repetition suppression effect by canceling stimulus priming (Gerbier and Toppino, 2015). TMS revealed that primary and supplementary motor areas are involved in motor memory consolidation (Censor and Cohen, 2011), which might be facilitated by spaced practice. Further, learning and physical activity have been linked to hippocampal neurogenesis (Praag et al., 1999). Animal studies also suggest that spaced practice facilitates long-term memory formation (Okamoto et al., 2011; Yamazaki et al., 2015) by fostering the survival of cells in the dentate gyrus that are important for learning and memory (Sisti et al., 2007). Also, *in vivo* spacing of electrical stimulation facilitates the recruitment of protein-synthesis-dependent processes, which facilitates late LTP effects (Scharf et al., 2002; Gerbier and Toppino, 2015).

In the clinical field, only a few studies have investigated the effect of spacing on post-stroke recovery. A clinical study that investigated whether a CIMT protocol could be distributed over more days with less therapy time per day showed improvement in motor outcomes that were similar to previous CIMT protocols and superior outcomes in long-term quality of life (Dettmers et al., 2005).

## Dosage

Unlike in pharmacology, dosage is an ill-defined term in rehabilitation (Dobkin, 2005; Kwakkel, 2009). Generally, it is operationalized as the number of hours spent in therapy (Kwakkel, 2009; Birkenmeier et al., 2010; Veerbeek et al., 2014; Basso and Lang, 2017), the frequency of training sessions and the duration of a session (Dobkin, 2005), or the training amount required to stimulate learning (Wadden et al., 2017). High dosages are often equated with high intensity of training (Kwakkel et al., 2015). However, the intensity of training could also be operationalized as the metabolic cost, work rate, or perceived intensity through exertion (Billinger, 2015; Hasan et al., 2016), which are rarely measured in standard therapies except in fitness and aerobic protocols (Kwakkel, 2009).

Typically, inpatients receive only 22 (Verbeek et al., 2014) to 60 min of training a day, with fewer minutes at later stages (Schaechter, 2004). There is some evidence that increasing therapy hours would be beneficial to speeding up functional recovery (Lohse et al., 2014; Veerbeek et al., 2014). At least 16 h of extra training (e.g., 71 more minutes per day for 3 months) within the first 6 months seem to be required for functional gains (Kwakkel et al., 2004; Veerbeek et al., 2014). However, there is some controversy over the benefits of increased training early after stroke (Schaechter, 2004; Dromerick et al., 2009; Kwakkel, 2009), and a pooled analysis revealed no evidence of an effect of additional doses (Hayward et al., 2014). Hence, the exact dose-response for different therapies at different stages post-stroke needs to be determined (Kwakkel, 2009; Basso and Lang, 2017). Also, it seems that motor performance needs to reach an asymptotic level in the first session to facilitate delayed performance gains across sessions

or days. Therefore, delayed performance gains seem not to depend on repetition or over-night consolidation, but on the amount of training that induces asymptote in the individual's performance (Hauptmann et al., 2005). Neurologically, high-dose rehabilitation protocols with extended training hours possibly induce structural plastic changes as well as a reorganization of neural networks (summarized by Kwakkel et al., 2015), increase cortical excitability and improve motor function and use (Liepert et al., 2000; Veerbeek et al., 2014). Several studies observed a normalization in ipsilesional cortex activity, which could underlie the functional gains (Schaechter, 2004).

## Task-Specific Practice

Task-specific practice postulates that changing the conditions of a task might require a change in the abilities needed to execute it; conditions during training should match the conditions during testing (Schmidt and Lee, 2011). Thus, the specific conditions of practice shape the internal sensorimotor representation of the skill learned (Nudo et al., 1996; Ridderinkhof et al., 2004), leading potentially to highly specialized skills (Keetch et al., 2005) whose performance is superior in transfer tasks that meet the training conditions (Schmidt and Lee, 2011). Grounded in this principle, conventional rehabilitation protocols focus their training on the execution of ADL, as they are deemed meaningful to the patient (Hubbard et al., 2009). Since the main target of rehabilitation is to enable the patient to perform ADL independently (Winstein et al., 2014), therapy might not prioritize the restoration of pre-stroke movement patterns but allows the patient to acquire compensatory movement skills.

One study with a large sample size found that task-specific practice appears to be similar to standard therapy in improving motor functionality (Winstein et al., 2016). On the other hand, smaller fMRI studies found that task-specific training facilitated motor learning and retention (Boyd et al., 2010) and induced a change in the laterality index, which was confirmed in other studies as well (Jang et al., 2003; Wilkins et al., 2017). However, while two studies found reduced activity in the contralesional cortex, one (Jang et al., 2003) found changes in neuronal activity patterns in both hemispheres. A study with TMS demonstrated a trend toward reduced interhemispheric inhibition following task-specific training (Singer et al., 2013).

## Goal-Oriented Practice

Since a given goal (e.g., throwing a ball into the basket) could be accomplished by many different motor synergies, it is assumed that movement control is achieved through the coupling of goal-specific functional movements. Goal-oriented practice, therefore, does not emphasize primarily individual muscles or movement patterns involved in execution but requires the patient to explore the couplings that are suitable to achieve the task (Horak, 1991). In general, motor skill performance and learning are enhanced if attention is directed to the effect of movement instead to the movement itself (Wulf and Prinz, 2001). Goal-oriented movements appear to produce a better reaching performance than the same movements without a goal (Wu et al., 2000), and setting specific, difficult goals leads to higher motor learning performance than non-specific

goals (Gauggel and Fischer, 2001). It appears that probing a skill in a goal-directed fashion after overnight consolidation promotes better performance than probing the skill by drawing attention to finger movements (Cohen et al., 2005). Evidence from studies looking into tool-use in animals and humans suggest that, neurologically, action goals are represented as effector-dependent in the anterior intraparietal sulcus and primary motor areas, and as effector-independent in the ventral intraparietal sulcus and premotor cortex (Gallivan and Culham, 2015). Goal-oriented movements produce higher activity in sensorimotor areas (Nathan et al., 2012).

There is some evidence that goal-oriented practice is beneficial for recovery (Bosch et al., 2014). However, the described interventions seem to be confounded by other principles that are sometimes ascribed to goal-oriented training (Harvey, 2009).

## Variable Practice

Variable practice can be achieved in two ways: (1) by providing variability within a training sequence, a method termed as *variability of practice* (Schmidt, 1975), or (2) by randomizing the presentation of individual training sequences, a method termed as *random practice* or *contextual interference* (Battig, 1966; Shea and Morgan, 1979). Both methods have been shown to lead to better retention (Shea and Kohl, 1991) and enhanced generalization to similar but untrained tasks (McCracken and Stelmach, 1977) or movements (Shea and Morgan, 1979; Mulder and Hochstenbach, 2001; Park et al., 2016), despite hampering initial performance (Shea and Morgan, 1979). However, a random presentation of information might be detrimental to motor learning (Mulder and Hochstenbach, 2001). Imaging studies have shed some light on the mechanisms supporting these effects. fMRI and TMS studies in humans indicate that improved performance due to variable practice correlates with increased neuronal activity and connectivity in the areas of the motor learning network during acquisition, which is associated with better performance at retention stages (Lage et al., 2015). Also, the motor cortex showed greater excitability during retention. These results point to more efficient retrieval of motor memory due to variable practice (Lin et al., 2011). More complex bimanual visuomotor tasks that were practiced randomly have shown modality-specific activation patterns that led to the recruitment of areas related to visual processing (Pauwels et al., 2018). The effect of variable practice might be related to the strong link between the neuromodulatory systems that control neuronal plasticity and novelty, for instance, the dopaminergic (Redgrave and Gurney, 2006), cholinergic (Hasselmo et al., 1996), and noradrenergic systems (Vankov et al., 1995), which are used by the brainstem activation system for controlling the global state of arousal (Gur et al., 2007).

In the clinical context, one study that investigated random versus blocked practice failed to find an effect (Hayward et al., 2014). It seems that this principle is rarely studied explicitly in clinical studies (Darekar et al., 2015; Nielsen et al., 2015), but instead applied in conjunction with other principles to overcome boredom (Birkenmeier et al., 2010).

## Increasing Difficulty

According to Guadagnoli and Lee (2004) and based on the ideas from Marteniuk (1976), task difficulty can be described by the training requirements and conditions that are pertinent to the task, called the nominal task difficulty, and by how challenging the training is relative to the skill of the performer, called the functional task difficulty. Practice leads to fewer prediction errors and less need to process error information. Increasing the nominal task difficulty hence increases prediction errors and error processing demands. The optimal challenge point lies where functional task difficulty leads to a balance between information processing demands and performance, which is optimal for learning (Marteniuk, 1976; Guadagnoli and Lee, 2004). It has been shown that training with difficulty levels personalized to the learner's capabilities leads to superior learning outcomes than when increases in difficulty are fixed (Wickens et al., 2013). Further, if subjects can control the task difficulty by themselves, their motor performance during acquisition and retention is significantly better (Andrieux et al., 2012). However, if difficulty surpasses one's perceived ability to succeed, it might lead to detrimental effects on performance (Gendolla, 1999). Brain imaging studies showed increased activity in lateralized pre-motor and sensorimotor areas, but with an even more pronounced increase in parietal areas, pointing to a specialization of that area for task complexity (Wexler et al., 1997; Winstein et al., 1997). Potentially, noradrenergic neurons keep track of high or low task performance due to difficulty by switching their activity pattern preceding behavior (Rajkowski et al., 2004; Aston-Jones and Cohen, 2005).

In stroke rehabilitation, task difficulty has been partly investigated through *shaping* or *graded practice*. Shaping is a concept that was initially used by behaviorists studying operant conditioning in animals and that was successfully transferred from animals to humans by making it part of CIMT (Taub, 1976; Taub and Uswatte, 2003): The use of the impaired limb is augmented by progressively increasing the complexity of the required movement (Taub et al., 1994; Kwakkel et al., 2015). Although shaping appears to be one of the essential components of CIMT, its particular effect on motor recovery has not been studied on its own (Kwakkel et al., 2015). Increasing difficulty has been successfully used in standard care studies (Woldag et al., 2010), robot-assisted therapy (Lucca, 2009), and VR-based systems (Cameirão et al., 2012; Ballester Rubio et al., 2016), all of which showed beneficial effects on motor recovery. Task difficulty appears to be implicitly present in many tasks that investigate motor learning without being explicitly operationalized.

## Multisensory Stimulation

The perception and integration of multiple senses are fundamental abilities of the brain. Because sensory information is noisy, the integration of various modalities requires probabilistic estimations to enhance perception (Knill and Pouget, 2004). Studies in the cat superior colliculus showed that a single neuron could be responsive to several sensory modalities (Meredith and Stein, 1986; Wallace and Stein, 1996). In primates, the classic areas associated with multisensory processing are the



superior temporal sulcus, the intraparietal cortex, and the frontal cortex, with newer studies confirming multisensory processing also in areas that were previously thought to be mainly unisensory (Ghazanfar and Schroeder, 2006). One sensory input (e.g., touch) can influence how another sensory modality is perceived (e.g., vision) (Driver and Noesselt, 2008); therefore, exposure to multisensory feedback can enhance the ability to detect, discriminate and recognize sensory information (Driver and Noesselt, 2008; Shams and Seitz, 2008; Gentile et al., 2011). For instance, active physical exploration of multisensory stimuli led to greater accuracy in an associative recognition task showing enhanced connectivity between sensory and motor cortices (Butler et al., 2011). Animal studies demonstrated that sensory feedback is crucial in motor learning. Monkeys with an ablated primary sensory hand area had no problems in executing a previously known task but were unable to learn new skills (Pavlidis et al., 1993). Providing multisensory stimulation during goal-oriented action execution might help to establish sensorimotor contingencies (McGann, 2010). Muscle vibrations appear to influence the sensorimotor organization, whereas paired associative stimulation with TMS increases motor-evoked-potentials (Rosenkranz and Rothwell, 2006).

Of specific interest for rehabilitation is the integration of visual and proprioceptive information to perform movements. It has been shown that vision and proprioception are weighted differently at various stages during motor planning (Sober and Sabes, 2003), suggesting a target for multisensory manipulations. Concurrent haptic feedback during motor imagery appears to enhance the classification accuracy of brain-computer interfaces when decoding movement intention, indicating that it can aid in closing the sensorimotor loop (Gomez-Rodriguez et al., 2011). Multisensory stimulation training might help patients to recover from unimodal deficits, for instance, visual deficits or auditory localization deficits (Làdavvas, 2008).

## Rhythmic Cueing

Neuroentrainment encompasses the study of the temporal relationship between the body's movements and the rhythmic stimulation emerging from the environment. Any sensory modality (auditory, visual, tactile, or vestibular) can be used for entrainment (Ross and Balasubramaniam, 2014). To date, there is not much literature about visual entrainment, possibly because the auditory-motor synchronization appears to be mainly driving internal rhythmic movement control (Ross and Balasubramaniam, 2014). Hence, mainly auditory cues are used to synchronize movements to rhythmic patterns (Rossignol and Jones, 1976; Schaefer, 2014). Rhythmic patterns act like a template whose sequence can be anticipated (Nombela et al., 2013). The regularity detection and tempo tracking of rhythmic patterns increases the activity in motor network areas and cerebellum (Schaefer, 2014) and creates a mental representation of the rhythm, the so-called auditory model, which enables motor movements to anticipate the rhythmic pattern. The pooled evidence provided in the reviews by Grahn (2012) and Nombela et al. (2013) suggests that there are neuronal interactions between auditory and motor systems (Grahn, 2012; Nombela et al., 2013), and auditory-cued motor

training can change their mutual structural connectivity (Moore et al., 2017). The auditory-motor action coupling relies on a subcortico-thalamic-cortical circuitry that can be activated through extrinsic cueing (Grahn, 2012; Nombela et al., 2013). Cerebellar patients cannot consciously perceive rhythm changes and show high variable motor responses. However, rhythmic synchronization, respectively, motor entrainment remains intact (Molinari et al., 2003), suggesting that the cerebellum might control the rhythmic auditory-motor synchronization by monitoring rhythmic patterns. Even without cueing, repetitive movements become periodic over time, as observed when analyzing gait patterns. The gait impairment observed in PD is ascribed to a deficiency of the internal timing ability that disturbs coordinated rhythmic locomotion, and which can be improved with rhythmical auditory stimulation (Thaut et al., 1996). Besides, rhythmic somatosensory cueing of stride frequency through vibrotactile stimulation at the wrist could improve qualitative walking performance in PD (van Wegen et al., 2006).

There is evidence that auditorily paced treadmill walking can improve gait coordination in stroke patients as well (Thaut and Abiru, 2010). Further, bilateral arm training with rhythmic auditory cueing enhances functional motor performance, which is maintained long-term (Whitall et al., 2000) and induces cortical and cerebellar changes (Luft et al., 2004). Meta-analyses found large effects that rhythmic auditory cueing improves walking velocity, cadence, and stride length (Yoo and Kim, 2016) and beneficial effects on improving upper limb impairment and function (Ghai, 2018) after stroke.

## Explicit Feedback/Knowledge of Results

KR has been defined as verbal, terminal and augmented feedback about goal achievement (Salmoni et al., 1984). Although the finding that extrinsic feedback can effectively create simple stimulus-response associations was brought forward by animal research in reinforcement learning, KR signifies more than just extrinsic rewards (Winstein, 1991; Schmidt and Lee, 2011). KR contributes to learning through cognitive processing, not through conditioning (Salmoni et al., 1984). KR is provided through explicit feedback. Explicit feedback is given on quantitative or qualitative task outcomes, e.g., correctness, exactness, success, or failure (Mazzoni and Krakauer, 2006; Subramanian et al., 2010; Schmidt and Lee, 2011). This feedback does not have to be verbal. For instance, when failing to reach for a target, the subject can hear unpleasant tones or see that the failed targets change color (Taylor et al., 2014). Also, explicit feedback about kinematic outcomes can be KR, e.g., playing back a recorded movement after execution. However, this feedback supports learning only if the movement features that led to the outcome are pointed out to the subject (Salmoni et al., 1984). Explicit feedback seems to activate explicit learning mechanisms and shows only subtle effects on implicit learning mechanisms (Taylor et al., 2014). While implicit learning appears to increase the cortical motor output maps of the involved movement initially, they return to baseline topography once the learned content can be explicitly declared. Possibly through explicit feedback a global motor plan is learned

that is represented by higher-order neuronal networks, which influence the cortical sensorimotor representations differently (Pascual-Leone et al., 1994). Rewarding or punishing feedbacks appear to have dissociative effects on skilled motor learning. Punishment can speed up motor learning, whereas rewards ensure long-term retention (Abe et al., 2011; Galea et al., 2015). The reinforcement of positive outcomes appears to foster a success-driven learning system, which limits decay after learning, possibly by mobilizing the dopaminergic system (Wickens et al., 2003). Reward expectations modulate the activity of caudate neurons (striatal projection neurons), which receive reward-related information through the dopaminergic input from substantia nigra and spatial information through the cortico-striatal connection. Consequently, they modulate the inhibitory output of the basal ganglia, biasing attention to rewarded items. Either reward-driven activity of caudate neurons is a result of cerebral plasticity, or activity in the cerebral cortex is influenced by caudate neurons through the output nuclei of basal ganglia (Kawagoe et al., 1998). Dopamine has a gradual build-up and can persist for longer time courses; it might support long-term memory formation of motor actions (Abe et al., 2011).

KR has been used to reinforce adherence to CIMT (Taub et al., 1994). Meta-analyses often analyze KR together with KP under the umbrella term augmented feedback (van Dijk et al., 2005; Hayward et al., 2014). A meta-analysis analyzing different feedback types reported positive effects on motor function for KR (Molier et al., 2010). However, this evidence is based on one study (Eckhouse et al., 1990), whose intervention included other principles as well. It can, therefore, not be established whether KR is effective for motor recovery.

## Implicit Feedback/Knowledge of Performance

KP was defined as feedback given about movement execution in the form of verbal descriptions, demonstrations, or replays of recordings (Gentile, 1972). Advances in technology made it possible that KP can be delivered online, in an implicit manner and concurrent during movement execution, providing verbal or non-verbal feedback about ongoing intrinsic somatic processes and movement kinematics (Salmoni et al., 1984; Winstein, 1991). For instance, feedback in the form of sounds and colors can be given while trunk displacements surpass a threshold (Subramanian et al., 2007). Biofeedback uses physiological sources like electromyograms to provide patients with real-time visual or auditory signals about their motor activity (Huang et al., 2006). Ultimately arm movements can be visualized and augmented using VR representations (Ballester Rubio et al., 2015b; Ferreira dos Santos et al., 2016).

Implicit sensory feedback enhances learning from sensorimotor prediction errors, which for instance can aid the adaptation to unexpected perturbations (Shadmehr et al., 2010), possibly by contributing to implicit learning mechanisms (Taylor et al., 2014). Concurrent implicit feedback leads to lasting adaptations to visuomotor rotations, which are not (Hinder et al., 2008) or less observed (Taylor et al., 2014)

when feedback about movement outcome, e.g., KR is given. Although KP appears to be beneficial during training, there is evidence that subjects can become dependent on it, showing inferior performance when feedback is removed (Ronsse et al., 2011). Ronsse et al. (2011) compared the effects of providing concurrent visual to concurrent auditory feedback during the acquisition of a bimanual movement pattern. The authors found that subjects that had obtained visual KP showed poorer performance during retention testing than subjects that were given auditory KP. During acquisition, the visual feedback increased the activity in vision/sensorimotor-specific areas, which was maintained during retention testing even in the absence of feedback. On the contrary, the concurrent auditory feedback reduced the activity in temporo-parieto-frontal areas and deactivated task-specific sensory areas during retention testing without feedback. These results suggest that subjects can become dependent on concurrent visual feedback, but not on concurrent auditory feedback because they rely on sensory processing areas that have become tuned to visual information during practice. The auditory feedback, on the other hand, might foster the formation of an internal controller, evidenced by the stronger activation of prefrontal areas. Alternatively, auditory feedback might promote reliance on proprioception and is consequently ignored during training (Ronsse et al., 2011). Results from cerebellar patients that were exposed to force-field learning tasks propose that the cerebellum may play an important role in using implicit information to correct and adapt motor commands to changed limb dynamics, and in forming internal controllers (Nezafat et al., 2001; Smith and Shadmehr, 2005; Tseng et al., 2007). In contrast to explicit error signals mediated through midbrain dopamine neurons in basal ganglia, implicit sensorimotor errors are possibly encoded by cerebellar climbing fibers and manifest in complex spikes in Purkinje cells during reaching tasks (Kitazawa et al., 1998). Computational modeling of adaptation to visuomotor rotations following concurrent visual feedback points to narrowly tuned neurons in the cortex that are driven by a prediction error that is computed by the cerebellum (Tanaka et al., 2009).

Stroke patients experienced a significant recovery in motor function and showed increased activation in the ipsilesional primary sensorimotor cortex after 4 weeks of training with a VR system that provided them with implicit feedback about their upper-limb movement (Jang et al., 2005). However, the system also included several other principles. In addition, the provision of KP has been shown to recover impaired movement patterns (Cirstea and Levin, 2007), to reduce learned non-use (Ballester Rubio et al., 2015b), and to lead to longer-lasting recovery effects (Subramanian et al., 2010). A meta-analysis found a beneficial effect for KP on motor function (Molier et al., 2010); however, the effect was based on two studies only.

## Modulate Effector Selection

In the acute stage after stroke, patients typically suppress the use of the affected limb due to pain, weakness, or malfunctioning (Taub and Uswatte, 2003). As a consequence, they are prone

to overuse the non-paretic limb, and the resulting under-usage of the impaired limb can cause a loss of behavioral and neuronal function (Andrews and Stewart, 1979; Taub et al., 2006). Some authors argue that this compensation strategy, called learned non-use, emerges because the spontaneous use of the paretic limb does not cross a threshold level (Han et al., 2008). Although standard therapy focuses on improving the functionality of the impaired limb, the improvement does not transfer to increased use of the arm for ADLs (Smania et al., 2012; Kwakkel et al., 2015).

Of those therapeutic approaches that were successful in counteracting learned non-use CIMT is the most common and most successful one (Kwakkel et al., 2015). An fMRI study revealed changes in brain activity patterns due to paretic arm use in patients that underwent a 2 weeks CIMT program at home where the non-affected arm was constrained for 90% of the waking time. Increased grip strength in the affected limb correlated significantly with increased fMRI signal change in ipsilesional cortico-cerebellar areas (Johansen-Berg et al., 2002). However, a meta-analysis did not find a pooled effect that forcing the use of the paretic arm alone is effective (Hayward et al., 2014). Other approaches aimed at promoting paretic arm-use through positive reinforcement during bilateral arm training (Ballester Rubio et al., 2016) or through wearable devices (e.g., bracelets) that provide feedback about performance of ADLs (Ballester Rubio et al., 2015a).

## Action Observation/Embodied Practice

Action observation (Martens et al., 1976) gained increased attention after the discovery of mirror neurons (Rizzolatti and Sinigaglia, 2010): in monkeys, some neurons discharged not only when the animal executed a motor command but also when it observed another individual executing it. In humans, subjects who first observed other individuals performing a novel task performed better in the same task than control subjects that did not observe other individuals or observed a slightly different task (Mattar and Gribble, 2005). It is thought that in monkeys, as in humans, action observation relies on the frontoparietal network (Rizzolatti and Sinigaglia, 2010). Indeed, a meta-analysis showed that in humans, movement observation, as well as movement execution, recruits mainly the premotor and parietal areas. Movement observation, however, exclusively activated the visual cortex, whereas execution activated the primary motor cortex (Hardwick et al., 2018). Therefore, action observation might facilitate movement execution and motor learning by facilitating the excitability of the motor system (Mulder, 2007). Indeed, TMS during action observation elicited increased muscle activation patterns (Fadiga et al., 1995). For practical reasons, action observation could be especially beneficial for stroke patients with severe hemiparesis or complete paralysis. There is some clinical evidence that action observation therapy can reduce impairment and increase brain activation in the frontoparietal network and bilateral cerebellum (Ertelt et al., 2007).

Besides internalizing someone else's movement, humans can also ascribe ownership and agency to body parts that do not pertain to them (Botvinick and Cohen, 1998). The discovery of rubber hand illusions (Botvinick and Cohen, 1998) led to

insights about the mechanisms underlying agency. Both the sense of agency (Sato and Yasuda, 2005) and ownership are susceptible to manipulations (Slater et al., 2010), that have been used for therapeutic purposes, for instance, in mirror therapy (Ramachandran and Rogers-Ramachandran, 1996). Similar to action observation, mirror therapy appears to rely on the frontoparietal circuit (Harmsen et al., 2015), which is why its motor learning effects are partly explained by the same mechanisms (Hamzei et al., 2012). However, contrary to movement observation, mirror therapy robustly activates the primary motor cortex and visual processing areas ipsilateral to the mirrored movement. Also, mirror therapy seems to increase functional connectivity between cortical motor areas and to excite the neural connection between the two hemispheres (Hamzei et al., 2012; Arya, 2016). A meta-analysis attests mirror therapy a significant long-term effect on motor function, the ADLs, the reduction of pain and the reduction of visuospatial neglect (Thieme et al., 2012).

If the impairment of the limb impedes active movement, visual illusions could be presented to the patients to simulate movements with the paretic arm. The error-prediction mechanism driven by the cerebellum could be equally activated through the alternative representation (Fiorio et al., 2014). Possibly, the stronger the visual illusion, the more agency is ascribed to it, which could explain the difference in brain activation patterns between action observation and mirror therapy. The sense of agency seems to be important when learning from sensorimotor prediction errors (Tsakiris et al., 2007), respectively agency is reduced when prediction and outcome do not match (Sato and Yasuda, 2005). However, there is no consensus on the definition of ownership and agency, which makes their operationalization in clinical practice difficult.

## Mental Practice/Motor Imagery

Mental practice and motor imagery rely on the ability to simulate actions mentally without overt behavior, as summarized by the simulation theory (Jeannerod, 2001). Motor imagery can be seen as a mental rehearsal of future movements and motor plans (Naito et al., 2002; Schmidt and Lee, 2011), that can be beneficial for motor learning (Di Rienzo et al., 2016). However, actual physical practice shows superior effects on learning (Hird et al., 1991). A meta-analysis compared the brain areas that are active during mental imagery and movement execution. Both seem to recruit premotor areas, somatosensory cortex, and subcortical areas. Also, activation in the mid-cingulate cortex was found, with motor imagery activating more the anterior region that is linked to the cognitive aspects of motor control, whereas motor execution recruiting more the posterior region that is associated with basic motor functions. While motor imagery appears to activate more the parietal cortex, movement execution appears to recruit more classic sensorimotor regions like the primary motor cortex and cingulate motor areas (Hardwick et al., 2018). These findings are in line with studies showing that lesions in the frontoparietal system can diminish the ability of motor imagery (Johnson, 2000; Danckert et al., 2002). Motor imagery and physical practice also appear to induce similar learning-dependent brain changes (Di Rienzo et al., 2016). Not



surprisingly, the activation pattern of motor imagery appears to be similar to the one identified in action observation and mirror therapy.

The learning effects of motor imagery and mental practice have been extensively studied in sports, whereas research regarding their clinical efficacy and efficiency is sparse and relatively recent (Mulder, 2007). However, motor imagery is thought to be advantageous for stroke recovery, especially for severely impaired patients (Mulder, 2007). Since patients retain the ability to imagine movements with the paretic limb, mental motor practice might facilitate functional reorganization (Johnson, 2000). A meta-analysis looking into the effectiveness of mental practice also found some trends for positive outcomes. However, pooled effects could not be estimated because only a few Class I studies exist, and their protocols, measurements, and interventions vary widely (Braun et al., 2006).

## Social Interaction

Social interaction has been defined as a behavior in which the participants' actions are both a response to and a stimulus for the counterpart's behavior (Rubin et al., 2006). Many ADL implicate social interaction, and a failure to perform them might lead to an undesired dependence on others (Lilja et al., 2003). The level of self-efficacy influences motor skill performance and learning, and in turn, is influenced by the appraisal or discouragement from others (Wulf et al., 2012). fMRI recordings of a subject experiencing a live social interaction revealed activations in areas commonly identified in the perception of social cues besides other regions involved in goal-directed and visual attention as well as reward processing (Redcay et al., 2010).

Animals that are allowed social interaction when recovering from an artery occlusion show higher functional improvement (Johansson and Ohlsson, 1996), increased recovery of behavior, and lower mortality, especially if the interaction partner was healthy (Venna et al., 2014). Including and investigating the impact of social interaction as part of the rehabilitation experience seems an important but missed opportunity. We found no study that was evaluating this specific aspect in a randomized controlled trial. One study evaluating enriched environments that included social interaction found positive results in terms of activity (Janssen et al., 2014).

## DISCUSSION

This synthesis aimed at identifying a set of principles that should guide the design of effective neurorehabilitation protocols for post-stroke recovery. We identified 15 principles based on existing work on motor learning and recovery: massed practice/repetitive practice, spaced practice, dosage/duration, task-specific practice, task-oriented practice, variable practice, increasing difficulty, multisensory information, rhythmic cueing, explicit feedback/knowledge of results, implicit feedback/knowledge of performance, modulate effector selection, action observation/embodied practice, mental practice, and social interaction. Where possible, we identified the therapeutic

and neurological effects of these principles from experimental work and clinical studies and commented on their limitations.

Our motivation for this analysis is twofold. Firstly, we are confident that the quality of evidence from clinical work and its interpretation would be enhanced if interventions are described along with the included principles. Reviews or meta-analyses with ambiguous effects often state that the included protocols remained vague on the exact experience provided to the patients, which makes the comparison and interpretation difficult (Veerbeek et al., 2014; Renton et al., 2017). By focusing solely on the ingredients of therapeutic interventions and compiling their current neuroscientific evidence, we aim to raise awareness of their importance. Also, this work might serve as a guide for clinicians and researchers to construct or identify the active ingredients in their interventions and to discover evidence currently missing. Secondly, we believe that there is a need to create a link between the principles of motor learning and their current operationalization in clinical studies and practice. We have identified several difficulties and shortcomings that do not aid in obtaining a common understanding of these principles and hence complicate the clinical investigation.

It seems that many principles are poorly operationalized in clinical trials. For instance, when massed practice is investigated, the repetitions performed within a session and during the treatment duration are rarely quantified (Lang et al., 2007) such that recovery effects due exclusively to repetition cannot be singled out. Also, the clinical research of spaced practice and dosage/duration would benefit if the parameters were quantified in a standardized way. Particularly dosage should be explicitly described in treatment minutes per session in order to be able to establish a dose-response due to training (Dobkin, 2005). Furthermore, dosage/duration should not be equated with intensity since the intensiveness of training cannot be estimated through treatment minutes only (Billinger, 2015). Intensity should be an independent principle that needs to be investigated separately. Task-specific and goal-oriented practice appear to be often used interchangeably (Winstein et al., 2014; Fu et al., 2015; Yue et al., 2017) although their training target is different. While task-specific practice focuses on the acquisition of a specific skill (Keetch et al., 2005) for ADL, goal-oriented practice permits the use of any movement or skill that is deemed suitable to achieve the goal (Horak, 1991), fostering the exploration of alternative movement patterns. Variability appears to be included inherently in many protocols (Darekar et al., 2015), possibly because it renders the training less repetitive and, therefore, less boring (Birkenmeier et al., 2010), which could counteract low adherence. However, this link has not been explicitly studied. Increasing the difficulty during practice is part of many intervention protocols as well; however, personalizing the difficulty level in order to provide training at the optimal challenge point seems to be rarely addressed. Concerning multisensory integration, it would be interesting to explore whether the presence of more than two sensory stimulations could enhance learning (Sánchez et al., 2013). Similarly, rhythmic entrainment could be extended with protocols exploring if visual or haptic entrainment might aid recovery of impaired movements (Penhune et al., 1998). Explicit feedback and implicit feedback are often investigated together



under the umbrella term of augmented feedback, as evidenced by the sourced meta-analysis and clinical studies (Molier et al., 2010). However, their aim and the neuronal mechanisms that they appear to stimulate are different. While explicit feedback provides terminal feedback about movement outcome, implicit feedback provides concurrent error-signals during movement execution fostering possibly different learning mechanisms. Meta-analyses also appear to interpret the sensory modality of the feedback, e.g., if it is visual, auditory or haptic as a feedback type. However, the sensory modality is a separate layer that is added to feedback. Explicit feedback, as well as implicit feedback, can be unisensory or multisensory. Action observation and mirror-therapy appear to be well studied therapeutic ingredients, whereas mental practice is only addressed in a few studies, and social interaction remains unexplored territory so far. If the principles would be better operationalized, it would not only help to identify their contribution to the recovery of motor functions, but also other learning outcomes such as cognitive or language improvements.

The neuronal changes found within each principle allow us to draw some general conclusions for the advancement of neurorehabilitation. While some principles appear to modulate more specific brain areas (massed practice, dosage, variable practice, task-specific practice, modulate effector selection, multisensory stimulation) within the motor areas of the cortex others appear to recruit or rely more on networks of brain regions (goal-oriented practice, increasing difficulty, action observation, motor imagery, mirror therapy, rhythmic cueing, implicit feedback/knowledge of results, social interaction). An effective rehabilitation approach should thus incorporate principles of both types in order to counteract neuronal degradation and promote improvement. Firstly, a training that addresses only a limited subset of the neuronal circuitry underlying a general function might limit transfer to other behaviors that depend on the same circuitry (Kleim and Jones, 2008). Secondly, not all principles are equally applicable to all patients. Some principles might be more beneficial early after stroke, whereas others benefit patients with less severe damage. Spontaneous biological recovery and activity-dependent plasticity appear to interact differently at different stages after stroke, which, aside from other factors like severity, predicts recovery (Reinkensmeyer et al., 2016; Hylin et al., 2017). It seems that in acute patients the sensorimotor cortex activity is highly abnormal, and the normalization in activity patterns is linked to better recovery (Schaechter, 2004). Principles like task-oriented practice that promote localized changes, might therefore be more beneficial at the acute stage after stroke (Schaechter, 2004), whereas

therapies like CIMT, where the forced use of the impaired limb is paired with increasing difficulty and further principles, have been shown to be more suitable at later stages after stroke and for less impaired patients (Dromerick et al., 2009). More severely impaired patients, on the other hand, might benefit from action observation, mirror therapy and motor imagery (Dohle et al., 2009; Sun et al., 2013). Future studies will show the optimal combinations of principles that stimulate plasticity in a way that learning of preexisting or novel functions is enhanced.

We are aware that the view proposed here is strongly influenced by knowledge mainly derived from clinical work with hemiparetic stroke patients. However, the literature indicates that other diseases, for instance, PD (Rossiter et al., 2014) or Alzheimer's disease (Kalaria, 2002), show similar cognitive, functional, and neuronal alterations even though they may have different pathologies. Therefore, these principles of neurorehabilitation could be potentially applied beyond the field of stroke. As our main goal was to provide a synthesis that is informative and practical, in-depth analysis of each principle and its neurological underpinnings lie outside of the scope of this work. In future work, we will unify the principles addressed here in a theoretical framework to show how each of them contributes to the restoration of sensorimotor contingencies (Verschure, 2011).

In summary, our review provides a synthesis of effective therapeutic ingredients that could be beneficial in aiding recovery after stroke. We hope that future work will extend the evidence presented here by implementing and investigating the principles of neurorehabilitation in novel rehabilitation protocols for stroke and other patient populations.

## AUTHOR CONTRIBUTIONS

MM, BB, and PV contributed to the design of the review. MM wrote the first draft of the manuscript and performed the computerized search. All authors contributed to the manuscript revision, read and approved the submitted version.

## FUNDING

This study was supported by EIT Health ID 19277 (RGS@HOME) under H2020, SANAR (MINECO, TIN2013-44200), cDAC (ERC 2013 ADG 341196), and socSMCs (Grant Number EC, H2020-641321).

## REFERENCES

- Abe, M., Schambra, H., Wassermann, E. M., Luckenbaugh, D., Schweighofer, N., and Cohen, L. G. (2011). Reward improves long-term retention of a motor memory through induction of offline memory gains. *Curr. Biol.* 21, 557–562. doi: 10.1016/j.cub.2011.02.030
- Ammons, R. B. (1947). Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.* 54, 263–281. doi: 10.1037/h0062455
- Ammons, R. B., and Willig, L. (1956). Acquisition of motor skill: IV. Effects of repeated periods of massed practice. *J. Exp. Psychol.* 51, 118–126. doi: 10.1037/h0048337
- Andrews, K., and Stewart, J. (1979). Stroke recovery: he can but does he? *Rheumatol. Rehabil.* 18, 43–48. doi: 10.1093/rheumatology/18.1.43
- Andrieux, M., Danna, J., and Thon, B. (2012). Self-Control of task difficulty during training enhances motor learning of a complex coincidence-anticipation task. *Res. Q. Exerc. Sport* 83, 27–35. doi: 10.1080/02701367.2012.10599822

- Arya, K. N. (2016). Underlying neural mechanism of mirror therapy: implications for motor rehabilitation in stroke. *Neurol. Individ.* 64, 38–44. doi: 10.1056/NEJMe0902377
- Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus Coeruleus-Norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709
- Ballester Rubio, B., Lathe, A., Duarte, E., Duff, A., and Verschure, P. F. M. J. (2015a). “A wearable bracelet device for promoting arm use in stroke patients,” in *Proceedings of the 3rd International Congress on Neurotechnology, Electronics and Informatics (NEUROTECHNIX 2015)*, Portugal, 24–31. doi: 10.5220/0005662300240031
- Ballester Rubio, B., Oliva, L. S., Duff, A., and Verschure, P. F. M. J. (2015b). “Accelerating motor adaptation by virtual reality based modulation of error memories,” in *Proceedings of the 2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*, Singapore, 623–629. doi: 10.1109/ICORR.2015.7281270
- Ballester Rubio, B., Maier, M., San Segundo, M. R., Castañeda, V., Duff, A., and Verschure, P. F. M. J. (2016). Counteracting learned non-use in chronic stroke patients with reinforcement-induced movement therapy. *J. Neuroeng. Rehabil.* 13, 1–15. doi: 10.1186/s12984-016-0178-x
- Basso, D. M., and Lang, C. E. (2017). Consideration of dose and timing when applying interventions after stroke and spinal cord injury. *J. Neurol. Phys. Ther.* 41, 1–19. doi: 10.1097/NPT.000000000000165.Consideration
- Battig, W. F. (1966). “Facilitation and interference,” in *Acquisition of Skill*, ed. E. A. Bilodeau, (New York, NY: Academic Press).
- Billinger, S. A. (2015). Does aerobic exercise and the FITT principle fit into stroke. *Curr. Neurol. Neurosci. Rep.* 15, 1–14. doi: 10.1007/s11910-014-0519-8
- Birkenmeier, R. L., Prager, E. M., and Lang, C. E. (2010). Translating animal doses of task-specific training to people with chronic stroke in 1-hour therapy sessions: a proof-of-concept study. *Neurorehabil. Neural Repair* 24, 620–635. doi: 10.1177/1545968310361957
- Bosch, J., O'Donnell, M. J., Barreca, S., Thabane, L., and Wishart, L. (2014). Does task-oriented practice improve upper extremity motor recovery after stroke? A systematic review. *ISRN Stroke* 2014, 1–10. doi: 10.1155/2014/504910
- Botvinick, M., and Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature* 391:756. doi: 10.1016/0005-2736(86)90402-5
- Boyd, L. A., Vidoni, E. D., and Wessel, B. D. (2010). Motor learning after stroke: is skill acquisition a prerequisite for contralateral neuroplastic change? *Neurosci. Lett.* 482, 21–25. doi: 10.1016/j.neulet.2010.06.082
- Braun, S. M., Beurskens, A. J., Borm, P. J., Schack, T., and Wade, D. T. (2006). The effects of mental practice in stroke rehabilitation: a systematic review. *Arch. Phys. Med. Rehabil.* 87, 842–852. doi: 10.1016/j.apmr.2006.02.034
- Brown, M. W., Wilson, F. A. W., and Riches, I. P. (1987). Neuronal evidence that inferomedial temporal cortex is more important than in certain processes underlying recognition memory. *Brain Res.* 409, 158–162. doi: 10.1016/0006-8993(87)90753-0
- Butler, A. J., James, T. W., and James, K. H. (2011). Enhanced multisensory integration and motor reactivation after active motor learning of audiovisual associations. *J. Cogn. Neurosci.* 23, 3515–3528. doi: 10.1162/jocn\_a\_00015
- Cameirão, M. S., Badia, S. B., Duarte, E., Frisoli, A., and Verschure, P. F. M. J. (2012). The combined impact of virtual reality neurorehabilitation and its interfaces on upper extremity functional recovery in patients with chronic stroke. *Stroke* 43, 2720–2728. doi: 10.1161/STROKEAHA.112.653196
- Carter, A. R., Connor, L. T., and Dromerick, A. W. (2010). Rehabilitation after stroke: current state of the science. *Curr. Neurol. Neurosci. Rep.* 10, 158–166. doi: 10.1007/s11910-010-0091-9.Rehabilitation
- Censor, N., and Cohen, L. G. (2011). Using repetitive transcranial magnetic stimulation to study the underlying neural mechanisms of human motor learning and memory. *J. Physiol.* 589, 21–28. doi: 10.1113/jphysiol.2010.198077
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J., and Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychol. Bull.* 132, 354–380. doi: 10.1037/0033-2909.132.3.354
- Cirstea, M. C., and Levin, M. F. (2007). Improvement of arm movement patterns and endpoint control depends on type of feedback during practice in stroke survivors. *Neurorehabil. Neural Repair* 21, 398–411. doi: 10.1177/1545968306298414
- Cohen, D. A., Pascual-Leone, A., Press, D. Z., and Robertson, E. M. (2005). Off-line learning of motor skill memory: a double dissociation of goal and movement. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18237–18241. doi: 10.1073/pnas.0506072102
- Danckert, J., Ferber, S., Doherty, T., Steinmetz, H., Nicolle, D., and Goodale, M. A. (2002). Selective, non-lateralized impairment of motor imagery following right parietal damage. *Neurocase* 8, 194–205. doi: 10.1093/neucas/8.3.194
- Darekar, A., McFadyen, B. J., Lamontagne, A., and Fung, J. (2015). Efficacy of virtual reality-based intervention on balance and mobility disorders post-stroke: a scoping review. *J. Neuroeng. Rehabil.* 12, 1–14. doi: 10.1186/s12984-015-0035-3
- Dettmers, C., Teske, U., Hamzei, F., Uswatte, G., Taub, E., and Weiller, C. (2005). Distributed form of constraint-induced movement therapy improves functional outcome and quality of life after stroke. *Arch. Phys. Med. Rehabil.* 86, 204–209. doi: 10.1016/j.apmr.2004.05.007
- Di Rienzo, F., Debarnot, U., Daligault, S., Saruco, E., Delpuech, C., Doyon, J., et al. (2016). Online and offline performance gains following motor imagery practice: a comprehensive review of behavioral and neuroimaging studies. *Front. Hum. Neurosci.* 10:315. doi: 10.3389/fnhum.2016.00315
- Dobkin, B. H. (2004). Strategies for stroke rehabilitation. *Lancet Neurol.* 3, 528–536. doi: 10.1016/S1474-4422(04)00851-8
- Dobkin, B. H. (2005). Rehabilitation and functional neuroimaging dose-response trajectories for clinical trials. *Neurorehabil. Neural Repair* 19, 276–282. doi: 10.1177/1545968305281892
- Dohle, C., Püllen, J., Nakaten, A., Küst, J., Rietz, C., and Karbe, H. (2009). Mirror therapy promotes recovery from severe hemiparesis: a randomized controlled trial. *Neurorehabil. Neural Repair* 23, 209–217. doi: 10.1177/1545968308324786
- Driver, J., and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on “Sensory-Specific” brain regions, neural responses, and judgments. *Neuron* 57, 11–23. doi: 10.1016/j.neuron.2007.12.013
- Dromerick, A. W., Lang, C. E., Birkenmeier, R. L., Wagner, J. M., Miller, J. P., Videen, T. O., et al. (2009). Very early constraint-induced movement during stroke rehabilitation (VECTORS). *Neurology* 73, 195–201. doi: 10.1212/WNL.0b013e3181ab2b27
- Eckhouse, R. H., Morash, R. P., and Maulucci, R. A. (1990). Sensory feedback and the impaired motor system. *J. Med. Syst.* 14, 93–105. doi: 10.1007/BF00996909
- Ertelt, D., Small, S., Solodkin, A., Dettmers, C., McNamara, A., Binkofski, F., et al. (2007). Action observation has a positive impact on rehabilitation of motor deficits after stroke. *Neuroimage* 36, 164–173. doi: 10.1016/j.neuroimage.2007.03.043
- Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *J. Neurophysiol.* 73, 2608–2611. doi: 10.1152/jn.1995.73.6.2608
- Ferreira dos Santos, L., Christ, O., Mate, K., Schmidt, H., Krüger, J., and Dohle, C. (2016). Movement visualisation in virtual reality rehabilitation of the lower limb: a systematic review. *Biomed. Eng. Online* 15, 75–88. doi: 10.1186/s12938-016-0289-4
- Fiorio, M., Mariotti, C., Panzeri, M., Antonello, E., Classen, J., and Tinazzi, M. (2014). The role of the cerebellum in dynamic changes of the sense of body ownership: a study in patients with Cerebellar degeneration. *J. Cogn. Neurosci.* 24, 712–721. doi: 10.1162/jocn
- French, B., Thomas, L. H., Coupe, J., McMahon, N. E., Connell, L., Harrison, J., et al. (2016). Repetitive task training for improving functional ability after stroke (Review). *Cochrane Database Syst. Rev.* 11, CD006073.
- Fu, M. J., Knutson, J., and Chae, J. (2015). Stroke rehabilitation using virtual environments. *Phys. Med. Rehabil. Clin. N. Am.* 26, 747–757. doi: 10.1016/j.pmr.2015.06.001.Stroke
- Furlan, L., Conforto, A. B., Cohen, L. G., and Sterr, A. (2016). Upper limb immobilisation: a neural plasticity model with relevance to poststroke motor rehabilitation. *Neural Plast.* 2016:8176217. doi: 10.1155/2016/8176217
- Galea, J. M., Mallia, E., Rothwell, J., and Diedrichsen, J. (2015). The dissociable effects of punishment and reward on motor learning. *Nat. Neurosci.* 18, 597–602. doi: 10.1038/nn.3956
- Gallivan, J. P., and Culham, J. C. (2015). Neural coding within human brain areas involved in actions. *Curr. Opin. Neurobiol.* 33, 141–149. doi: 10.1016/j.conb.2015.03.012

- Guggel, S., and Fischer, S. (2001). The effect of goal setting on motor performance and motor learning in brain-damaged patients. *Neuropsychol. Rehabil.* 11, 33–44. doi: 10.1080/09602010042000150
- Gendolla, G. H. E. (1999). Self-relevance of performance, task difficulty, and task engagement assessed as cardiovascular response. *Motiv. Emot.* 23, 45–66. doi: 10.1023/A:1021331501833
- Gentile, A. M. (1972). A working model of skill acquisition with application to teaching. *Quest* 17, 3–23. doi: 10.1080/00336297.1972.10519717
- Gentile, G., Petkova, V. I., and Ehrsson, H. H. (2011). Integration of visual and tactile signals from the hand in the human brain?: an fMRI study. *J. Neurophysiol.* 105, 910–922. doi: 10.1152/jn.00840.2010
- Gerbier, E., and Toppino, T. C. (2015). The effect of distributed practice: neuroscience, cognition, and education. *Trends Neurosci. Educ.* 4, 49–59. doi: 10.1016/j.tine.2015.01.001
- Ghai, S. (2018). Effects of real-time (sonification) and rhythmic auditory stimuli on recovering arm function post stroke: a systematic review and meta-analysis. *Front. Neurol.* 9:488. doi: 10.3389/fneur.2018.00488
- Ghazanfar, A. A., and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* 10, 278–285. doi: 10.1016/j.tics.2006.04.008
- Gomez-Rodriguez, M., Peters, J., Hill, J., Schölkopf, B., Gharabaghi, A., and Grosse-Wentrup, M. (2011). Closing the sensorimotor loop: haptic feedback facilitates decoding of motor imagery. *J. Neural Eng.* 8:036005. doi: 10.1088/1741-2560/8/3/036005
- Grahn, J. A. (2012). Neural mechanisms of rhythm perception: current findings and future perspectives. *Top. Cogn. Sci.* 4, 585–606. doi: 10.1111/j.1756-8765.2012.01213.x
- Guadagnoli, M. A., and Lee, T. D. (2004). Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *J. Mot. Behav.* 36, 212–224. doi: 10.3200/JMBR.36.2.212-224
- Gur, R. C., Turetsky, B. I., Loughhead, J., Waxman, J., Snyder, W., Ragland, J. D., et al. (2007). Hemodynamic responses in neural circuitries for detection of visual target and novelty: an event-related fMRI study. *Hum. Brain Mapp.* 28, 263–274. doi: 10.1002/hbm.20319
- Hamzei, F., Lämpchen, C. H., Glauche, V., Mader, I., Rijntjes, M., and Weiller, C. (2012). Functional plasticity induced by mirror training: the mirror as the element connecting both hands to one hemisphere. *Neurorehabil. Neural Repair* 26, 484–496. doi: 10.1177/1545968311427917
- Han, C. E., Arbib, M. A., and Schweighofer, N. (2008). Stroke rehabilitation reaches a threshold. *PLoS Comput. Biol.* 4:e1000133. doi: 10.1371/journal.pcbi.1000133
- Hardwick, R. M., Caspers, S., Eickhoff, S. B., and Swinnen, S. P. (2018). Neural correlates of action: comparing meta-analyses of imagery, observation, and execution. *Neurosci. Biobehav. Rev.* 94, 31–44. doi: 10.1016/j.neubiorev.2018.08.003
- Harmsen, W. J., Bussmann, J. B. J., Selles, R. W., Hurkmans, H. L. P., and Ribbers, G. M. (2015). A mirror therapy-based action observation protocol to improve motor learning after stroke. *Neurorehabil. Neural Repair* 29, 509–516. doi: 10.1177/1545968314558598
- Harvey, R. L. (2009). Improving poststroke recovery: neuroplasticity and task-oriented training. *Curr. Treat. Options Cardiovasc. Med.* 11, 251–259. doi: 10.1007/s11936-009-0026-4
- Hasan, S. M. M., Rancourt, S. N., Austin, M. W., and Ploughman, M. (2016). Defining optimal aerobic exercise parameters to affect complex motor and cognitive outcomes after stroke?: a systematic review and synthesis. *Neural Plast.* 2016, 1–12. doi: 10.1155/2016/2961573
- Hasselmo, M. E., Wyble, B. P., and Wallenstein, G. V. (1996). Encoding and retrieval of episodic memories: role of cholinergic and GABAergic modulation in the hippocampus. *Hippocampus* 6, 693–708. doi: 10.1002/(sici)1098-1063(1996)6:6<693::aid-hipo12>3.0.co;2-w
- Hauptmann, B., Reinhart, E., Brandt, S. A., and Karni, A. (2005). The predictive value of the leveling off of within-session performance for procedural memory consolidation. *Cogn. Brain Res.* 24, 181–189. doi: 10.1016/j.cogbrainres.2005.01.012
- Hayward, K. S., Barker, R. N., Carson, R. G., and Brauer, S. G. (2014). The effect of altering a single component of a rehabilitation programme on the functional recovery of stroke patients: a systematic review and meta-analysis. *Clin. Rehabil.* 28, 107–117. doi: 10.1177/0269215513497601
- Henson, R., Shallice, T., and Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science* 287, 1269–1272. doi: 10.1126/science.287.5456.1269
- Henson, R. N., Rylands, A., Ross, E., Vuilleumier, P., and Rugg, M. D. (2004). The effect of repetition lag on electrophysiological and haemodynamic correlates of visual object priming. *Neuroimage* 21, 1674–1689. doi: 10.1016/j.neuroimage.2003.12.020
- Henson, R. N. A. (2003). Neuroimaging studies of priming. *Prog. Neurobiol.* 70, 53–81. doi: 10.1016/S0301-0082(03)00086-8
- Hinder, M. R., Tresilian, J. R., Riek, S., and Carson, R. G. (2008). The contribution of visual feedback to visuomotor adaptation: how much and when? *Brain Res.* 1197, 123–134. doi: 10.1016/j.brainres.2007.12.067
- Hird, J. S., Landers, D. M., Thomas, J. R., and Horan, J. J. (1991). Physical practice is superior to mental practice in enhancing cognitive and motor task performance. *J. Sport Exerc. Psychol.* 8, 281–293. doi: 10.1123/jsep.13.3.281
- Horak, F. B. (1991). “Assumptions underlying motor control for neurologic rehabilitation,” in *Proceedings of the Contemporary Management of Motor Problems: Proceedings of the II Step Conference*, (Alexandria, VA: Foundation for Physical Therapy), 11–27.
- Huang, H., Wolf, S. L., and He, J. (2006). Recent developments in biofeedback for neuromotor rehabilitation. *J. Neuroeng. Rehabil.* 3:11. doi: 10.1186/1743-0003-3-11
- Hubbard, I. J., Parsons, M. W., Neilson, C., and Carey, L. M. (2009). Task-specific training: evidence for and translation to clinical practice. *Occup. Ther. Int.* 16, 190–203. doi: 10.1002/oti
- Hylin, M. J., Kerr, A. L., and Holden, R. (2017). Understanding the mechanisms of recovery and / or compensation following injury. *Neural Plast.* 2017:7125057. doi: 10.1155/2017/7125057
- Jang, S. H., Cho, S., Lee, J., and Park, J. (2003). Cortical reorganization induced by task-oriented training in chronic hemiplegic stroke patients. *Mot. Syst.* 14, 7–11. doi: 10.1097/01.wnr.0000051544.96524.f2
- Jang, S. H., You, S. H., Hallett, M., Cho, Y. W., Park, C.-M., Cho, S.-H., et al. (2005). Cortical reorganization and associated functional motor recovery after virtual reality in patients with chronic stroke: an experimenter-blind preliminary study. *Arch. Phys. Med. Rehabil.* 86, 2218–2223. doi: 10.1016/j.apmr.2005.04.015
- Janssen, H., Ada, L., Bernhardt, J., McElduff, P., Pollack, M., Nilsson, M., et al. (2014). An enriched environment increases activity in stroke patients undergoing rehabilitation in a mixed rehabilitation unit: a pilot non-randomized controlled trial. *Disabil. Rehabil.* 36, 255–262. doi: 10.3109/09638288.2013.788218
- Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage* 14, 103–109. doi: 10.1006/nimg.2001.0832
- Johansen-Berg, H., Dawes, H., Guy, C., Smith, S. M., Wade, D. T., and Matthews, P. M. (2002). Correlation between motor improvements and altered fMRI activity after rehabilitative therapy. *Brain* 125, 2731–2742. doi: 10.1093/brain/awf282
- Johansson, B. B., and Ohlsson, A. L. (1996). Environment, social interaction, and physical activity as determinants of functional outcome after cerebral infarction in the rat. *Exp. Neurol.* 139, 322–327. doi: 10.1006/exnr.1996.0106
- Johnson, S. H. (2000). Imagining the impossible: intact motor representations in hemiplegics. *Neuroreport* 11, 729–732. doi: 10.1097/00001756-200003200-00015
- Joo, H. W., Hyun, J. K., Kim, T. U., Chae, S. H., Lee, Y. I., and Lee, S. J. (2012). Influence of constraint-induced movement therapy upon evoked potentials in rats with cerebral infarction. *Eur. J. Neurosci.* 36, 3691–3697. doi: 10.1111/ejn.12014
- Kalaria, R. (2002). Similarities between Alzheimer’s disease and vascular dementia. *J. Neurol. Sci.* 20, 29–34. doi: 10.1016/S0022-510X(02)00256-3
- Kawagoe, R., Takikawa, Y., and Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.* 1, 411–416. doi: 10.1038/1625
- Keetch, K. M., Lee, T. D., Schmidt, R. A., and Young, D. E. (2005). Especial skills: their emergence with massive amounts of practice. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 970–978. doi: 10.1037/0096-1523.31.5.970



- Kitazawa, S., Kimura, T., and Yin, P.-B. (1998). Cerebellar complex spikes encode both destinations and errors in arm movements. *Nature* 392, 494–497. doi: 10.1038/33141
- Kleim, J. A., and Jones, T. A. (2008). Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. *J. Speech Lang. Hear. Res.* 51:S225. doi: 10.1044/1092-4388(2008/018)
- Knill, D. C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j.tics.2004.10.007
- Kollen, B. J., Lennon, S., Lyons, B., Wheatley-Smith, L., Scheper, M., Buurke, J. H., et al. (2009). The effectiveness of the bobath concept in stroke rehabilitation what is the evidence? *Stroke* 40:e89–97. doi: 10.1161/STROKEAHA.108.533828
- Körding, K. P., and Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends Cogn. Sci.* 10, 319–326. doi: 10.1016/j.tics.2006.05.003
- Kwakkel, G. (2009). Intensity of practice after stroke: more is better. *Schweizer Arch. fur Neurol. und Psychiatr.* 160, 295–298. doi: 10.1080/09638280500534861
- Kwakkel, G., van Peppen, R., Wagenaar, R. C., Wood Dauphinee, S., Richards, C., Ashburn, A., et al. (2004). Effects of augmented exercise therapy time after stroke. *Stroke* 35, 2529–2539. doi: 10.1161/01.str.0000143153.76460.7d
- Kwakkel, G., Veerbeek, J. M., van Wegen, E. E. H., and Wolf, S. L. (2015). Constraint-induced movement therapy after stroke. *Lancet Neurol.* 14, 224–234. doi: 10.1016/S1474-4422(14)70160-7
- Làdavas, E. (2008). Multisensory-based approach to the recovery of unisensory deficit. *Ann. N. Y. Acad. Sci.* 1124, 98–110. doi: 10.1196/annals.1440.008
- Lage, G. M., Ugrinowitsch, H., Apolinário-Souza, T., Vieira, M. M., Albuquerque, M. R., and Benda, R. N. (2015). Repetition and variation in motor practice: a review of neural correlates. *Neurosci. Biobehav. Rev.* 57, 132–141. doi: 10.1016/j.neubiorev.2015.08.012
- Lang, C. E., MacDonald, J. R., and Gnip, C. (2007). Counting repetitions: an observational study of outpatient therapy for people with hemiparesis post-stroke. *J. Neurol. Phys. Ther.* 31, 3–10. doi: 10.1097/01.NPT.0000260568.31746.34
- Langhorne, P., Bernhardt, J., and Kwakkel, G. (2011). Stroke rehabilitation. *Lancet* 377, 1693–1702. doi: 10.1016/S0140-6736(11)60325-5
- Langhorne, P., Coupar, F., and Pollock, A. (2009). Motor recovery after stroke? a systematic review. *Lancet Neurol.* 8, 741–754. doi: 10.1016/S1474-4422(09)70150-4
- Laver, K. E., Lange, B., George, S., Deutsch, J. E., Saposnik, G., and Crotty, M. (2017). Virtual reality for stroke rehabilitation. *Cochrane Database Syst. Rev.* 11:CD008349. doi: 10.1002/14651858.CD008349.pub4
- Lee, T. D., and Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: learning and performance effects reconsidered. *Res. Q. Exerc. Sport* 59, 277–287. doi: 10.1080/02701367.1988.10609373
- Levin, M. F., Weiss, P. L., and Keshner, E. A. (2015). Emergence of virtual reality as a tool for upper limb rehabilitation: incorporation of motor control and motor learning principles. *Phys. Ther.* 95, 415–425. doi: 10.2522/ptj.20130579
- Liepert, J., Bauder, H., Miltner, W. H. R., Taub, E., and Weiller, C. (2000). Treatment-induced cortical reorganization after stroke in humans. *Stroke* 31, 1210–1216. doi: 10.1161/01.STR.31.6.1210
- Lilja, M., Bergh, A., Johansson, L., and Nygård, L. (2003). Attitudes towards rehabilitation needs and support from assistive technology and the social environment among elderly people with disability. *Occup. Ther. Int.* 10, 75–93. doi: 10.1002/oti.178
- Lin, C. H., Knowlton, B. J., Chiang, M.-C., Iacoboni, M., Udompholkul, P., and Wu, A. D. (2011). Brain-behavior correlates of optimizing learning through interleaved practice. *Neuroimage* 56, 1758–1772. doi: 10.1016/j.neuroimage.2011.02.066
- Livingston-Thomas, J., Nelson, P., Karthikeyan, S., Antonescu, S., Jeffers, M. S., Marzolini, S., et al. (2016). Exercise and environmental enrichment as enablers of task-specific neuroplasticity and stroke recovery. *Neurotherapeutics* 13, 395–402. doi: 10.1007/s13311-016-0423-9
- Lohse, K. R., Lang, C. E., and Boyd, L. A. (2014). Is more better? Using metadata to explore dose-response relationships in stroke rehabilitation. *Stroke* 45, 2053–2058. doi: 10.1161/STROKEAHA.114.004695
- Lucca, L. F. (2009). Virtual reality and motor rehabilitation of the upper limb after stroke: a generation of progress? *J. Rehabil. Med.* 41, 1003–1006. doi: 10.2340/16501977-0405
- Luft, A. R., McCombe-Waller, S., Whithall, J., Forrester, L. W., Macko, R., Sorkin, J. D., et al. (2004). Repetitive bilateral arm training and motor. *JAMA* 292, 1853–1862.
- Maier, M., Ballester Rubio, B., Duff, A., Duarte Oller, E., and Verschure, P. F. M. J. (2019). Effect of specific over nonspecific VR-based rehabilitation on poststroke motor recovery: a systematic meta-analysis. *Neurorehabil. Neural Repair* 33, 112–129. doi: 10.1177/1545968318820169
- Marteniuk, R. G. (1976). *Information Processing in Motor Skills*, ed. R. Holt, (Toronto, CA: Holt).
- Martens, R., Burwitz, L., and Zuckerman, J. (1976). Modeling effects on motor performance. *Res. Quarterly. Am. Alliance Heal. Phys. Educ. Recreat.* 47, 277–291. doi: 10.1080/10671315.1976.10615372
- Mattar, A. A. G., and Gribble, P. L. (2005). Motor learning by observing. *Neuron* 46, 153–160. doi: 10.1016/j.neuron.2005.02.009
- Mazzoni, P., and Krakauer, J. W. (2006). An implicit plan overrides an explicit strategy during visuomotor adaptation. *J. Neurosci.* 26, 3642–3645. doi: 10.1523/JNEUROSCI.5317-05.2006
- McCracken, H. D., and Stelmach, G. E. (1977). A test of the schema theory of discrete motor learning. *J. Mot. Behav.* 9, 193–201. doi: 10.1080/00222895.1977.10735109
- McGann, M. (2010). perceptual modalities: modes of presentation or modes of interaction? *J. Conscious. Stud.* 17, 72–94.
- Meredith, M. A., and Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.* 56, 640–662. doi: 10.1152/jn.1986.56.3.640
- Middleton, E. L., and Schwartz, M. F. (2012). Errorless learning in cognitive rehabilitation: a critical review. *Neuropsychol. Rehabil.* 22, 138–168. doi: 10.1016/j.neuroimage.2013.08.045
- Molier, B. I., Van Asseldonk, E. H. F., Hermens, H. J., and Jannink, M. J. A. (2010). Nature, timing, frequency and type of augmented feedback; Does it influence motor relearning of the hemiparetic arm after stroke? A systematic review. *Disabil. Rehabil.* 32, 1799–1809. doi: 10.3109/09638281003734359
- Molinari, M., Leggio, M. G., De Martin, M., Cerasa, A., and Thaut, M. (2003). Neurobiology of rhythmic motor entrainment. *Ann. N. Y. Acad. Sci.* 999, 313–321. doi: 10.1196/annals.1284.042
- Moore, E., Schaefer, R. S., Bastin, M. E., Roberts, N., and Overy, K. (2017). Diffusion tensor MRI tractography reveals increased fractional anisotropy (FA) in arcuate fasciculus following music-cued motor training. *Brain Cogn.* 116, 40–46. doi: 10.1016/j.bandc.2017.05.001
- Mulder, T. (2007). Motor imagery and action observation: cognitive tools for rehabilitation. *J. Neural Transm.* 114, 1265–1278. doi: 10.1007/s00702-007-0763-z
- Mulder, T., and Hochstenbach, J. (2001). Adaptability and flexibility of the human motor system: implications for neurological rehabilitation. *Neural Plast.* 8, 131–140. doi: 10.1155/NP.2001.131
- Mulligan, M., Guess, D., Holvoet, J., and Brown, F. (1980). The individualized curriculum sequencing model (I): implications from research on massed, distributed, or spaced trial training. *Res. Pract. Pers. Sev. Disabil.* 5, 325–336. doi: 10.1177/154079698000500403
- Naito, E., Kochiyama, T., Kitada, R., Nakamura, S., Matsumura, M., Yonekura, Y., et al. (2002). Internally simulated movement sensations during motor imagery activate cortical motor areas and the cerebellum. *J. Neurosci.* 22, 3683–3691. doi: 10.1523/jneurosci.22-09-03683.2002
- Nathan, D. E., Prost, R. W., Guastello, S. J., Jeutter And, D. C., and Reynolds, N. C. (2012). Investigating the neural correlates of goal-oriented upper extremity movements. *Neuro Rehabil.* 31, 421–428. doi: 10.3233/NRE-2012-00812
- Nezafat, R., Shadmehr, R., and Holcomb, H. H. (2001). Long-term adaptation to dynamics of reaching movements: a PET study. *Exp. Brain Res.* 140, 66–76. doi: 10.1007/s002210100787
- Nielsen, J. B., Willerslev-Olsen, M., Christiansen, L., Lundbye-Jensen, J., and Lorentzen, J. (2015). Science-based neurorehabilitation: recommendations for neurorehabilitation from basic science. *J. Mot. Behav.* 47, 7–17. doi: 10.1080/00222895.2014.931273



- Nombela, C., Hughes, L. E., Owen, A. M., and Grahn, J. A. (2013). Into the groove: can rhythm influence Parkinson's disease? *Neurosci. Biobehav. Rev.* 37, 2564–2570. doi: 10.1016/j.neubiorev.2013.08.003
- Nudo, R. J., Milliken, G. W., Jenkins, W. M., and Merzenich, M. M. (1996). Use-dependent primary motor alterations of movement representations cortex of adult squirrel monkeys. *J. Neurosci.* 16, 785–807. doi: 10.1523/jneurosci.16-02-00785.1996
- Okamoto, T., Endo, S., Shirao, T., and Nagao, S. (2011). Role of cerebellar cortical protein synthesis in transfer of memory trace of cerebellum-dependent motor learning. *J. Neurosci.* 31, 8958–8966. doi: 10.1523/JNEUROSCI.1151-11.2011
- Park, H., Kim, S., Winstein, C. J., Gordon, J., and Schweighofer, N. (2016). Short-duration and intensive training improves long-term reaching performance in individuals with Chronic stroke. *Neurorehabil. Neural Repair* 30, 551–561. doi: 10.1177/1545968315606990
- Pascual-Leone, A., Grafman, J., and Hallett, M. (1994). Modulation of cortical motor output maps during development of implicit and explicit knowledge. *Science* 263, 1287–1289. doi: 10.1126/science.8122113
- Pauwels, L., Chalavi, S., Gooijers, J., Maes, C., Albouy, G., Sunaert, S., et al. (2018). Challenge to promote change: the neural basis of the contextual interference effect in young and older adults. *J. Neurosci.* 38, 3333–3345. doi: 10.1523/JNEUROSCI.2640-17.2018
- Pavlidis, C., Miyashita, E., and Asanuma, H. (1993). Projection from the sensory to the motor cortex is important in learning motor skills in the monkey. *J. Neurophysiol.* 70, 733–741. doi: 10.1016/j.jneuron.2011.07.029
- Penhune, V. B., Zatorre, R. J., and Evans, A. C. (1998). Cerebellar contributions to motor timing: a PET study of auditory and visual rhythm reproduction. *J. Cogn. Neurosci.* 10, 752–765. doi: 10.1162/089892998563149
- Plautz, E. J., Milliken, G. W., and Nudo, R. J. (2000). Effects of repetitive motor training on movement representations in adult squirrel monkeys: role of use versus learning. *Neurobiol. Learn. Mem.* 74, 27–55. doi: 10.1006/nlme.1999.3934
- Praag, H., Van Kempermann, G., and Gage, F. H. (1999). Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. *Nat. Neurosci.* 2, 266–270. doi: 10.1038/6368
- Proffitt, R., and Lange, B. (2015). Considerations in the efficacy and effectiveness of virtual reality interventions for stroke rehabilitation: moving the field forward. *Phys. Ther.* 95, 441–448. doi: 10.2522/ptj.20130571
- Rajkowski, J., Majczynski, H., Clayton, E., and Aston-Jones, G. (2004). Activation of monkey Locus Coeruleus neurons varies with difficulty and performance in a target detection task. *J. Neurophysiol.* 92, 361–371. doi: 10.1152/jn.00673.2003
- Ramachandran, V. S., and Rogers-Ramachandran, D. (1996). Synaesthesia in Phantom Limbs induced with mirrors. *Proc. Biol. Sci.* 263, 377–386. doi: 10.1098/rspb.1996.0058
- Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., et al. (2010). Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *Neuroimage* 50, 1639–1647. doi: 10.1038/jid.2014.371
- Redgrave, P., and Gurney, K. (2006). The short-latency dopamine signal: a role in discovering novel action. *Nat. Rev. Neurosci.* 7, 967–975. doi: 10.1038/nrn2022
- Reinkensmeyer, D. J., Burdet, E., Casadio, M., Krakauer, J. W., Kwakkel, G., Lang, C. E., et al. (2016). Computational neurorehabilitation: modeling plasticity and learning to predict recovery. *J. Neuroeng. Rehabil.* 13:42. doi: 10.1186/s12984-016-0148-3
- Renton, T., Tibbles, A., and Topolovec-Vranic, J. (2017). Neurofeedback as a form of cognitive rehabilitation therapy following stroke: a systematic review. *PLoS One* 12:e0177290. doi: 10.1371/journal.pone.0177290
- Ridderinkhof, K. R., van den Wildenberg, W. P. M., Segalowitz, S. J., and Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain Cogn.* 56, 129–140. doi: 10.1016/j.bandc.2004.09.016
- Rizzolatti, G., and Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat. Rev. Neurosci.* 11, 264–274. doi: 10.1038/nrn2805
- Ronsse, R., Puttemans, V., Coxon, J. P., Goble, D. J., Wagemans, J., Wenderoth, N., et al. (2011). Motor learning with augmented feedback: modality-dependent behavioral and neural consequences. *Cereb. Cortex* 21, 1283–1294. doi: 10.1093/cercor/bhq209
- Rosenkranz, K., and Rothwell, J. C. (2006). Differences between the effects of three plasticity inducing protocols on the organization of the human motor cortex. *Eur. J. Neurosci.* 23, 822–829. doi: 10.1111/j.1460-9568.2006.04605.x
- Ross, J. M., and Balasubramaniam, R. (2014). Physical and neural entrainment to rhythm: human sensorimotor coordination across tasks and effector systems. *Front. Hum. Neurosci.* 8:576. doi: 10.3389/fnhum.2014.00576
- Rossignol, S., and Jones, G. M. (1976). Audio-spinal influence in man studied by the H-reflex and its possible role on rhythmic movements synchronized to sound. *Electroencephalogr. Clin. Neurophysiol.* 41, 83–92. doi: 10.1016/0013-4694(76)90217-0
- Rossiter, H. E., Boudrias, M.-H., and Ward, N. S. (2014). Do movement-related beta oscillations change after stroke? *J. Neurophysiol.* 112, 2053–2058. doi: 10.1152/jn.00345.2014
- Rubin, K. H., Bukowski, W., and Parker, J. G. (2006). “Peer interactions, relationships, and groups,” in *Handbook of Child Psychology: Vol. 3. Social, Emotional, and Personality Development*, eds W. Damon, M. Lerner, and N. Eisenberg, (New York, NY: Wiley).
- Salmoni, A. W., Schmidt, R. A., and Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychol. Bull.* 95, 355–386. doi: 10.1037/0033-2909.95.3.355
- Sánchez, A., Millán-Calenti, J. C., Lorenzo-López, L., and Maseda, A. (2013). Multisensory stimulation for people with Dementia. *Am. J. Alzheimer's Dis.* 28, 7–14. doi: 10.1177/1533317512466693
- Sato, A., and Yasuda, A. (2005). Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition* 94, 241–255. doi: 10.1016/j.cognition.2004.04.003
- Savion-Lemieux, T., and Penhune, V. B. (2005). The effects of practice and delay on motor skill learning and retention. *Exp. Brain Res.* 161, 423–431. doi: 10.1007/s00221-004-2085-9
- Schachter, J. D. (2004). Motor rehabilitation and brain plasticity after hemiparetic stroke. *Prog. Neurobiol.* 73, 61–72. doi: 10.1016/j.pneurobio.2004.04.001
- Schaefer, R. S. (2014). Auditory rhythmic cueing in movement rehabilitation: findings and possible mechanisms. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369:20130402. doi: 10.1098/rstb.2013.0402
- Scharf, M. T., Woo, N. H., Lattal, K. M., Young, J. Z., Nguyen, P. V., and Abel, T. (2002). Protein synthesis is required for the enhancement of long-term potentiation and long-term memory by spaced training. *J. Neurophysiol.* 87, 2770–2777. doi: 10.1177/0093854806288066
- Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychol. Rev.* 82, 225–260. doi: 10.1037/h0076770
- Schmidt, R. A., and Lee, T. D. (2011). *Motor Control and Learning: A Behavioral Emphasis*, 5th Edn. Champaign, IL: Human Kinetics.
- Shadmehr, R., Smith, M. A., and Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annu. Rev. Neurosci.* 33, 89–108. doi: 10.1146/annurev-neuro-060909-153135
- Shams, L., and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cogn. Sci.* 12, 411–417. doi: 10.1016/j.tics.2008.07.006
- Shea, C. H., and Kohl, R. M. (1991). Composition of practice: influence on the retention of motor skills. *Res. Q. Exerc. Sport* 62, 187–195. doi: 10.1080/02701367.1991.10608709
- Shea, J. B., and Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *J. Exp. Psychol. Hum. Learn. Mem.* 5, 179–187.
- Singer, B., Vallence, A.-M., Cleary, S., Cooper, I., and Loftus, A. (2013). The effect of EMG triggered electrical stimulation plus task practice on arm function in chronic stroke patients with moderate-severe arm deficits. *Restor. Neurol. Neurosci.* 31, 681–691. doi: 10.3233/RNN-130319
- Sisti, H. M., Glass, A. L., and Shors, T. J. (2007). Neurogenesis and the spacing effect: learning over time enhances memory and the survival of new neurons. *Learn. Mem.* 14, 368–375. doi: 10.1101/lm.488707
- Slater, M., Spanlang, B., Sanchez-Vives, M. V., and Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS One* 5:e0010564. doi: 10.1371/journal.pone.0010564

- Smania, N., Gandolfi, M., Paolucci, S., Iosa, M., Ianes, P., Recchia, S., et al. (2012). Reduced-intensity modified constraint-induced movement therapy versus conventional therapy for upper extremity rehabilitation after stroke: a multicenter trial. *Neurorehabil. Neural Repair* 26, 1035–1045. doi: 10.1177/1545968312446003
- Smith, M. A., and Shadmehr, R. (2005). Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. *J. Neurophysiol.* 93, 2809–2821. doi: 10.1152/jn.00943.2004
- Sober, S. J., and Sabes, P. N. (2003). Multisensory integration during motor planning. *J. Neurosci.* 23, 6982–6992. doi: 10.1523/jneurosci.23-18-06982.2003
- Subramanian, S., Knaut, L. A., Beaudoin, C., McFadyen, B. J., Feldman, A. G., and Levin, M. F. (2007). Virtual reality environments for post-stroke arm rehabilitation. *J. Neuroeng. Rehabil.* 4, 1–5. doi: 10.1186/1743-0003-4-20
- Subramanian, S. K., Massie, C. L., Malcolm, M. P., and Levin, M. F. (2010). Does provision of extrinsic feedback result in improved motor learning in the upper Limb Poststroke? a systematic review of the evidence. *Neurorehabil. Neural Repair* 24, 113–124. doi: 10.1177/1545968309349941
- Sun, L., Yin, D., Zhu, Y., Fan, M., Zang, L., Wu, Y., et al. (2013). Cortical reorganization after motor imagery training in chronic stroke patients with severe motor impairment: a longitudinal fMRI study. *Neuroradiology* 55, 913–925. doi: 10.1007/s00234-013-1188-z
- Tanaka, H., Sejnowski, T. J., and Krakauer, J. W. (2009). Adaptation to visuomotor rotation through interaction between posterior parietal and motor cortical areas. *J. Neurophysiol.* 102, 2921–2932. doi: 10.1152/jn.90834.2008
- Taub, E. (1976). Movement in nonhuman primates deprived of somatosensory feedback. *Exerc. Sport Sci. Rev.* 4, 335–374.
- Taub, E., Crago, J. E., Burgio, L. D., Groomes, T. E., Cook, E. W., DeLuca, S. C., et al. (1994). An operant approach to rehabilitation medicine: overcoming learned nonuse by shaping. *J. Exp. Anal. Behav.* 61, 281–293. doi: 10.1901/jeab.1994.61-281
- Taub, E., and Uswatte, G. (2003). Constraint-Induced Movement therapy: bridging from the primate laboratory to the stroke rehabilitation laboratory. *J. Rehabil. Med.* 35, 34–40. doi: 10.1080/16501960310010124
- Taub, E., Uswatte, G., Mark, V. W., and Morris, D. M. (2006). The learned nonuse phenomenon: implications for rehabilitation. *Eura. Medicophys.* 42, 241–256.
- Taub, E., Uswatte, G., and Pidikiti, R. (1999). Constraint-induced movement therapy: a new family of techniques with broad application to physical rehabilitation - a clinical review. *J. Rehabil. Res. Dev.* 36, 237–251.
- Taylor, J. A., Krakauer, J. W., and Ivry, R. B. (2014). Explicit and implicit contributions to learning in a sensorimotor adaptation task. *J. Neurosci.* 34, 3023–3032. doi: 10.1523/JNEUROSCI.3619-13.2014
- Thaut, M. H., and Abiru, M. (2010). rhythmic auditory stimulation in rehabilitation of movement disorders: a review of current research. *Muscle Percept.* 27, 263–270.
- Thaut, M. H., McIntosh, G. C., Rice, R. R., Miller, R. A., Rathbun, J., and Brault, J. M. (1996). Rhythmic auditory stimulation in gait training for Parkinson's disease patients. *Mov. Disord.* 11, 193–200. doi: 10.1002/mds.870110213
- Thieme, H., Mehrholz, J., Pohl, M., Behrens, J., and Dohle, C. (2012). Mirror therapy for improving motor function after stroke (Review). *Cochrane Database Syst. Rev.* 14:CD008449. doi: 10.1002/14651858.CD008449
- Thomas, L. H., French, B., Coupe, J., McMahon, N., Connell, L., Harrison, J., et al. (2017). Repetitive task training for improving functional ability after stroke. *Stroke* 48, 102–104. doi: 10.1161/STROKEAHA.117.016503
- Tsakiris, M., Schütz-Bosbach, S., and Gallagher, S. (2007). On agency and body-ownership: phenomenological and neurocognitive reflections. *Conscious. Cogn.* 16, 645–660. doi: 10.1016/j.concog.2007.05.012
- Tseng, Y., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., and Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *J. Neurophysiol.* 98, 54–62. doi: 10.1152/jn.00266.2007
- van Dijk, H., Jannink, M. J. A., and Hermens, H. J. (2005). Effect of augmented feedback on motor function of the affected upper extremity in rehabilitation patients: a systematic review of randomized controlled trials. *J. Rehabil. Med.* 37, 202–211. doi: 10.1080/16501970510030165
- van Wegen, E., de Goede, C., Lim, I., Rietberg, M., Nieuwboer, A., Willems, A., et al. (2006). The effect of rhythmic somatosensory cueing on gait in patients with Parkinson's disease. *J. Neurol. Sci.* 248, 210–214. doi: 10.1016/j.jns.2006.05.034
- Vankov, A., Hervé-Minvielle, A., and Sara, S. J. (1995). Response to novelty and its rapid habituation in Locus Coeruleus neurons of the freely exploring rat. *Eur. J. Neurosci.* 7, 1180–1187. doi: 10.1111/j.1460-9568.1995.tb01108.x
- Veerbeek, J. M., van Wegen, E., van Peppen, R., van der Wees, P. J., Hendriks, E., Rietberg, M., et al. (2014). What is the evidence for physical therapy poststroke? a systematic review and meta-analysis. *PLoS One* 9:e0087987. doi: 10.1371/journal.pone.0087987
- Venna, V. R., Xu, Y., Doran, S. J., Patrizzi, A., and McCullough, L. D. (2014). Social interaction plays a critical role in neurogenesis and recovery after stroke. *Transl. Psychiatry* 4:e351. doi: 10.1038/tp.2013.128
- Verschure, P. F. M. J. (2011). “Neuroscience, virtual reality and neurorehabilitation: brain repair as a validation of brain theory,” *Paper Presented at the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Boston, 2254–2257. doi: 10.1109/IEMBS.2011.6090428
- Wadden, K. P., Asis, K., De Mang, C. S., Neva, J. L., Peters, S., Lakhani, B., et al. (2017). Predicting motor sequence learning in individuals with chronic stroke. *Neurorehabil. Neural Repair* 3, 95–104. doi: 10.1177/1545968316662526
- Wallace, M. T., and Stein, B. E. (1996). Sensory organization of the superior colliculus in cat and monkey. *Prog. Brain Res.* 112, 301–311. doi: 10.1016/s0079-6123(08)63337-3
- Wexler, B. E., Fulbright, R. K., Lacadie, C. M., Skudlarski, P., Kelz, M. B., Constable, R. T., et al. (1997). An fMRI study of the human cortical motor system response to increasing functional demands. *Magn. Reson. Imaging* 15, 385–396. doi: 10.1016/S0730-725X(96)00232-9
- Whitall, J., Waller, S. M., Silver, K. H. C., and Macko, R. F. (2000). Cueing improves motor function in chronic Hemiparetic stroke. *Stroke* 31, 2390–2395. doi: 10.1161/01.STR.31.10.2390
- Wickens, C. D., Hutchins, S., Carolan, T., and Cumming, J. (2013). Effectiveness of part-task training and increasing-difficulty training strategies: a meta-analysis approach. *Hum. Factors* 55, 461–470. doi: 10.1177/0018720812451994
- Wickens, J. R., Reynolds, J. N. J., and Hyland, B. I. (2003). Neural mechanisms of reward-related motor learning. *Curr. Opin. Neurobiol.* 13, 685–690. doi: 10.1016/j.conb.2003.10.013
- Wilkins, K. B., Owen, M., Ingo, C., Carmona, C., Dewald, J. P. A., and Yao, J. (2017). Neural plasticity in moderate to severe chronic stroke following a device-assisted task-specific arm/hand intervention. *Front. Neurol.* 8:284. doi: 10.3389/fneur.2017.00284
- Winstein, C. J. (1991). Knowledge of results and motor learning - Implications for physical therapy. *Phys. Ther.* 71, 140–149. doi: 10.1093/ptj/71.2.140
- Winstein, C. J., Grafton, S. T., and Pohl, P. S. (1997). Motor task difficulty and brain activity: investigation of goal-directed reciprocal aiming using positron emission tomography. *J. Neurophysiol.* 77, 1581–1594. doi: 10.1152/jn.1997.77.3.1581
- Winstein, C. J., Lewthwaite, R., Blanton, S. R., Wolf, L. B., and Wishart, L. (2014). Infusing motor learning research into neurorehabilitation practice: a historical perspective with case exemplar from the accelerated skill acquisition program. *J. Neuologic Phys. Ther.* 38, 190–200. doi: 10.1097/NPT.000000000000046
- Winstein, C. J., Wolf, S. L., Dromerick, A. W., Lane, C. J., Nelsen, M. A., Lewthwaite, R., et al. (2016). Effect of a task-oriented rehabilitation program on upper extremity recovery following motor stroke the ICARE randomized clinical trial. *JAMA* 315, 571–581. doi: 10.1001/jama.2016.0276
- Woldag, H., Stupka, K., and Hummelsheim, H. (2010). Repetitive training of complex hand and arm movements with shaping is beneficial for motor improvement in patients after stroke. *J. Rehabil. Med.* 42, 582–587. doi: 10.2340/16501977-0558
- Wu, C. Y., Trombly, C. A., Lin, K. C., and Tickle-Degnen, L. (2000). A kinematic study of contextual effects on reaching performance in persons with and without stroke: influences of object availability. *Arch. Phys. Med. Rehabil.* 81, 95–101. doi: 10.1053/apmr.2000.0810095
- Wulf, G., and Prinz, W. (2001). Directing attention to movement effects enhances learning: a review. *Psychon. Bull. Rev.* 8, 648–660. doi: 10.3758/BF03196201
- Wulf, G., Chiviawsky, S., and Lewthwaite, R. (2012). Altering mindset can enhance motor learning in older adults. *Psychol. Aging* 27, 14–21. doi: 10.1037/a0025718

- Yamazaki, T., Nagao, S., Lennon, W., and Tanaka, S. (2015). Modeling memory consolidation during posttraining periods in cerebellovestibular learning. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3541–3546. doi: 10.1073/pnas.1413798112
- Yoo, G. E., and Kim, S. J. (2016). Rhythmic auditory cueing in motor rehabilitation for stroke patients: systematic review and meta-analysis. *J. Music Ther.* 53, 149–177. doi: 10.1093/jmt/thw003
- Yue, Z., Zhang, X., and Wang, J. (2017). Hand rehabilitation robotics on poststroke motor recovery. *Behav. Neurol.* 2017, 1–20. doi: 10.1155/2017/3908135
- Zhang, J., Yu, J., Bao, Y., Xie, Q., Xu, Y., Zhang, J., et al. (2017). Constraint-induced aphasia therapy in post-stroke aphasia rehabilitation?: a systematic review and meta-analysis of randomized controlled trials. *PLoS One* 12:e0183349. doi: 10.1371/journal.pone.0183349
- Zhang, Y., Cai, J., Zhang, Y., Ren, T., Zhao, M., and Zhao, Q. (2016). Improvement in stroke-induced motor dysfunction by music-supported therapy: a systematic review and meta-analysis. *Sci. Rep.* 6:38521. doi: 10.1038/srep38521
- Zhao, S., Zhao, M., Xiao, T., Jolkkonen, J., and Zhao, C. (2013). Constraint-induced movement therapy overcomes the intrinsic axonal growth-inhibitory signals in stroke rats. *Stroke* 44, 1698–1705. doi: 10.1161/STROKEAHA.111.000361
- Conflict of Interest:** PV is the founder and interim CEO of Eodyne Systems S.L., which aims at bringing scientifically validated neurorehabilitation technology to society.
- The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2019 Maier, Ballester and Verschure. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Molecular Organization of Self-awareness: Paralimbic Dopamine-GABA Interaction

Hans C. Lou<sup>1\*</sup>, Kristine Rømer Thomsen<sup>2</sup> and Jean-Pierre Changeux<sup>3</sup>

<sup>1</sup>Center of Functionally Integrative Neuroscience, Institute for Clinical Medicine, Aarhus University, Aarhus, Denmark,

<sup>2</sup>Department of Psychology and Behavioral Sciences, Center for Alcohol and Drug Research, School of Business and Social Sciences, Aarhus, Denmark, <sup>3</sup>Institut Pasteur, Paris, France

The electrophysiology of the paralimbic network (“default mode”) for self-awareness has drawn much attention in the past couple of decades. In contrast, knowledge of the molecular organization of conscious experience has only lately come into focus. We here review newer data on dopaminergic control of awareness in humans, particularly in self-awareness. These results implicate mainly dopaminergic neurotransmission and the control of GABAergic function directly in the paralimbic network. The findings are important for understanding addiction, developmental disorders, and dysfunctional consciousness.

**Keywords:** self-awareness, default mode network, conscious experience, GABA, dopamine, addiction

## OPEN ACCESS

### Edited by:

Olivia Gosseries,  
University of Liège, Belgium

### Reviewed by:

M. Gustavo Murer,  
University of Buenos Aires, Argentina  
Pere Berbel,  
Universidad Miguel Hernández de  
Elche, Spain

### \*Correspondence:

Hans C. Lou  
hanslou1@gmail.com

**Received:** 04 October 2019

**Accepted:** 10 January 2020

**Published:** 28 January 2020

### Citation:

Lou HC, Rømer Thomsen K and Changeux J-P (2020) The Molecular Organization of Self-awareness: Paralimbic Dopamine-GABA Interaction. *Front. Syst. Neurosci.* 14:3. doi: 10.3389/fnsys.2020.00003

## INTRODUCTION

Self-awareness is a conscious experience with the self as an object. It is an essential part of conscious experience of the world, and is a tool for conscious self-monitoring and for controlling behavior. Its default may have grave consequences. The “neural correlates” of self-awareness have been studied by several investigators, including Devue and Brédart (2011) and D’Argembeau (2013). Generally, a cortical, paralimbic network has been proposed as a correlate for self-awareness. It often included prefrontal and medial parietal regions, but it remained unknown whether these regions were indeed instrumental in that function.

## THE PARALIMBIC NETWORK: FROM “CORRELATIONS” TO BEING “INSTRUMENTAL” IN SELF-AWARENESS

Based on the effect of temporary dysfunction of brain regions induced by transcranial magnetic stimulation (TMS), we have realized that a paralimbic network is indeed instrumental in self-awareness: the network is not only *active* during self-awareness (Kjaer et al., 2002; Lou et al., 2004), but is also *causal* for that function (Lou et al., 2004; Luber et al., 2012). Its core cortical regions, medial prefrontal and medial parietal are interneuron-rich hubs with multiple connections (Freund, 2003). They are usually linked with activity in the angular gyri, insula, and subcortical regions including striatum (Slagter et al., 2017) and thalamus (Rømer Thomsen et al., 2013; Bachman and Hudetz, 2014). The network is often termed the “default mode network” because it defaults when attention is turned away from the self to other tasks (Lou et al., 2017). The medial prefrontal and medial parietal regions are “hubs” of the network. They interact *via* GABA-induced synchronized gamma oscillations (Rømer Thomsen et al., 2013; Joansson et al., 2015).



## DOPAMINE-GABA INTERACTION IN CONSCIOUS EXPERIENCE

Functional brain imaging has indicated that abnormal conscious experiences in schizophrenia, like hallucinations and delusions, are associated with abnormal dopaminergic neurotransmission (Changeux and Lou, 2011; Palmiter, 2011). In order to test if this association is causal, we used signal detection tasks and subsequent subjective interpretation, i.e., by setting a criterion for when the sensory signal-to-noise ratio provides confidence that a stimulus is present (Lou et al., 2011). We examined the effect of increasing dopamine activation and showed that dopaminergic stimulation with the D<sub>1</sub> and D<sub>2</sub> receptor agonist pergolide is effective in increasing confidence in seeing words and in improving performance in a forced-choice word recognition task. This demonstrates neurotransmitter regulation of subjective conscious experience of perception and provides the first direct evidence that dopamine is instrumental in conscious experience (Lou et al., 2011; **Figure 1**). Such activation depends on fast-spiking GABAergic interneurons (Changeux and Lou, 2011), which intermittently inhibit pyramidal cell activity to produce gamma oscillations necessary for effective neurotransmission (Freund, 2003; Lou et al., 2011; Rømer Thomsen et al., 2013; Joensson et al., 2015).

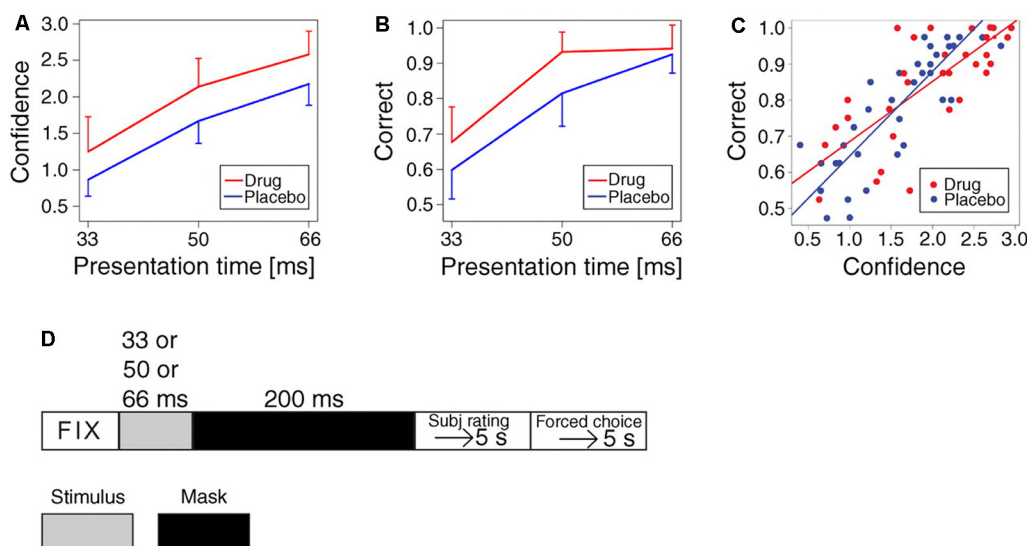
## SITE OF DOPAMINE-GABA INTERACTION

The site of dopamine-GABA interaction for self-awareness in the human brain was unknown until recently. To clarify this issue, we have used a PET ligand for GABA receptor binding.

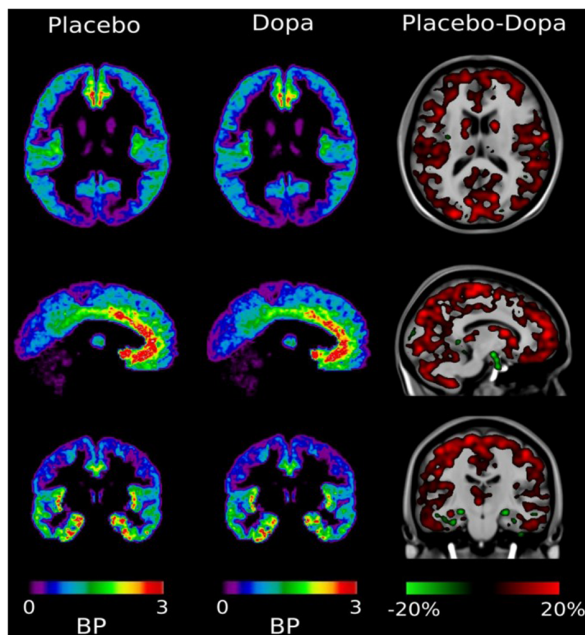
With this ligand, we were able to detect changes in dopamine-induced GABA binding under different physiological conditions at well-defined brain sites (Lou et al., 2016). It was found that GABA ligand binding was maximal in the medial anterior paralimbic region (cingulate gyrus), indicating empty GABA binding sites in these locations. Orally given dopamine reduced free GABA ligand binding sites throughout the cortex including the paralimbic system (**Figure 2**). In other words: dopamine increases GABA binding directly in the human paralimbic cortex, concomitantly with increased self-awareness and conscious experience (Lou et al., 2016). The effect is faltering in problem gambling, reflecting faltering self-monitoring and self-awareness (Møller et al., 2019). This development brings the properties of the GABA receptor molecule into focus.

## FUNCTIONAL ANATOMY OF GABA RECEPTORS

The GABA receptors are constructed as ligand ion channels. According to Stephens et al. (2017), they are “organized as a pore between five protein complexes. The pore allows passage of negative chloride ions, and, hence, the generation of electrical pulses when stabilized in an open conformation. This stabilization occurs by binding of GABA to the complex. The binding is not specific for GABA. The affinity of other molecules, whether physiologic, including dopamine, and foreign molecules depends on the protein composition of the five pentameric molecules constituting the pore and determine the function of the synapse. The subtype composition of the pentameric pore is abnormal in addiction.” This may explain the abnormal



**FIGURE 1 |** Effect of dopaminergic activation on confidence in seeing words and on accuracy in word recognition task. **(A)** Increase in the subjective rating of confidence (scale 0–3) for words presented at 33 ms, 50 ms, and 66 ms (means and standard deviations of means,  $p = 0.0018$ ). **(B)** Increase in accuracy (percentage of correct responses, in word recognition task by forced choice, one distractor). The expectation from chance: 50%. Note ceiling effect ( $p = 0.006$ ). **(C)** All observations of correct answers as a function of confidence, showing a significant effect ( $p < 0.0001$ , regression with random effect). The pergolide-treated group and the placebo groups had identical regression lines statistically. **(D)** Experimental timeline (Lou et al., 2011, p. 3). Adapted with permission from Lou et al. (2011).



**FIGURE 2 |** Distribution of [11C] Ro15-4513 GABA receptor ligand. After the placebo, the ligand was mainly bound in the medial-inferior prefrontal cingulate cortex, and right and left insula. After dopamine challenge ligand binding is reduced, with a general reduction of between 5% and 20% throughout the cortex. In hippocampal regions there were small foci of increased binding potential. Adapted with permission from Lou et al. (2016).

interaction between dopamine and GABA receptors in gambling disorder, a prime example of dysfunctional self-monitoring and self-control (Møller et al., 2019). Related to this, a number of studies suggest that altered GABA neurotransmission plays an important role in substance addiction (Lingford-Hughes et al., 2012, 2016) or gambling (Mick et al., 2017).

## SELF-AWARENESS NETWORK DURING INFANCY AND CHILDHOOD

In premature infants, functional MR imaging together with diffusion tensor imaging-based tractography has been used to study the relationship between performance on the Bayley Scales of Infant Development and early myelination (Cui et al., 2017). The Bayley Scales of Infant Development is a developmental play task that derives a developmental quotient (rather than an intelligence quotient). The authors confirmed such a link by showing that scores on the Bayley Scales of Infant Development were significantly associated with cingulate fractional anisotropy and radial diffusivity (Cui et al., 2017). This suggests that interconnecting axonal pathways within the default mode network are of critical importance already in the early neurocognitive development of infants.

Newborn infants already show a form of “basic consciousness” by establishing rudimentary eye contact with their mother (Lagercrantz and Changeux, 2009). A classical test for the presence of self-awareness in infants (and animals) is the mirror recognition test, where the infant is placed before

a mirror with a spot marked on his/her forehead. A positive response is usually present at about 2 years of age. It requires the infant to try to remove the spot on his/her head, and not on the mirror (Anderson, 1984).

A recent review showed that the default mode network follows an inverse U-shape, where it is weaker in children and elderly and stronger in adults. Cognitive function is positively correlated with default mode network functional connectivity (Mak et al., 2017).

## DEFICIENT SELF-AWARENESS AND PATHOLOGY

In later childhood and adulthood, disturbance of the paralimbic network is linked to severe pathology. Thus, deficient GABA neurotransmission is prominent in disorders with poor self-awareness and self-monitoring such as addiction (Lingford-Hughes et al., 2012, 2016; Mick et al., 2017; Møller et al., 2019), autism (Hashemi et al., 2017), anosognosia, hallucinations, and delusions (Therriault et al., 2018), progressing to schizophrenia (Rikandi et al., 2018) major depression (Rikandi et al., 2018), and/or dementia (Lin et al., 2015). Even in the vegetative state (also termed unresponsive wakefulness syndrome, Laureys et al., 2010), recovery of the paralimbic network is tightly linked to clinical recovery (Thibaut et al., 2019). Finally, a large clinical study of the possible therapeutic effect of apomorphine, a dopaminergic drug, is now underway (Sanz et al., 2019).

To determine if deficient GABA neurotransmission in pathology could be a primary event or secondary to toxic or pathological effects in more complex disorders, we examined if deficient dopamine-GABA neurotransmission was present in a relatively mono-symptomatic disorder such as gambling disorder (Møller et al., 2019). This was indeed the case. Therefore, normal interaction between these transmitters in the medial paralimbic system seems to be fundamental for brain function.

## DOPAMINE ACTIVITY LINKED WITH OTHER TRANSMITTERS

The basal forebrain part of the system is not only regulated by *dopamine*. It is also a prominent site of *cholinergic* activity. Interaction between the cholinergic and dopaminergic system *via* GABA receptors has been well described (Changeux and Lou, 2011; Takács et al., 2018). Thereby it promotes functions related to conscious experiences such as attention, learning and memory, and sleep-wake alternation (Lin et al., 2015).

## VULNERABILITY OF PARALIMBIC NETWORK

The widespread dysfunction of self-awareness in disease is likely to be a consequence of the exceedingly high oxygen demand of the paralimbic network. The high oxygen requirement is considered to be the result of dense concentrations of parvalbumin GABAergic interneurons in the richly connected hubs of the paralimbic network. In particular, the fast gamma

oscillations are susceptible to metabolic disruptions because of their high energy-demand (Kann et al., 2016).

## DISCUSSION

Until recently, conscious experience and self-awareness were considered off-limits for the natural sciences. Neurobiological research shunned the “hard question” of how conscious experience and self-awareness arise from a physical basis. Hence, it has been fashionable to limit neuroscience to try to identify neural “correlates” of conscious experience and self-awareness. The risk is evident for arriving at two parallel worlds: a mental and a physical, without understanding how they interact. This limitation has impeded our understanding of the biological function of self-awareness, and how it may account for disease. We have here reported data showing that self-awareness and conscious experience can be disturbed by electrophysiological manipulation of the paralimbic network (Lou et al., 2004; Lubner et al., 2012). Therefore, we may conclude that the network is instrumental for these functions. Newer data reviewed here

also strengthen the conclusion by showing that dopaminergic agents may stimulate conscious experience *via* GABA receptors in the paralimbic network. This finding has already resulted in a large and promising study of disabled persons with faltering self-awareness and consciousness (Sanz et al., 2019).

## AUTHOR CONTRIBUTIONS

HL conceptualized the review and wrote the initial draft. KR and J-PC participated in writing the manuscript.

## FUNDING

The authors received no financial support for the research and authorship of this article.

## ACKNOWLEDGMENTS

We thank Henriette Vuust, CFIN, Aarhus University, for help with figures.

## REFERENCES

- Anderson, J. R. (1984). The development of self-recognition: a review. *Dev. Psychobiol.* 17, 35–49. doi: 10.1016/s0273-2297(84)80006-4
- Bachman, T., and Hudetz, A. G. (2014). It is time to combine the two main traditions in the research on the neural correlates of consciousness:  $C = L \times D$ . *Front. Psychol.* 5:940. doi: 10.3389/fpsyg.2014.00940
- Changeux, J. P., and Lou, H. C. (2011). Emergent pharmacology of conscious experience. *FASEB J.* 25, 2098–2108. doi: 10.1096/fj.11-0702ufm
- Cui, J., Tymoflyeva, O., Desikan, R., Flynn, T., Kim, H., Gano, D., et al. (2017). Microstructure of the default mode network in the preterm infant. *AJNR Am. J. Neuroradiol.* 38, 343–348. doi: 10.3174/ajnr.A4997
- D'Argebeau, A. (2013). On the role of the ventromedial prefrontal cortex in self-processing: the valuation hypothesis. *Front. Hum. Neurosci.* 7:372. doi: 10.3389/fnhum.2013.00372
- Deve, C., and Brédart, S. (2011). Neural correlates of visual self-recognition. *Conscious. Cogn.* 20, 40–51. doi: 10.1016/j.concog.2010.09.007
- Freund, T. F. (2003). Interneuron diversity series: rhythm and mood in perisomatic inhibition. *Trends Neurosci.* 26, 489–495. doi: 10.1016/s0166-2236(03)00227-3
- Hashemi, E., Ariza, J., Rogers, N., Noctor, S. C., and Martínez-Cardeno, V. (2017). The number of parvalbumin-expressing interneurons is decreased in the prefrontal cortex in autism. *Cereb. Cortex* 27, 1931–1943. doi: 10.1093/cercor/bhw021
- Joensson, M., Rømer Thomsen, K., Andersen, L. M., Gross, J., Mouridsen, K., Sandberg, K., et al. (2015). Making sense: dopamine activates conscious self-monitoring through medial prefrontal cortex. *Hum. Brain Mapp.* 36, 1866–1877. doi: 10.1002/hbm.22742
- Kann, O., Hollnagel, J. O., Elzoheiry, S., and Schneider, J. (2016). Energy and potassium homeostasis during  $\gamma$  oscillations. *Front. Mol. Neurosci.* 9:47. doi: 10.3389/fnmol.2016.00047
- Kjaer, T. W., Nowak, M., and Lou, H. C. (2002). Reflective self-awareness and representation of the mental self. *NeuroImage* 17, 1080–1086. doi: 10.1006/nimg.2002.1230
- Lagercrantz, H., and Changeux, J. P. (2009). The emergence of human consciousness: from fetal to neonatal life. *Pediatr. Res.* 65, 255–260. doi: 10.1203/pdr.0b013e3181973b0d
- Laureys, S., Celesia, G. G., Cohandon, F., Lavrijnsen, J., León-Carrión, J., Sannita, W. G., et al. (2010). Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. *BMC Med.* 8:68. doi: 10.1186/1741-7015-8-68
- Lin, S. C., Brown, R. E., Hussain Shuler, M. G., Petersen, C. C., and Kepecs, A. (2015). Optogenetic dissection of the basal forebrain. *J. Neurosci.* 35, 13896–13901. doi: 10.1523/JNEUROSCI.2590-15.2015
- Lingford-Hughes, A., Myers, J., Watson, B., Reid, A. G., Kalk, N., Feeney, A., et al. (2016). Using [ $^{11}\text{C}$ ]Ro15 4513 PET to characterize GABA-benzodiazepine receptors in opiate addiction: similarities and differences with alcoholism. *Neuroimage* 132, 1–7. doi: 10.1016/j.neuroimage.2016.02.005
- Lingford-Hughes, A., Reid, A. G., Myers, J., Feeney, A., Hammers, A., Taylor, L. G., et al. (2012). A [ $^{11}\text{C}$ ]Ro15 4513 PET study suggests that alcohol dependence in man is associated with reduced  $\alpha 5$  benzodiazepine receptors in limbic regions. *J. Psychopharmacol.* 26, 273–281. doi: 10.1177/0269881110379509
- Lou, H. C., Changeux, J. P., and Rosenstand, A. (2017). Towards a cognitive neuroscience of self-awareness. *Neurosci. Biobehav. Rev.* 83, 765–773. doi: 10.1016/j.neubiorev.2016.04.004
- Lou, H. C., Lubner, B., Crupain, M., Keenan, J. P., Nowak, M., Kjaer, T. W., et al. (2004). Parietal cortex representation of the mental Self. *Proc. Natl. Acad. Sci. U S A* 101, 6827–6832. doi: 10.1073/pnas.0400049101
- Lou, H. C., Rosentand, A., Brooks, D. J., Bender, D., Jakobsen, S., Blicher, J. U., et al. (2016). Exogenous dopamine reduces GABA receptor availability in the human brain. *Brain Behav.* 6:e00484. doi: 10.1002/brb3.484
- Lou, H. C., Skewes, J. C., Thomsen, K. R., Overgaard, M., Law, H. C., Mouridsen, K., et al. (2011). Dopaminergic stimulation enhances confidence and accuracy in seeing rapidly presented words. *J. Vis.* 11:15. doi: 10.1167/11.2.15
- Lubner, B., Lou, H. C., Keenan, J. P., and Lisanby, S. H. (2012). Self-enhancement in the default network. *Exp. Brain Res.* 223, 177–187. doi: 10.1007/s00221-012-3249-7
- Mak, L. E., Minuzzi, L., MacQueen, G., Hall, G., Kennedy, S. H., and Milev, R. (2017). The default mode network in healthy individuals. *Brain Connect.* 7, 25–33. doi: 10.1089/brain.2016.0438
- Mick, I., Ramos, A. C., Myers, J., Stokes, P. R., Chandrasekera, S., Mendez, M. A., et al. (2017). Evidence for GABA-A receptor dysregulation in gambling disorder: correlation with impulsivity. *Addict. Biol.* 22, 1601–1609. doi: 10.1111/adb.12457
- Møller, A., Rømer Thomsen, K., Brooks, D. J., Blicher, J. U., Hansen, K. V., Mouridsen, K., et al. (2019). Attenuation of dopamine-induced GABA release in problem gamblers. *Brain Behav.* 9:e01239. doi: 10.1002/brb3.1239
- Palmiter, R. D. (2011). Dopamine signaling as a neural correlate of consciousness. *Neuroscience* 198, 213–220. doi: 10.1016/j.neuroscience.2011.06.089
- Rikandi, E., Mäntylä, T., Lindgren, M., Kieseppä, T., Suvisaari, J., and Raji, T. T. (2018). Connectivity of the precuneus-posterior cingulate cortex with the anterior cingulate-medial prefrontal cortex differs consistently between control

- subjects and first-episode psychosis patients during a movie stimulus. *Schizophr. Res.* 199, 235–242. doi: 10.1016/j.schres.2018.03.018
- Rømer Thomsen, K., Joensson, M., Lou, H. C., Møller, A., Gross, J., Kringelbach, M. L., et al. (2013). Altered paralimbic interaction in behavioral addiction. *Proc. Natl. Acad. Sci. U S A* 110, 4744–4749. doi: 10.1073/pnas.1302374110
- Sanz, L. R. D., Lejeune, N., Blandiaux, S., Bonin, E., Thibaut, E., Stenter, J., et al. (2019). Treating disorders of consciousness with apomorphine: protocol for a double-blind randomized controlled trial using multimodal assessments. *Front. Neurol.* 10:248. doi: 10.3389/fneur.2019.00248
- Slagter, H. A., Mazaheri, A., Reteig, L. C., Smolders, R., Figue, M., Mantione, M., et al. (2017). Contributions of the ventral striatum to conscious perception: an intracranial EEG study of the attentional blink. *J. Neurosci.* 37, 1081–1089. doi: 10.1523/JNEUROSCI.2282-16.2016
- Stephens, D. N., King, S. L., Lambert, J. J., Beileli, D., and Duka, T. (2017). GABAA receptor subtype involvement in addictive behavior. *Genes Brain Behav.* 16, 149–184. doi: 10.1111/gbb.12321
- Takács, V. T., Cserép, C., Schlingloff, D., Pósfai, B., Szonyi, A., Sos, K. E., et al. (2018). Co-transmission of acetylcholine and GABA regulates hippocampal states. *Nat. Commun.* 9:2848. doi: 10.1038/s41467-018-05136-1
- Therriault, J., Ng, K. P., Pascoal, T. A., Mathotaarachchi, S., Kang, M. S., Struys, H., et al. (2018). Anosognosia predicts default mode network hypometabolism and clinical progression to dementia. *Neurology* 90, e932–e939. doi: 10.1212/wnl.0000000000005120
- Thibaut, A., Schiff, N., Glacino, J., Laureys, S., and Gosseries, O. (2019). Therapeutic interventions in patients with prolonged disorders of consciousness. *Lancet Neurol.* 18, 600–614. doi: 10.1016/s1474-4422(19)30031-6

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Lou, Rømer Thomsen and Changeux. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Multiple Network Disconnection in Anosognosia for Hemiplegia

Elena Monai<sup>1,2</sup>, Francesca Bernocchi<sup>1,2</sup>, Marta Bisio<sup>1,2</sup>, Antonio Luigi Bisogno<sup>1,2</sup>, Alessandro Salvalaggio<sup>1,2</sup> and Maurizio Corbetta<sup>1,2,3\*</sup>

<sup>1</sup> Department of Neuroscience, Neurological Clinic, University of Padua, Padua, Italy, <sup>2</sup> Padova Neuroscience Center, University of Padua, Padua, Italy, <sup>3</sup> Department of Neurology, Radiology, and Neuroscience, Washington University in St. Louis, St. Louis, MO, United States

## OPEN ACCESS

### Edited by:

Maria V. Sanchez-Vives,  
August Pi i Sunyer Biomedical  
Research Institute (IDIBAPS), Spain

### Reviewed by:

Ines R. Violante,  
University of Surrey, United Kingdom  
Valentina Moro,  
University of Verona, Italy

### \*Correspondence:

Maurizio Corbetta  
maurizio.corbetta@unipd.it;  
mau@npg.wustl.edu

**Received:** 01 August 2019

**Accepted:** 30 March 2020

**Published:** 29 April 2020

### Citation:

Monai E, Bernocchi F, Bisio M,  
Bisogno AL, Salvalaggio A and  
Corbetta M (2020) Multiple Network  
Disconnection in Anosognosia  
for Hemiplegia.  
Front. Syst. Neurosci. 14:21.  
doi: 10.3389/fnsys.2020.00021

Anosognosia for hemiplegia (AHP) is a complex syndrome whose neural correlates are still under investigation. One hypothesis, mainly based on lesion mapping studies, is that AHP reflects a breakdown of neural systems of the right hemisphere involved in motor function. However, more recent theories have suggested that AHP may represent a disorder of cognitive systems involved in belief updating, self-referential or body processing. Two recent studies, using a method to estimate the degree of white matter disconnection from lesions, have indeed shown that patients with AHP suffer from damage of several long-range white matter pathways in association cortex. Here, we use a similar indirect disconnection approach to study a group of patients with motor deficits without anosognosia (hemiparesis or hemiplegia, HP,  $n = 35$ ), or motor deficits with AHP ( $n = 28$ ). The HP lesions came from a database of stroke patients, while cases of AHP were selected from the published literature. Lesions were traced into an atlas from illustrations of the publications using a standard method. There was no region in the brain that was more damaged in AHP than HP. In terms of structural connectivity, AHP patients had a similar pattern of disconnection of motor pathways to HP patients. However, AHP patients also showed significant disconnection of the right temporo-parietal junction, right insula, right lateral and medial prefrontal cortex. These associative cortical regions were connected through several white matter tracts, including superior longitudinal fasciculus III, arcuate, fronto-insular, frontal inferior longitudinal, and frontal aslant. These tracts connected regions of different cognitive networks: default, ventral attention, and cingulo-opercular. These results were not controlled for clinical variables as concomitant symptoms and other disorders of body representation were not always available for co-variate analysis. In conclusion, we confirm recent studies of disconnection demonstrating that AHP is not limited to dysfunction of motor systems, but involves a much wider set of large-scale cortical networks.

**Keywords:** anosognosia, hemiplegia, stroke, structural disconnection, network, awareness

## INTRODUCTION

*Anosognosia*, or lack of awareness of having a disorder or disability (Mograbi and Morris, 2018) represents an impressive phenomenon whose neural correlates have not been completely clarified.

*Anosognosia for hemiplegia (AHP)* refers to a syndrome in which a patient, typically following a stroke, fails to recognize his motor deficit. Patients may explain away the motor deficits with

other reasons (a fall, arthritis, unwillingness to move at this time, etc.). AHP after stroke is the most investigated form of anosognosia, and it is clinically relevant due to its negative impact on motor rehabilitation (Vocat et al., 2010). Other forms of anosognosia have been described: for visual stimuli (hemianopia and Anton's syndrome), spatial and body processing, and cognitive deficits (Goldenberg et al., 1995; Anton, 1898; Barrett et al., 2005; Spinazzola et al., 2008; Vossel et al., 2012; Ronchi et al., 2014; Baier et al., 2015).

Traditionally, anosognosia is considered an alteration of monitoring systems specific for the involved function. For instance, monitoring systems for movement in the case of AHP. This view is supported by clinical reports showing dissociation of awareness between different types of deficits in the same patient. For instance, a patient may ignore one type of deficit, but be aware of other types of deficits. Also anosognosia can be separated from general cognitive impairment (Bisiach et al., 1986; Berti et al., 1996; Spinazzola et al., 2008).

Motor control and movement awareness of intended motor acts have been proposed to reside within a frontal-parietal circuit (Haggard, 2005), involving mainly premotor cortex (PMC), supplementary motor area (SMA), posterior parietal cortex, and prefrontal cortex (PFC; Frith et al., 2000; Desmurget and Sirigu, 2009). These regions code intended, predicted, and actual movement states through feedforward and feedback signals.

Anatomo-clinical studies of AHP have reported damage of several fronto-temporo-parietal regions, as well as insula and subcortical regions. As an example, some studies have emphasized the importance of impaired sensory feedback with spared motor intentions or lack of a feedforward motor plan or an impairment of a motor “comparator” between an internal prediction and actual action (Heilman, 1991; Berti et al., 2005; Fotopoulou et al., 2008; Spinazzola et al., 2008; Heilman, 2014) involving premotor, sensory-motor regions, temporo-parietal junction (TPJ), basal ganglia, PFC, and insular cortex.

However, more recently, other theories have suggested that, at least for AHP, other systems beyond the motor system must be involved: e.g., systems involved in perspective taking (Besharati et al., 2016), reality checking and belief updating (Vuilleumier, 2004; Jenkinson et al., 2010; Vocat et al., 2013), and top-down and bottom-up processes on the prediction of the current state of the body (Fotopoulou, 2014). For instance, damage of the insula in AHP has been interpreted as a body representation and body schema disorder (Craig, 2002; Karnath, 2005; Craig, 2009). Beyond sensory-motor systems Vocat et al. (2013) have suggested a deficit of cognitive control mechanisms through damage of PFC and fronto-striatal circuits.

Recent advances in network neuroscience have emphasized the critical role of distributed interacting cognitive systems. Similarly, cognitive deficits in stroke have been associated with widespread network dysfunction, and white matter pathways damage Kang et al. (2003); Wessels et al. (2006); Boes et al. (2015); Corbetta et al. (2015); Ramsey et al. (2017). Similarly, for AHP a network view has been proposed (Fotopoulou, 2014) in which multiple cognitive system contribute in addition to the sensory-motor system to the syndrome.

Two recent studies of AHP are particularly relevant to our own work. Moro et al. (2016) identified using a voxel-based lesion analysis damage to several white matter regions near/at corona radiata, arcuate fasciculus and ventral superior longitudinal fasciculus, along with cortical damage of the rolandic operculum (ventral premotor cortex), insula, Heschl and superior temporal gyrus, and subcortically, basal ganglia. Persistent anosognosia chronically was associated with damage of fronto-temporal cortex (ventral premotor cortex, Heschl and temporal superior cortex), as well as long white matter tracts (cortico-spinal tract, anterior segment of fasciculus arcuate, and corpus callosum).

Pacella et al. (2019) used a structural disconnection approach (described below) to study network correlates of AHP. They found disconnection of white matter pathways that connected three sets of networks: a premotor loop, the limbic system, and the ventral attention network (VAN). In terms of pathways, AHP lesions disconnected primarily cingulate bundle (limbic), SLF III (VAN), and tracts connecting with pre-SMA such as frontal aslant and fronto-striatal (premotor loop).

Our study along the same lines aims to study the neural correlates of AHP both in terms of lesion topography and white matter disconnection. First, we searched the literature for reports on AHP in which the picture of the lesions were shown (see section “Materials and Methods” for search criteria). Next, we traced each lesion in an atlas space (MNI-152) using the approach of Boes et al. (2015) validated for functional connectivity in several recent publications (Laganier et al., 2016; Darby et al., 2017, 2018; Cohen et al., 2019). Neuropsychological information on these cases was incomplete, and could not be used for analysis. Third, as in Pacella et al. (2019), we used an indirect estimate of white matter disconnection obtained by embedding a patient's lesion into an atlas of white matter connections derived from healthy subjects. This approach, developed by Foulon et al. (2018), allows not only to estimate group-wide differences in structural disconnection, but also to identify the cortical regions corresponding to the disconnected white matter pathways. To identify AHP specific lesions or disconnection correlates we compared AHP patients with a group of stroke controls with significant motor impairment (hemiparesis or hemiplegia) from Corbetta et al. (2015) for whom the lesion volumes were available.

## MATERIALS AND METHODS

### Subjects

The electronic database MEDLINE (accessed through PubMed) was searched. The search was conducted in April 2018 and consisted of all the following terms related to several body-self disorders: */anosognosia, stroke//anosognosia, hemiplegia, stroke//asomatognosia stroke//misoplegia stroke/,out of body experience stroke//personal neglect stroke//somatoparaphrenia stroke/*. Seven hundred and nineteen articles were examined from which we included patients with anosognosia for left hemiplegia after right-hemisphere stroke and with available pictures of lesions in the report.

Twenty-eight subjects from 21 papers were included. Two subjects were not selected because of the poor quality of the published images. Among the selected articles, 20 described single cases, providing demographical and neurological scores for each patient, whereas one article included a group analysis with mean demographical and neurological scores, but included lesion imaging for each patient (Cappa et al., 1987; House and Hodges, 1988; Berti and Frassinetti, 2000; Tei, 2000; Morin et al., 2003; Venneri and Shanks, 2004; Turnbull et al., 2005; Fotopoulou et al., 2008, 2009, 2011; van Stralen et al., 2011; Coglianò et al., 2012; Venneri et al., 2012; Besharati et al., 2015; Di Vita et al., 2015; Moro et al., 2015; Piedimonte et al., 2015, 2016; Salvato et al., 2016; Facchin and Beschin, 2018).

Among the 28 selected subjects (mean age  $69.85 \pm 12$  years, range 51–93; 11 women, 11 men and 6 not specified, 14 lesions were documented with original CT scans, 6 lesions with original MRI, 6 with reconstruction on an MNI template, two of them in Damasio atlas spaces. All 28 subjects had single lesions, of which 12 ischemic, 7 hemorrhagic, while for 9 subjects the etiology was not specified.

A control group of first time stroke subjects with motor impairment, hemiparesis or hemiplegia (HP), was selected from the cohort ( $n = 132$ ) described in Corbetta et al. (2015). We selected subjects with an overall motor impairment greater than one standard deviation from control subjects based on a principal component analysis of motor scores across different tasks of strength, coordination and dexterity (see Corbetta et al., 2015 for how the overall motor score was calculated). There were 19 subjects with right hemisphere lesions and 16 subjects with left hemisphere lesions. Left hemisphere lesions were flipped to the right hemisphere for comparison with the AHP group that was comprised all of right hemisphere lesions. The control (HP) group in the end included  $n = 35$  subjects (mean age  $55.6 \pm 9.0$  years, range 38–83, 18 women, 17 men, Table 1). All 35 control subjects underwent an MRI protocol as specified in Corbetta et al. (2015).

Scanning was performed with a Siemens 3T Tim-Trio scanner at the School of Medicine of the Washington University in

St. Louis, and included sagittal MP-RAGE T1-weighted image. Lesion segmentation was described in Corbetta et al. (2015). Individual T1 MRI images were registered to the Montreal Neurological Institute brain using FSL (FMRIB Software Library) FNIRT (FMRIB non-linear imaging registration tool<sup>1</sup>). Lesions were manually segmented using the Analyze biomedical imaging software system<sup>2</sup> (Robb and Hanson, 1991).

More comprehensive demographical, clinical and radiological data of AHP patients and HP controls are reported in Tables 1, 2 and Supplementary Table 1.

## Drawing of the Lesions

Brain lesions were drawn onto a standard brain template from FSL<sup>3</sup> (MNI152 brain,  $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$ ) using itk-SNAP as lesion mapping software.<sup>4</sup> Lesions from published papers were traced manually in the 2D axial plane of the template brain using neuroanatomical landmarks for guidance. The template was re-oriented to fit the plan of the published picture to increase accuracy. The lesions were segmented by medical students (FB, ALB) and neurology residents (EM, AS) and were all checked by a board-certified neurologist (MC). In case of lesion with multiple slices they were traced in the same atlas space (see Supplementary Material). The drawn lesions were subsequently down-sampled to a  $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$  3D space<sup>5</sup> (MNI152 brain,  $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$ ). Lesions of the control group from Corbetta et al. (2015) were already available having been previously segmented and normalized to the MNI152 brain template<sup>6</sup> (MNI152 brain,  $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$ ). Lesion volume was calculated for each subject in the AHP and HP groups.

## Lesion-Based Analyses

The distribution of the lesions was assessed at the voxel-level in the two groups<sup>7</sup> (FSL Version 5.0.11). A two-sample unpaired

<sup>1</sup><http://www.fmrib.ox.ac.uk/analysis/techrep>

<sup>2</sup>[www.mayoclinic.org](http://www.mayoclinic.org)

<sup>3</sup><http://fsl.fmrib.ox.ac.uk/fsl/downloads>

<sup>4</sup><http://www.itk-snap.org/pmwiki/pmwiki.php?n=Downloads.SNAP3>

<sup>5</sup><http://fsl.fmrib.ox.ac.uk/fsl/downloads>

<sup>6</sup><http://fsl.fmrib.ox.ac.uk/fsl/downloads>

<sup>7</sup><http://www.fmrib.ox.ac.uk/fsl/>

**TABLE 1 |** Demographical, clinical and radiological data of AHP and HP patients.

Demographic, clinical and radiological data	AHP patients	HP patients
Selected subjects	28	35
Side of lesion R/L	28/0	19/16
Mean age (Year)	$69.85 \pm 12$	$55.6 \pm 9.0$
Sex F/M	11/11, 6 NS	17/18
CT scan	14	NS
MRI	6	35
Reconstruction (MNI/Damasio)	8	0
Mean imaging- delay after stroke (D)	$91.0 \pm 191.7$	$14.00 \pm 5.57$
Lesion Volume ( $\text{mm}^3$ )	$26610.8 \pm 28693.1$	$74334.4 \pm 84587.2$
Type of stroke I/H	12/7, 9 NS	25/8, 2 I + H

NS, not specified; I, ischemic; H, haemorrhagic; I + H, ischemic with haemorrhagic infarction.

**TABLE 2 |** Clinical features of AHP and HP patients.

Clinical feature	AHP patients	HP patients
Selected subjects	28	35
Sensory impairment	14	20
Visual impairment	12	8
Neglect	22	12
Somatoparaphrenia	7	0
Asomatognosia	7	0
Anton syndrome/anosognosia for visual impairment	2	0
Anosognosia for hemianesthesia	3	0
Others*	6	0

\*e.g., anosognosia for neglect, allochiria, anosodiaphoria, hyperkinetic motor behavior, alien hand, and misoplegia. 0 if absent or not specified.

*T*-test with lesion size as a co-variate was performed between AHP and HP groups applying the general linear model (GLM) and randomize function (with 5000 permutations; Winkler et al., 2014), tools from the FMRIB Software Library<sup>8</sup> (FSL Version 5.0.11) package (Smith et al., 2004). A threshold-free cluster enhancement approach was applied to control for family-wise error.

## Structural Disconnection Analyses (Disconnectome)

Disconnectome maps were calculated using BCBtoolkit (Foulon et al., 2018). In this approach a dataset of 164 diffusion weighted imaging data set from healthy controls from the Human Connectome Project (Rojkova et al., 2016) was used to track fibers passing through each lesion. For each participant tractography was estimated as described in De Schotten et al. (2011). The lesions in MNI152 space were registered to each control native space using affine and diffeomorphic deformations (Klein et al., 2009; Avants et al., 2011) and used as seed for the tractography in Trackvis (Wang et al., 2007). Tractography from the lesions were transformed in visitation maps (De Schotten et al., 2011), binarized, and brought to the MNI152 using the inverse of precedent deformations. Finally, a percentage overlap map was produced by summing at each point in MNI space the normalized visitation map of each healthy subject. Hence, in the resulting disconnectome map, the value of each voxel takes into account the inter-individual variability of tract reconstructions in controls and indicate a probability of disconnection from 0 to 1 for a given lesion (Thiebaut de Schotten et al., 2015). Then a threshold of 0.5 was applied, considering a tract involved when the probability to be present in a given voxel was estimated above 50% (Thiebaut De Schotten et al., 2014).

This procedure was replicated for all AHP and HP lesions, allowing the construction of a disconnectome map for each patient in both groups. These steps were automatized in the tool/disconnectome/as part of the BCBtoolkit. A two-sample unpaired *T*-test with lesion size as covariate between groups was carried out controlling for family-wise error using a threshold-free cluster enhancement approach.

This analysis allows to obtain a map (*disconnection-based map*) of the statistically significant group differences in structural disconnection.

A concomitant analysis with the same procedure was also conducted comparing AHP patients with  $n = 19$  right hemisphere only HP patients (see **Supplementary Materials**).

To examine the overlap of the disconnection maps with the normal white matter tracts, cortical parcels, or subcortical parcels, two measures of overlap were computed as follows:

$$\begin{aligned}\% \text{ disconnected tract} &= \frac{N_d \cap N_t}{N_t} \times 100 \\ \% \text{ disco/parcel overlap} &= \frac{N_d \cap N_t}{N_d} \times 100\end{aligned}$$

where  $N_d$  is the number of voxels belonging to the whole *disconnection-based map* and  $N_t$  the number of voxels belonging to each specific white matter tract/brain region/parcel, specified by  $t$ .

% Disco/parcel overlap and % disconnected tract refer, respectively, to the overlap of cortical parcels computed in relation to the total number of voxels composing the *disconnection-based map* and the total number of voxels involved by the disconnection belonging to specific white matter tract.

The variable  $t$  defines regions of interest in three atlases, specifically: (i) a white matter tracts atlas (Rojkova et al., 2016); (ii) a subcortical atlas composed of thalamus, caudate, putamen, pallidum, brain-stem, hippocampus, amygdala and accumbens (part of FSL); and (iii) the Gordon–Laumann parcellation for the cortical surface (Gordon et al., 2016). All Gordon–Laumann networks were analyzed, only those that contribute above 2% to the whole disconnection were selected.

## RESULTS

### Lesions and Disconnection Maps in AHP and HP Group

The distribution of lesions was similar in the two groups, involving frontal, temporal, and parietal lobes, as well as insular cortex and subcortical regions in the basal ganglia and thalamus. The brainstem was involved only in the HP group.

The center of damage in the HP group was in the basal ganglia and central white matter similarly to Corbetta et al. (2015). The center of damage in the AHP group was more diffuse with the most common locations of damage in the frontal and parietal white matter (**Figure 1**). Note that the distribution of lesions is uneven and discontinuous across the brain volume in the AHP group due to the uneven sampling of the published pictures (as described in section “Materials and Methods”).

We computed the structural connectivity disconnection map that was common for each group, by thresholding the group maps at 75% overlap, i.e., 75% of the patients showed the same disconnection at that location. Both groups of patients showed a similar disconnection of the dorsal white matter pathways descending from the motor/premotor/parietal cortex to the internal capsule, and cerebral peduncle. There is also involvement of frontal and temporal white matter tracts, thalamo-cortical and basal ganglia-cortical tracts (**Figure 2**).

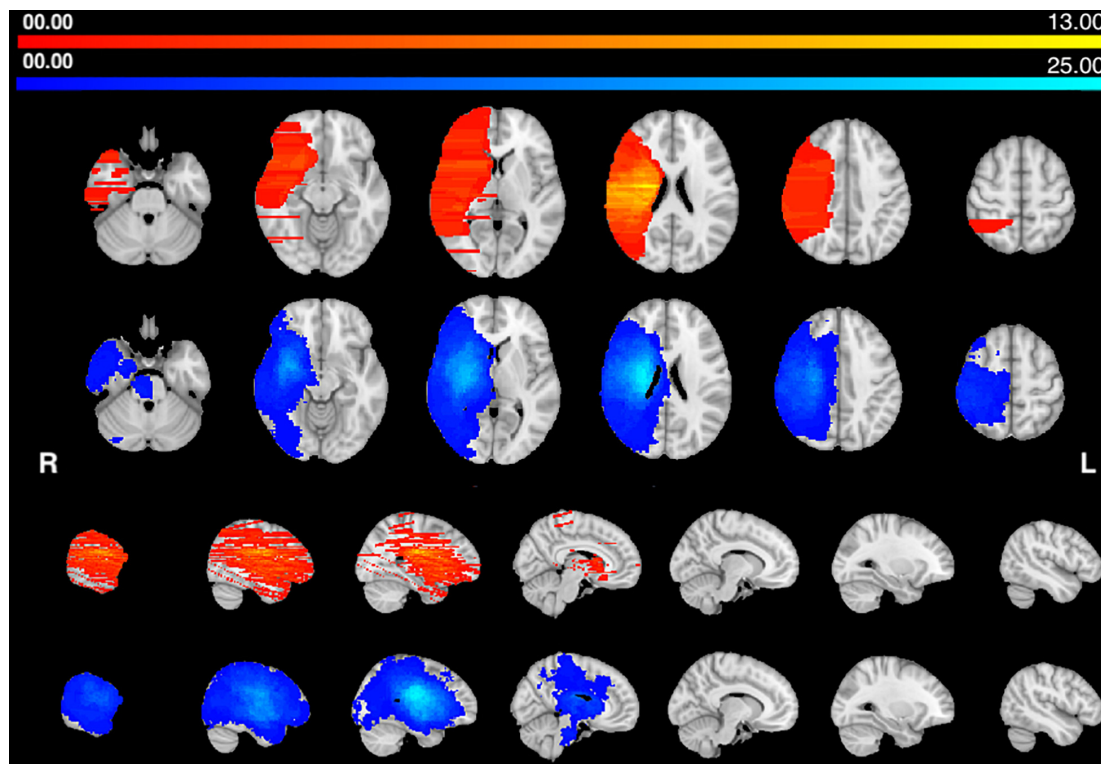
In summary, both groups have lesions located in the middle cerebral artery distribution, deep and superficial. They also show a common disconnection of descending motor pathways. In the HP group there is also a transcallosal (pre)-motor contralateral disconnection.

### AHP vs. HP Lesion vs. Disconnection Maps

Next, we statistically compared the two groups in terms of lesion location and volume ( $p < 0.05$  after correction for family-wise error). We did not find any significant difference in favor of the AHP group; however, there was significantly more damage in

<sup>8</sup><http://www.fmrib.ox.ac.uk/fsl/>





**FIGURE 1 |** Lesions frequency map: Voxel lesion overlap in AHP (orange-yellow) and HP groups (blue-teal). The color scale indicates the max number of patients with lesions in one voxel.

the dorsal corona radiata underlying the PMC in the HP group (Figure 3A).

In contrast, the structural disconnection analysis showed significant group differences ( $p < 0.05$  after correction for family-wise errors). HP patients showed greater disconnection of the contralateral dorsal white matter underlying premotor and motor cortex and cortico-spinal tracts (Figure 3B). The AHP patients showed greater disconnection that involved the white matter underlying the superior and inferior parietal lobule (SPL, IPL), superior and middle temporal gyrus (STG, MTG), PMC, insula, and anteriorly in the inferior frontal gyrus (IFG). There was also a slight bilateral disconnection of medial and orbito-frontal cortex through the anterior segment of the corpus callosum (genu; Figure 3B).

### Mapping Structural Disconnection Maps Onto White Matter Tracts, Subcortical Regions, and Cortical Networks

To identify the white matter tracts involved in the map of structural disconnection, we computed the overlap between the disconnection maps and an atlas of the white matter tracts. The tracts more affected in AHP included the superior longitudinal III and frontal inferior longitudinal fascicles (SLF III, FIL), the arcuate fascicle (AF, mainly the posterior segment, AP but also the anterior segment, AA), the fronto-insular tracts (FI), and fronto-aslant tract (FAT; Figures 4, 5A).

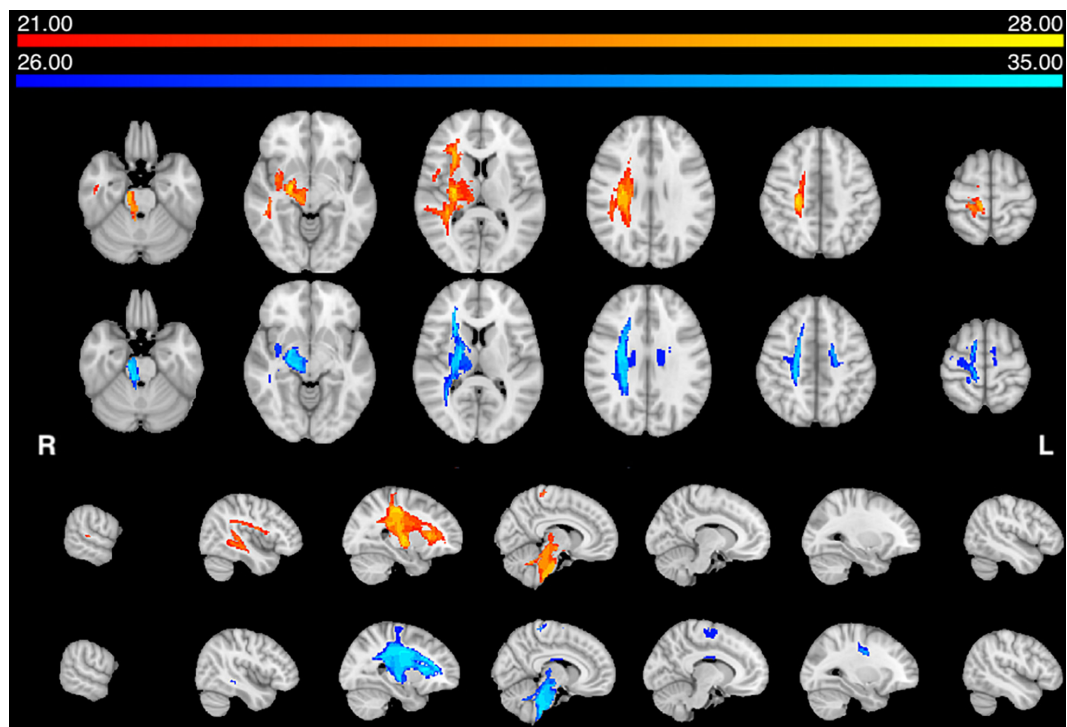
In contrast, in the HP group, the white matter tracts more affected localized almost to motor pathways (cortico-spinal, pons), transcallosal connections and superior longitudinal fasciculus I and II (SLF I-II; Figure 5B).

Finally, we analyzed the cortical projections of the affected tracts in terms of brain networks. Regions disconnected in the AHP group belonged mainly to the VAN, auditory, cingulo-opercular (CON) and default mode network (DMN; Figure 5C).

## DISCUSSION

### Methodological Issues and Limits of the Study

This study has many limitations. The first main limitation is that we drew manually the lesions in a template atlas from published pictures in the literature of different AHP cases and groups of patients. This approach was originally proposed by Boes et al. (2015) to investigate functional connectivity disconnection, what they call lesion symptom mapping (Fox, 2018), in rare syndromes or in patients in which direct measures of structural-functional connectivity are not available. We used these manually drawn lesions for a structural disconnection analysis, as proposed by Foulon et al. (2018). As far as we know, this is the first time these two methods have been combined, even though each method has already been separately applied.



**FIGURE 2 |** Structural disconnection frequency map (75%): Voxels of white matter tract overlap in AHP (orange–yellow) and HP (blue–teal) groups. The color scale indicates max number of patients per voxel.

In our hands, we first identified the best orientation of the atlas that would fit the published figure/s. This works best when multiple slices are displayed. Considerable uncertainty remains when only one slice is available. The pictures, first drawn by the medical students involved in the study (FB, ALB), were then checked by the neurology residents (EM, AS), and finally by a fully certified neurologist (MC). We did not perform formal test–retest analyses.

Usually published images of lesions include only one or few representative slides that do not capture the extent of the whole lesion. Therefore, lesion volume is on average underestimated. As a result, average lesion maps in the AHP group were discontinuous with gaps especially in the *z*-axis (**Figure 1**). This makes a statistical analysis at the group level problematic since the two groups do not uniformly cover the same brain volume. Moreover, on average, lesions in the HP group were larger in volume (**Table 1**). This difference is likely due to the differences in the way we identified the lesions, while in fact the opposite may be true. In fact lesions producing AHP (and hemiplegia) shall be larger, but also more variable in location than lesions producing HP only. The fact that AHP lesions, as drawn, are smaller and more variable shall favor the HP group in terms of lesion topography and disconnection. In fact, we found, for the HP group, more damage in the corona radiata, likely because of the more focal damage, and greater disconnection of motor/premotor/callosal connectivity in the central motor region. However, the disconnection for association pathways was greater in the AHP group suggesting a real biological effect.

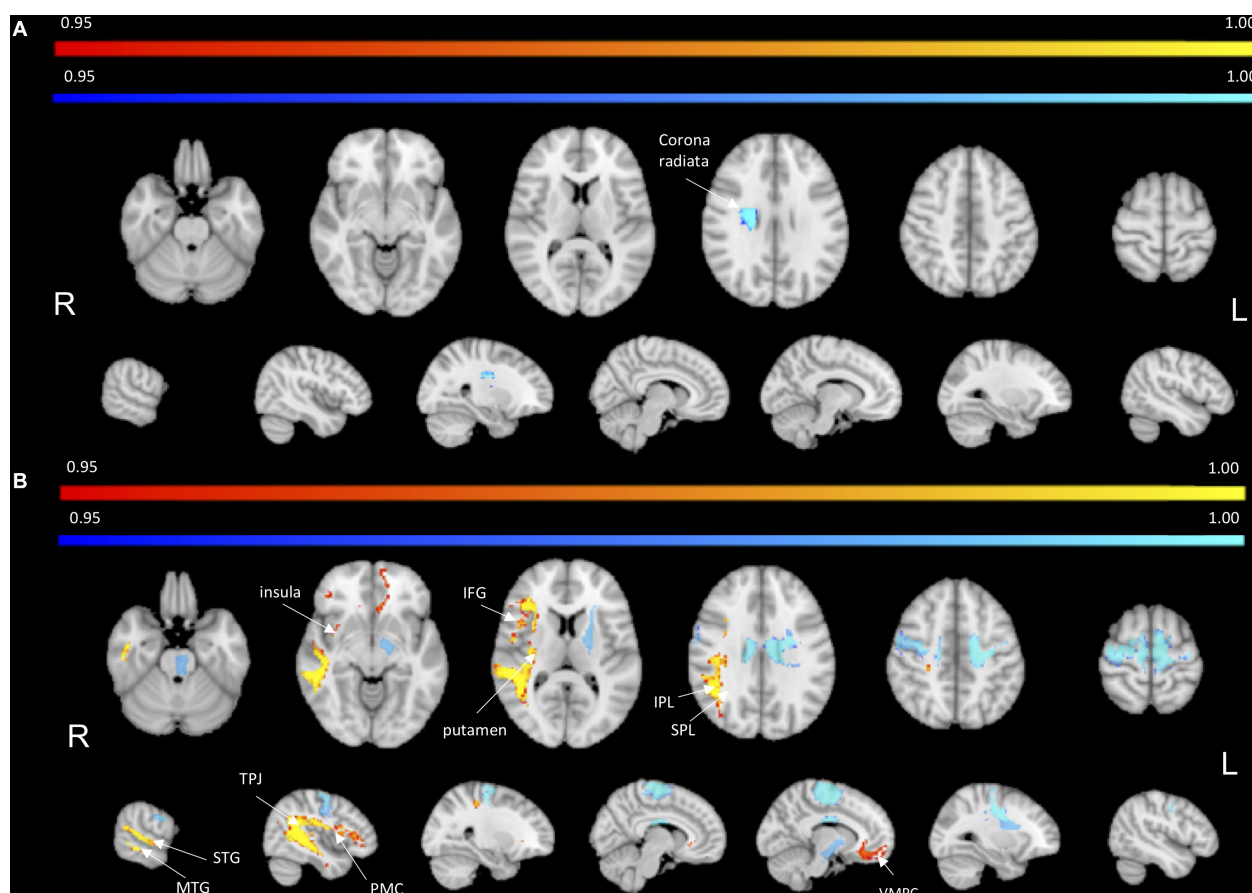
Another limitation is the scarcity of clinical information in the AHP group. The published reports we used for this investigation often did not contain detailed information on non-motor deficits, such as neglect symptoms. In addition, AHP is typically measured using clinical scales (e.g., Bisiach et al., 1986) that often were not even available in all reports. Therefore, it was not possible to perform any correlation with neuropsychological scores, also to covary out possible contribution from other deficits.

Some of AHP patients showed other forms of alteration of bodily awareness, e.g., asomatognosia, i.e., a loss of recognition or awareness of part of the body; somato-paraphrenia, i.e., the hallucinatory experience of attributing one own's arm to someone else or the subjective feeling of distortion or additional arms, and other forms of anosognosia (**Table 2**). So, a possible influence of the neural correlates of different syndromes and concomitant deficits with possible overlap or distinction could not be excluded from the performed analysis and represents a relevant limit of the present study.

Moreover, in the control group, anosognosia was not explicitly measured, even though it was reported by the expert clinical examiner of the trial. Most patients in the control group were examined at 2 weeks, a time in which AHP is disappearing.

Also, we cannot guarantee that the degree of motor deficit in AHP was the same as in the control group. For HP controls, we used a cut-off of  $z < -1$  that included both mild, moderate, and severe patients (mean *z* score  $-2.10$ , SD  $0.46$ ).

The timing of brain scans was not uniform between the two groups. It was consistently at two weeks in the



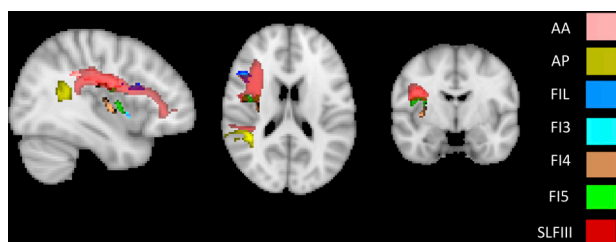
**FIGURE 3 |** Lesion vs. disconnection maps. **(A)** Lesion-based mapping: voxels with significant difference in lesion frequency between HP and AHP patients. HP > AHP ( $p < 0.05$ ) (blue-teal). No voxels for AHP > HP damage. **(B)** Disconnection-based maps: voxels showing significant difference in white matter disconnection. HP > AHP ( $p < 0.05$ ) (blue-teal). AHP > HP ( $p < 0.05$ ) (orange-yellow).

HP group, but was more variable in the AHP group (see **Supplementary Table 1**).

Finally, to increase the size of the control group we flipped the left side lesions onto the right hemisphere, hence abolishing potential hemispheric difference. However, a control analysis that compared directly AHP patients with  $n = 19$  right hemisphere

only HP patients yielded similar results (see **Supplementary Figures 2–5** and **Supplementary Tables 2, 3**).

We performed a group-comparison analysis, not a single subject analysis, after thresholding to 0.9, that is 90% of healthy subjects showed that tract. The disconnection of different tracts was computed both as% of the whole disconnection, an index which favors white matter tract that have larger volume, as well as% disconnection of different tract, an index that takes into account also small tracts. Issues of thresholds are relevant and is hard to account for the relative behavioral effect of complete vs. incomplete disconnection.

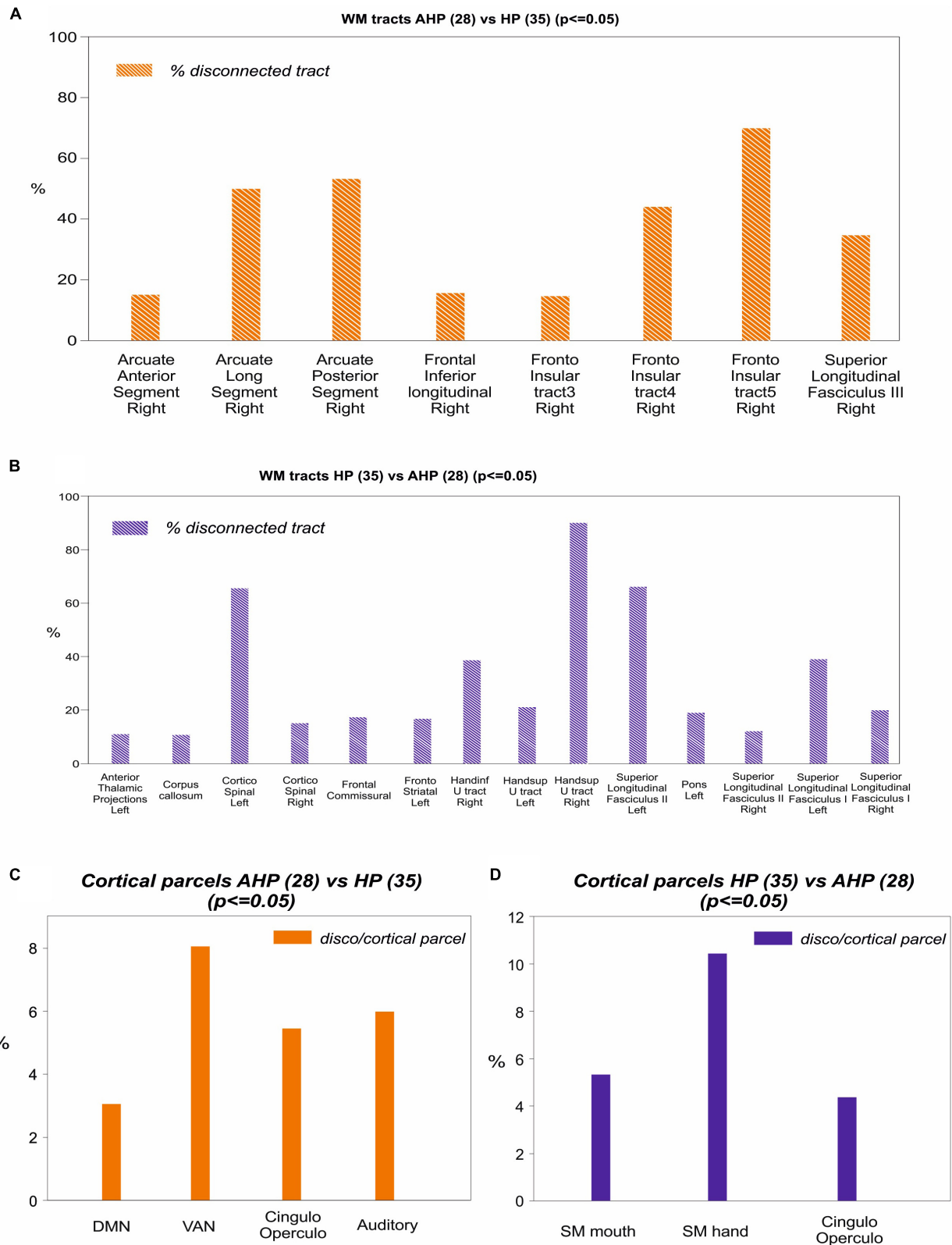


**FIGURE 4 |** White matter tracts more commonly disconnected in AHP: main white matter tracts involved in AHP disconnection map as compared to HP. FI: fronto-insular tract; SLF III: superior longitudinal fasciculus III; AA and AP: anterior and posterior segment of arcuate fasciculus; FIL: frontal inferior longitudinal fasciculus.

## A Network Perspective on AHP

Previous anatomical studies related AHP to lateral premotor and prefrontal, TPJ, insula, and basal ganglia lesions (Bisiach et al., 1986; Starkstein et al., 1992; Ellis and Small, 1997; Pia et al., 2004; Berti et al., 2005; Karnath, 2005; Vocat et al., 2010).

For instance a review by Pia et al. (2004, p. 367), who reviewed a large number of studies concluded: “It seems to be equally frequent when the damage is confined to frontal,



**FIGURE 5 |** Mapping of disconnection maps on white matter tracts and cortical networks. **(A)** AHP vs HP: individual white matter tracts' overlap with disconnectome map ( $> 10\%$  disconnected tract, orange bar). **(B)** Same for HP vs AHP (in blue). **(C)** AHP vs HP: % overlap  $> 2\%$  with cortical parcellation networks based on the Gordon–Laumann atlas (disco/cortical parcel, orange bar). **(D)** Same for HP vs AHP (in blue).



parietal or temporal cortical structures, and may also emerge as a consequence of subcortical lesions. Interestingly, the probability of occurrence of anosognosia is highest when the lesion involves parietal *and* frontal structures in combination, if compared to other combinations of lesioned areas. This pattern of lesions suggests the existence of a complex cortico-subcortical circuit underlying awareness of motor acts that, if damaged, can give rise to the anosognosic symptoms.”

Our distribution of lesions was compatible with the literature, and the center of damage was in the basal ganglia, insular and inferior frontal cortex and underlying fronto-temporal white matter tracts (**Figure 1**). However, the statistical analysis did not find any region that was differentially more affected in AHP than HP. In contrast, we found more frequent damage of the right dorsal corona radiata underlying motor and premotor regions in HP patients (**Figure 3A**). As noted above, this finding may be related to differences in the method used for lesion segmentation in the two groups. Interestingly, in the motor system, we also found some differences with greater disconnection of the corona radiata, and motor/premotor regions bilaterally in the HP group, consistent with the lesion results.

The most interesting result of this study was the difference in structural disconnection vis-a-vis a relative lack of difference in lesion location in favor of AHP. AHP patients showed a much more widespread pattern of white matter disconnection in the ipsilateral hemisphere reaching medial PFC bilaterally, lateral premotor and PFC, insula, TPJ, IPL and SPL, and putamen sub cortically (**Figure 3B**). The associated white matter tracts included SLF III connecting TPJ/IPL to ventral frontal cortex; arcuate fasciculus connecting temporal to frontal cortex; fronto-insular tracts connecting at short distance the insula with the inferior frontal lobe; and frontal inferior longitudinal fasciculus within the frontal lobe. The cortical networks most affected were association networks such as the VAN, CON and DMN, but also auditory (**Figures 4, 5**).

Herein, we will consider the possible contribution to AHP of differently disconnected regions/tracts within and beyond the motor system.

The insula is considered crucial for monitoring of internal body functions: the anterior part for human awareness of all feeling from the body containing not only interoceptive but also a somatotopic representation of subjective feeling of moving (Craig, 2009); while the posterior part for sense of ownership, agency and body schema (Karnath, 2005; Tsakiris et al., 2007; Karnath and Baier, 2010). The disconnection of fronto-insular tracts may mediate the abnormal body signal perception and awareness in AHP patients.

The inferior frontal gyrus includes nodes belonging to different networks such as the CON, a control network that includes insula and cingulate cortex and that is involved in task maintenance and shifting (Sestieri et al., 2014). Therefore, one aspect of AHP may be the inability to shift out of incorrect signals possibly related to motor planning or body schema.

Although the ventral premotor area was disconnected in AHP, pre-SMA and SMA were spared. This is in line with the observation of Berti et al. (2005) who claimed that AHP reflects

an inability to recognize sensory feedback notwithstanding a normal motor forward plan. However, damage to the CON and fronto-aslant tract does not rule a disconnection of pre-SMA and SMA.

The inferior frontal gyrus is also linked to the TPJ-supramarginal gyrus via SLFIII as part of the VAN. SLFIII was one of the most damaged white matter tracts. These regions/tracts may be relevant in AHP due to their role in integrating multi-modal body and visuospatial signals, and in switching between internal, bodily or self-perspective and external, or environmental, or others-perspective (Corbetta et al., 2008).

Regions in superior parietal and dorsal frontal cortex were also involved. These regions mediate top-down control signals for spatial attention (Corbetta et al., 2008). This damage may be related to the possible association of AHP with spatial neglect. Neglect symptoms were present indeed in a substantial part of our cohort of AHP patients (**Table 2**).

There is also disconnection of the angular gyrus and ventromedial PFC, bilaterally through the anterior corpus callosum. These are key nodes of the DMN, which might contribute to AHP through a number of internal processes including memory and self-referential behavior (Raichle et al., 2001) and emotional regulation (Raichle, 2015). Moreover, the bilateral involvement of ventromedial PFC could suggest that some aspects of anosognosia could not be related solely to right hemisphere dysfunction.

As previously described, for anosognosia a predominant role has always been recognized to the right hemisphere, nevertheless, this assumption may derive from its under-recognition in left hemisphere aphasic patients (Cocchini et al., 2009; Della Sala et al., 2009). Even though no left brain damaged AHP patient was present, we see evidence of bilateral prefrontal disconnection through the anterior corpus callosum.

Recently two studies have highlighted the role of white matter tracts and cortical regions around a wide fronto-temporo-parietal area. Moro et al. (2016) with a lesion voxels analysis on a large sample of AHP patients ( $n = 70$ ) found acute AHP to be associated with damage of ventral premotor cortex, insula and superior temporal gyrus, basal ganglia and white matter tracts as superior corona radiata, arcuate fasciculus and ventral part of SFL. Ventral premotor and superior temporal damage along with associated white matter tracts were present in chronic cases of anosognosia. We found similarly a lesion pattern and structural disconnection for AHP pointing to cortical areas such as ventral premotor cortex, insula and temporal regions, and among white matter tracts also arcuate fasciculus and SLFIII.

Pacella et al. (2019) used a similar structural disconnection approach on a cohort of 95 patients with AHP. Lesions were directly measured and there were available clinical and neuropsychological information for co-variate analyses. There are also differences between our analysis and Pacella et al. (2019) in the dataset of healthy subjects (164 vs 10), and the threshold (0.5 vs no threshold).

They found disconnection of white matter tracts like SLF III, frontal-aslant pathways and cingulum, and at the network

level disconnection of premotor/motor regions, limbic system, and VAN. Our analysis finds similar results: SLF III and FIL, AF, fronto-insular tracts, and fronto-aslant tract, with the most involved networks default mode (limbic), VAN, and cingulo-opercular that includes ventral premotor regions.

These studies jointly emphasize the neural correlates of AHP cannot be localized to a single brain region or functional system, e.g., motor-premotor. Rather, AHP as a syndrome involves several fronto-temporo-parietal areas and subcortical white matter, and multiple networks related to motor and body monitoring but also attention and self-referential processes. This in turn may connect with other theories on AHP as an aberrant predictive coding (Friston, 2005) due to a dynamic imbalance between prior beliefs, sensory feedbacks and prediction errors through large-scale networks (Fotopoulou, 2014).

One interesting question is whether this wide network disconnection is specific to AHP or partly generalizes to other forms of anosognosia. The current belief is that different forms of anosognosia are anatomically separate. It would be fascinating to find common anatomical areas or pathways with other forms, e.g., Anton or anosognosia for neglect.

Another interesting aspect that could be addressed in the future with functional studies of anosognosic patients are the temporal aspects of awareness. AHP is a fleeting syndrome that fluctuates over time. The factors controlling these fluctuations are unknown, but on-line monitoring of physiological activity through fMRI or EEG in parallel with behavioral assessments may show the dynamic loss of functional integration at moments of anosognosia.

## CONCLUSION

This study supports the hypothesis that AHP is a multicomponent network syndrome that includes multiple cognitive, as well as motor/premotor, networks.

## REFERENCES

- Anton, G. (1898). Ueber Herderkrankungen des Gehirnes, welche von Patienten selbst nicht wahrgenommen werden. *Wien. Klin. Wochenschr.* 11, 227–229.
- Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54, 2033–2044. doi: 10.1016/j.neuroimage.2010.09.025
- Baier, B., Geber, C., Müller-Forell, W., Müller, N., Dieterich, M., and Karnath, H. O. (2015). Anosognosia for obvious visual field defects in stroke patients. *Brain Struct. Funct.* 220, 1855–1860. doi: 10.1007/s00429-014-0753-5
- Barrett, A. M., Eslinger, P. J., Ballentine, N. H., and Heilman, K. M. (2005). Unawareness of cognitive deficit (cognitive anosognosia) in probable AD and control subjects. *Neurology* 64, 693–699. doi: 10.1212/01.wnl.0000151959.64379.1b
- Berti, A., Bottini, G., Gandola, M., Pia, L., Smania, N., Stracciari, A., et al. (2005). Neuroscience: shared cortical anatomy for motor awareness and motor control. *Science* 309, 488–491. doi: 10.1126/science.1110625
- Berti, A., and Frassinetti, F. (2000). When far becomes near: remapping of space by tool use. *J. Cogn. Neurosci.* 12, 415–420. doi: 10.1162/089892900562237
- Berti, A., Làdavas, E., and Della, C. M. (1996). Anosognosia for hemiplegia, neglect dyslexia, and drawing neglect: clinical findings and theoretical considerations. *J. Int. Neuropsychol. Soc.* 2, 426–440. doi: 10.1017/s135561770000151x
- Besharati, S., Forkel, S. J., Kopelman, M., Solms, M., Jenkinson, P. M., and Fotopoulou, A. (2016). Mentalizing the body: spatial and social cognition in anosognosia for hemiplegia. *Brain* 139(Pt 3), 971–985. doi: 10.1093/brain/awv390
- Besharati, S., Kopelman, M., Avesani, R., Moro, V., and Fotopoulou, A. (2015). Another perspective on anosognosia: self-observation in video replay improves motor awareness. *Neuropsychol. Rehabil.* 25, 1–34. doi: 10.1080/09602011.2014.923319
- Bisiach, E., Vallar, G., Perani, D., Papagno, C., and Berti, A. (1986). Unawareness of disease following lesions of the right hemisphere: anosognosia for hemiplegia and anosognosia for hemianopia. *Neuropsychologia* 24, 471–482. doi: 10.1016/0028-3932(86)90092-8
- Boes, A. D., Prasad, S., Liu, H., Liu, Q., Pascual-Leone, A., Caviness, V. S., et al. (2015). Network localization of neurological symptoms from focal brain lesions. *Brain* 138(Pt 10), 3061–3075. doi: 10.1093/brain/awv228
- Cappa, S., Sterzi, R., Vallar, G., and Bisiach, E. (1987). Remission of hemineglect and anosognosia during vestibular stimulation. *Neuropsychologia* 25, 775–782. doi: 10.1016/0028-3932(87)90115-1

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

MC and EM planned the research. EM, FB, AB, and AS reviewed the literature and drew the lesions. AS carried out the lesion-based and disconnection-based comparison analysis. MB carried out analyses related to white matter tract-overlap and cortical parcellation. EM, FB, AS, MB, and MC wrote the manuscript. MC provided supervision across the whole study.

## FUNDING

MC was supported by a starting grant of the University of Padua “FC-Neuro”, NIH RO1 NS095741, and ERA FLAG II: “Brainsynch-Hit”.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnsys.2020.00021/full#supplementary-material>

- Cocchini, G., Beschin, N., Cameron, A., Fotopoulou, A., and Della, S. S. (2009). Anosognosia for motor impairment following left brain damage. *Neuropsychology* 23, 223–230. doi: 10.1037/a0014266
- Cogliano, R., Crisci, C., Conson, M., Grossi, D., and Trojano, L. (2012). Chronic somatoparaphrenia: a follow-up study on two clinical cases. *Cortex* 48, 758–767. doi: 10.1016/j.cortex.2011.08.008
- Cohen, A. L., Soussand, L., Corrow, S. L., Martinaud, O., Barton, J. J. S., and Fox, M. D. (2019). Looking beyond the face area: lesion network mapping of prosopagnosia. *Brain* 142, 3975–3990. doi: 10.1093/brain/awz332
- Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324. doi: 10.1016/j.neuron.2008.04.017
- Corbetta, M., Ramsey, L., Callejas, A., Baldassarre, A., Hacker, C. D., Siegel, J. S., et al. (2015). Common behavioral clusters and subcortical anatomy in stroke. *Neuron* 85, 927–941. doi: 10.1016/j.neuron.2015.02.027
- Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* 3, 655–666. doi: 10.1038/nrn894
- Craig, A. D. (2009). How do you feel - now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70. doi: 10.1038/nrn2555
- Darby, R. R., Joutsma, J., Burke, M. J., and Fox, M. D. (2018). Lesion network localization of free will. *Proc. Natl. Acad. Sci. U.S.A.* 115, 10792–10797. doi: 10.1073/pnas.1814117115
- Darby, R. R., Laganier, S., Pascual-Leone, A., Prasad, S., and Fox, M. D. (2017). Finding the imposter: brain connectivity of lesions causing delusional misidentifications. *Brain* 140, 497–507. doi: 10.1093/brain/aww288
- De Schotten, M. T., Dell'Acqua, F., Forkel, S. J., Simmons, A., Vergani, F., Murphy, D. G. M., et al. (2011). A lateralized brain network for visuospatial attention. *Nat. Neurosci.* 14, 1245–1246. doi: 10.1038/nn.2905
- Della Sala, S., Cocchini, G., Beschin, N., and Cameron, A. (2009). Vata-m: visual-analogue test assessing anosognosia for motor impairment. *Clin. Neuropsychol.* 23, 406–427. doi: 10.1080/13854040802251393
- Desmurget, M., and Sirigu, A. (2009). A parietal-premotor network for movement intention and motor awareness. *Trends Cogn. Sci.* 13, 411–419. doi: 10.1016/j.tics.2009.08.001
- Di Vita, A., Palermo, L., Piccardi, L., and Guariglia, C. (2015). Peculiar body representation alterations in hemineglect: a case report. *Neurocase* 21, 697–706. doi: 10.1080/13554794.2014.974620
- Ellis, S., and Small, M. (1997). Localization of lesion in denial of hemiplegia after acute stroke. *Stroke* 28, 67–71. doi: 10.1161/01.str.28.1.67
- Facchin, A., and Beschin, N. (2018). Different impact of prism adaptation rehabilitation in spatial neglect and anosognosia for hemiplegia. *Ann. Phys. Rehabil. Med.* 61, 113–114. doi: 10.1016/j.rehab.2017.12.007
- Fotopoulou, A. (2014). Time to get rid of the “Modular” in neuropsychology: a unified theory of anosognosia as aberrant predictive coding. *J. Neuropsychol.* 8, 1–19. doi: 10.1111/jnp.12010
- Fotopoulou, A., Jenkinson, P. M., Tsakiris, M., Haggard, P., Rudd, A., and Kopelman, M. D. (2011). Mirror-view reverses somatoparaphrenia: dissociation between first- and third-person perspectives on body ownership. *Neuropsychologia* 49, 3946–3955. doi: 10.1016/j.neuropsychologia.2011.10.011
- Fotopoulou, A., Rudd, A., Holmes, P., and Kopelman, M. (2009). Self-observation reinstates motor awareness in anosognosia for hemiplegia. *Neuropsychologia* 47, 1256–1260. doi: 10.1016/j.neuropsychologia.2009.01.018
- Fotopoulou, A., Tsakiris, M., Haggard, P., Vagopoulou, A., Rudd, A., and Kopelman, M. (2008). The role of motor intention in motor awareness: an experimental study on anosognosia for hemiplegia. *Brain* 131(Pt 12), 3432–3442. doi: 10.1093/brain/awn225
- Foulon, C., Cerlini, L., Kinkingnehun, S., Levy, R., Rosso, C., Urbanski, M., et al. (2018). Advanced lesion symptom mapping analyses and implementation as BCToolkit. *Gigascience* 7, 1–17. doi: 10.1093/gigascience/giy004
- Fox, M. D. (2018). Mapping symptoms to brain networks with the human connectome. *N. Engl. J. Med.* 379, 2237–2245. doi: 10.1056/nejmra1706158
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 815–836.
- Frith, C. D., Blakemore, S. J., and Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philos. Trans. R. Soc. B Biol. Sci.* 355, 1771–1788. doi: 10.1098/rstb.2000.0734
- Goldenberg, G., Müllbacher, W., and Nowak, A. (1995). Imagery without perception-A case study of anosognosia for cortical blindness. *Neuropsychologia* 33, 1373–1382. doi: 10.1016/0028-3932(95)00070-j
- Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., and Petersen, S. E. (2016). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* 26, 288–303. doi: 10.1093/cercor/bhu239
- Haggard, P. (2005). Conscious intention and motor cognition. *Trends Cogn. Sci.* 9, 290–295. doi: 10.1016/j.tics.2005.04.012
- Heilman, K. M. (2014). Possible mechanisms of anosognosia of hemiplegia. *Cortex* 61, 30–42. doi: 10.1016/j.cortex.2014.06.007
- Heilman, M. (1991). “Anosognosia: possible neuropsychological mechanisms,” in *Awareness of Deficit after Brain Injury*, eds G. P. Prigatano, and D. L. Schacter (New York, NY: Oxford University Press), 371–384.
- House, A., and Hodges, J. (1988). Persistent denial of handicap after infarction of the right basal ganglia: a case study. *J. Neurol. Neurosurg. Psychiatry* 51, 112–115. doi: 10.1136/jnnp.51.1.112
- Jenkinson, P. M., Edeltyn, N. M. J., Drakeford, J. L., Roffe, C., and Ellis, S. J. (2010). The role of reality monitoring in anosognosia for hemiplegia. *Behav. Neurol.* 23, 241–243. doi: 10.1155/2010/572174
- Kang, D. W., Chalela, J. A., Ezzeddine, M. A., and Warach, S. (2003). Association of ischemic lesion patterns on early diffusion-weighted imaging with TOAST stroke subtypes. *Arch. Neurol.* 60, 1730–1734.
- Karnath, H.-O. (2005). Awareness of the functioning of one's own limbs mediated by the insular cortex? *J. Neurosci.* 25, 7134–7138. doi: 10.1523/jneurosci.1590-05.2005
- Karnath, H. O., and Baier, B. (2010). Right insula for our sense of limb ownership and self-awareness of actions. *Brain Struct. Funct.* 214, 411–417. doi: 10.1007/s00429-010-0250-4
- Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M. C., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46, 786–802. doi: 10.1016/j.neuroimage.2008.12.037
- Laganier, S., Boes, A. D., and Fox, M. D. (2016). Network localization of hemichorea-hemiballismus. *Neurology* 86, 2187–2195. doi: 10.1212/WNL.0000000000002741
- Mograbi, D. C., and Morris, R. G. (2018). Anosognosia. *Cortex* 103, 385–386.
- Morin, C., Durand, E., Marchal, F., Timsit, S., Manai, R., Pradat-Diehl, P., et al. (2003). Asomatognosie et troubles de l'oralité. Une lecture psychanalytique. *Ann. Readapt. Med. Phys.* 46, 12–23. doi: 10.1016/s0168-6054(02)00307-0
- Moro, V., Pernigo, S., Tsakiris, M., Avesani, R., Edeltyn, N. M. J., Jenkinson, P. M., et al. (2016). Motor versus body awareness: voxel-based lesion analysis in anosognosia for hemiplegia and somatoparaphrenia following right hemisphere stroke. *Cortex* 83, 62–77. doi: 10.1016/j.cortex.2016.07.001
- Moro, V., Scandola, M., Bulgarelli, C., Avesani, R., and Fotopoulou, A. (2015). Error-based training and emergent awareness in anosognosia for hemiplegia. *Neuropsychol. Rehabil.* 25, 593–616. doi: 10.1080/09602011.2014.951659
- Pacella, V., Foulon, C., Jenkinson, P., Scandola, M., Bertagnoli, S., Avesani, R., et al. (2019). Anosognosia for hemiplegia as a tripartite disconnection syndrome. *eLife* 8:e46075. doi: 10.7554/eLife.46075
- Pia, L., Neppi-Modona, M., Ricci, R., and Berti, A. (2004). The anatomy of anosognosia for hemiplegia: a meta-analysis. *Cortex* 40, 367–377. doi: 10.1016/s0010-9452(08)70131-x
- Piedimonte, A., Garbarini, F., Pia, L., Mezzanato, T., and Berti, A. (2016). From intention to perception: the case of anosognosia for hemiplegia. *Neuropsychologia* 87, 43–53. doi: 10.1016/j.neuropsychologia.2016.03.007
- Piedimonte, A., Garbarini, F., Rabuffetti, M., Pia, L., Montesano, A., Ferrarin, M., et al. (2015). Invisible grasps: grip interference in anosognosia for hemiplegia. *Neuropsychology* 29, 776–781. doi: 10.1037/neu0000182
- Raichle, M. E. (2015). The brain's default mode network. *Annu. Rev. Neurosci.* 38, 433–447.
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682.
- Ramsey, L. E., Siegel, J. S., Lang, C. E., Strube, M., Shulman, G. L., and Corbetta, M. (2017). Behavioural clusters and predictors of performance during recovery from stroke. *Nat. Hum. Behav.* 1:0038. doi: 10.1038/s41562-016-0038
- Robb, R. A., and Hanson, D. P. (1991). A software system for interactive and quantitative visualization of multidimensional biomedical images. *Australas Phys. Eng. Sci. Med.* 14, 9–30.
- Rojkova, K., Volle, E., Urbanski, M., Humbert, F., Dell'Acqua, F., and Thiebaut de Schotten, M. (2016). Atlas of the frontal lobe connections and their variability

- due to age and education: a spherical deconvolution tractography study. *Brain Struct. Funct.* 221, 1751–1766. doi: 10.1007/s00429-015-1001-3
- Ronchi, R., Bolognini, N., Gallucci, M., Chiapella, L., Algeri, L., Spada, M. S., et al. (2014). (Un)awareness of unilateral spatial neglect: a quantitative evaluation of performance in visuo-spatial tasks. *Cortex* 61, 167–182. doi: 10.1016/j.cortex.2014.10.004
- Salvato, G., Gandola, M., Veronelli, L., Agostoni, E. C., Sberna, M., Corbo, M., et al. (2016). The spatial side of somatoparaphrenia: a case study. *Neurocase* 22, 154–160. doi: 10.1080/13554794.2015.1077257
- Sestieri, C., Corbetta, M., Spadone, S., Romani, G. L., and Shulman, G. L. (2014). Domain-general signals in the cingulo-opercular network for visuospatial attention and episodic memory. *J. Cogn. Neurosci.* 26, 551–568. doi: 10.1162/jocn\_a\_00504
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23(Suppl. 1), S208–S219.
- Spinazzola, L., Pia, L., Folegatti, A., Marchetti, C., and Berti, A. (2008). Modular structure of awareness for sensorimotor disorders: evidence from anosognosia for hemiplegia and anosognosia for hemianaesthesia. *Neuropsychologia* 46, 915–926. doi: 10.1016/j.neuropsychologia.2007.12.015
- Starkstein, S. E., Fedoroff, P. J., Price, T. R., Leiguarda, R., and Robinson, R. G. (1992). Anosognosia in patients with cerebrovascular lesions a study of causative factors. *Stroke* 23, 1446–1453. doi: 10.1161/01.str.23.10.1446
- Tei, H. (2000). Right ipsilateral hypersensation in a case of anosognosia for hemiplegia and personal neglect with the patient's subjective experience. *J. Neurol. Neurosurg. Psychiatry* 69, 274–275. doi: 10.1136/jnnp.69.2.274
- Thiebaut de Schotten, M., Dell'Acqua, F., Ratiu, P., Leslie, A., Howells, H., Cabanis, E., et al. (2015). From phineas gage and monsieur Leborgne to H.M.: revisiting disconnection syndromes. *Cereb. Cortex* 25, 4812–4827. doi: 10.1093/cercor/bhv173
- Thiebaut De Schotten, M., Tomaiuolo, F., Aiello, M., Merola, S., Silvetti, M., Lecce, F., et al. (2014). Damage to white matter pathways in subacute and chronic spatial neglect: a group study and 2 single-case studies with complete virtual “in vivo” tractography dissection. *Cereb. Cortex* 24, 691–706. doi: 10.1093/cercor/bhs351
- Tsakiris, M., Hesse, M. D., Boy, C., Haggard, P., and Fink, G. R. (2007). Neural signatures of body ownership: a sensory network for bodily self-consciousness. *Cereb. Cortex* 17, 2235–2244. doi: 10.1093/cercor/bhl131
- Turnbull, O. H., Evans, C. E. Y., and Owen, V. (2005). Negative emotions and anosognosia. *Cortex* 41, 67–75. doi: 10.1016/s0010-9452(08)70179-5
- van Stralen, H. E., van Zandvoort, M. J. E., and Dijkerman, H. C. (2011). The role of self-touch in somatosensory and body representation disorders after stroke. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 3142–3152. doi: 10.1098/rstb.2011.0163
- Venneri, A., Pentore, R., Cobelli, M., Nichelli, P., and Shanks, M. F. (2012). Translocation of the embodied self without visuospatial neglect. *Neuropsychologia* 50, 973–978. doi: 10.1016/j.neuropsychologia.2012.02.004
- Venneri, A., and Shanks, M. F. (2004). Belief and awareness: reflections on a case of persistent anosognosia. *Neuropsychologia* 42, 230–238. doi: 10.1016/s0028-3932(03)00171-4
- Vocat, R., Saj, A., and Vuilleumier, P. (2013). The riddle of anosognosia: Does unawareness of hemiplegia involve a failure to update beliefs? *Cortex* 49, 1771–1781. doi: 10.1016/j.cortex.2012.10.009
- Vocat, R., Staub, F., Stroppini, T., and Vuilleumier, P. (2010). Anosognosia for hemiplegia: a clinical-anatomical prospective study. *Brain* 133, 385–386. doi: 10.1093/brain/awq297
- Vossel, S., Weiss, P. H., Eschenbeck, P., Saliger, J., Karbe, H., and Fink, G. R. (2012). The neural basis of anosognosia for spatial neglect after stroke. *Stroke* 43, 1954–1956. doi: 10.1161/STROKEAHA.112.657288
- Vuilleumier, P. (2004). Anosognosia: the neurology of beliefs and uncertainties. *Cortex* 40, 9–17. doi: 10.1016/s0010-9452(08)70918-3
- Wang, R., Benner, T., Sorensen, A. G., and Wedeen, V. J. (2007). Diffusion toolkit: a software package for diffusion imaging data processing and tractography. *Proc. Intl. Soc. Mag. Reson. Med.* 15:3720.
- Wessels, T., Wessels, C., Ellsiespen, A., Reuter, I., Trittman, S., Stolz, E., et al. (2006). Contribution of diffusion-weighted imaging in determination of stroke etiology. *Am. J. Neuroradiol.* 27, 35–39.
- Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage* 92, 381–397. doi: 10.1016/j.neuroimage.2014.01.060

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Monai, Bernocchi, Bisio, Bisogno, Salvalaggio and Corbetta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Perspective on the Multiple Pathways to Changing Brain States

Malinda L. S. Tantirigama, Timothy Zolnik, Benjamin Judkewitz,  
Matthew E. Larkum\* and Robert N. S. Sachdev\*

Institut für Biologie, NeuroCure Center for Excellence, Charité Universitätsmedizin Berlin & Humboldt Universität,  
Berlin, Germany

## OPEN ACCESS

### Edited by:

Maria V. Sanchez-Vives,  
August Pi i Sunyer Biomedical  
Research Institute (IDIBAPS), Spain

### Reviewed by:

Mario Rosanova,  
University of Milan, Italy  
Francisco Clasca,  
Autonomous University of Madrid,  
Spain

### \*Correspondence:

Matthew E. Larkum  
matthew.larkum@gmail.com  
Robert N. S. Sachdev  
bs387ster@gmail.com

**Received:** 23 December 2019

**Accepted:** 06 April 2020

**Published:** 08 May 2020

### Citation:

Tantirigama MLS, Zolnik T,  
Judkewitz B, Larkum ME and  
Sachdev RNS (2020) Perspective on  
the Multiple Pathways to Changing  
Brain States.  
*Front. Syst. Neurosci.* 14:23.  
doi: 10.3389/fnsys.2020.00023

In this review article, we highlight several disparate ideas that are linked to changes in brain state (i.e., sleep to arousal, Down to Up, synchronized to de-synchronized). In any discussion of the brain state, we propose that the cortical pyramidal neuron has a central position. EEG recordings, which typically assess brain state, predominantly reflect the activity of cortical pyramidal neurons. This means that the dominant rhythmic activity that characterizes a particular brain state ultimately has to manifest globally across the pyramidal neuron population. During state transitions, it is the long-range connectivity of these neurons that broadcast the resultant changes in activity to many subcortical targets. Structures like the thalamus, brainstem/hypothalamic neuromodulatory systems, and respiratory systems can also strongly influence brain state, and for many decades we have been uncovering bidirectional pathways that link these structures to state changes in the cerebral cortex. More recently, movement and active behaviors have emerged as powerful drivers of state changes. Each of these systems involve different circuits distributed across the brain. Yet, for a system-wide change in brain state, there must be a collaboration between these circuits that reflects and perhaps triggers the transition between brain states. As we expand our understanding of how brain state changes, our current challenge is to understand how these diverse sets of circuits and pathways interact to produce the changes observed in cortical pyramidal neurons.

**Keywords:** pyramidal neuron, brain state, cortico-thalamocortical, neuromodulation, active behavior, respiration

## INTRODUCTION

“Brain state,” particularly as it relates to consciousness, is usually recorded *via* EEG electrodes that predominantly measure the rhythmic activity of the cerebral cortex. Nevertheless, many circuits distributed across the brain contribute to and are related to the transition in brain states, and in this sense, the rhythmic activity of the cortex is a function of its bidirectional interaction with a host of subcortical structures. We know that changes in alertness, attention, arousal, and consciousness are associated with specific changes in levels and patterns of neuronal activity throughout the brain (Steriade and McCarley, 2005; Harris and Thiele, 2011). Explaining why and how the brain is in a given state at any given time is a formidable task. Like blind men considering the elephant (Ireland, 2007), this task has traditionally been approached from very different perspectives and has left the literature fragmented. A popular view is that system-wide transitions in brain state arise from ascending neuromodulatory input from the reticular activating system (reviewed in, Steriade and McCarley, 2005). Here, we offer a different

perspective. We bring together several disparate ideas and discuss: (1) specific and non-specific thalamic nuclei; (2) neuromodulation and hypothalamus; (3) behavior; and (4) respiration as paradigms for explaining brain state transitions. Although each paradigm involves very different systems, weaving them together are the cortical pyramidal neurons. We propose that both the effect of input to the cortical pyramidal neurons and the effect of their output, especially the layer 5 pyramidal neurons, are key to understanding system-wide brain state transitions.

## CORTEX, THALAMUS AND THE LOOP THAT LINK THEM TO BRAIN STATES

There are two overriding reasons to focus on pyramidal neurons. Firstly, the brain state as detected by the EEG signal is almost entirely determined by the electrical field dipoles across the apical dendritic axis of pyramidal neurons (Buzsáki et al., 2012; Larkum, 2013; Suzuki and Larkum, 2017). But even more importantly, pyramidal neurons make up the majority (80%) of all cortical neurons and are the principal long-range cortico-cortically and subcortically projecting neurons in the brain (DeFelipe and Fariñas, 1992; Molyneaux et al., 2007). Their aggregate activity is therefore central to both the local manifestation of the brain state and the causal explanation for system-wide transitions between brain states.

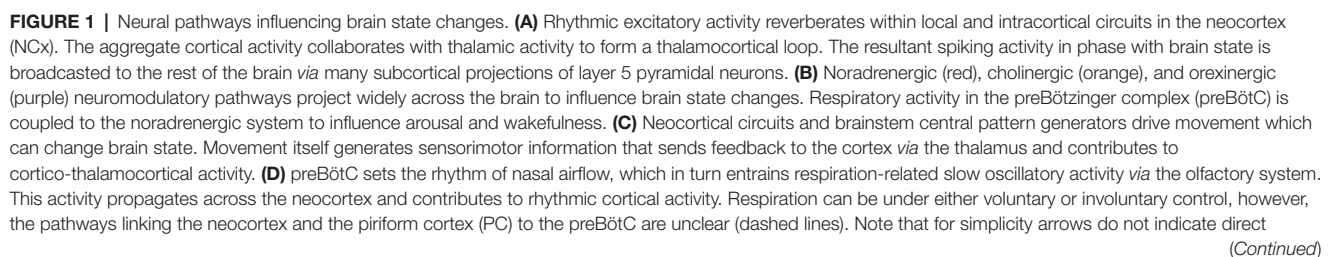
The elaborate long-range connectivity of layer 5 pyramidal neurons means that when cortical activity does change, activity in many subcortical areas follows the cortical rhythm. This effect of cortical activity is especially easy to monitor during slow-wave sleep, when activity spanning the entire cortical mantle fluctuates, and when sensory input and movement is minimal. When cortical activity fluctuates, the subcortical activity also shows phasic slow rhythmic activity in most targets of cortical axons (direct and multisynaptic) including thalamus (Steriade et al., 1993a,b; Contreras and Steriade, 1995; Contreras et al., 1996; Sherman and Guillery, 2011); striatum (Cowan and Wilson, 1994; Wilson and Kawaguchi, 1996); zona incerta (Barthó et al., 2007); subthalamus (Magill et al., 2004); globus pallidus (Magill et al., 2000; Tseng et al., 2001; Kasanetz et al., 2002); inferior olive (Rowland et al., 2010); locus coeruleus (Sara and Hervé-Minvielle, 1995); cholinergic pedunculopontine neurons (Roš et al., 2010; Motelow et al., 2015; Petzold et al., 2015); cerebellum and deep cerebellar nuclei (Ros et al., 2009; Rowland et al., 2010; **Figure 1A**). Most targets of pyramidal neurons show activity that is correlated with cortical slow waves, suggesting that in the quietly sitting animal, when input from the sensory world is negligible, these axons can drive changes in membrane potential, local field potentials, and even action potentials.

The thalamus is intimately connected to the cortex (**Figure 1A**): it is central to the thalamocortical loop, gates sensory input, and as mentioned above its activity fluctuates with cortical activity. However, the thalamus is not a unitary entity. For areas like the somatosensory cortex, multiple thalamic nuclei can affect cortical activity, including the ventroposterior medial

(VPM), the posterior medial nucleus (POm), and the midline intra-laminar thalamic nuclei (Sherman and Guillery, 2011; Saalmann, 2014). Neurons in these nuclei also receive cortical input and project back to the cortex. The close bidirectional connection with the cortex and the heterogeneity of the thalamus suggests that at least some thalamic inputs could trigger fast changes in the cortical state (Steriade, 2006)—and so argue against a cortico-centric view. In contrast to the original studies that ablated the thalamus (Gloor et al., 1977; Steriade et al., 1993c; Timofeev et al., 2000), subsequent loss-of-function studies have shown that thalamic inactivation (i.e., pharmacologically or lesioning inputs to the thalamus) can deactivate the persistent depolarization in the cortex and enhance slow cortical rhythms (Aguilar et al., 2010; Poulet et al., 2012; Zagha et al., 2013). Furthermore, whisker stimuli working through the thalamus or direct activation of the thalamus can trigger changes in cortical state (Petersen et al., 2003; Sachdev et al., 2004; Hasenstaub et al., 2007; Poulet et al., 2012; Poulet and Crochet, 2019).

The work mentioned above does not fully capture the role of the thalamus in modulating, triggering or maintaining changes in the brain-state. A substantial amount of recent evidence implicates a variety of midline thalamic nuclei in changing brain states globally, primarily in awakening the brain. Activating cholinergic input to reticular thalamus, which inhibits the reticular neurons, promotes slow-wave sleep (Ni et al., 2016), but remarkably direct inhibition of the reticular thalamus does not have the same effect. Neurons in the paraventricular thalamus increase their firing before mice wake up and optogenetic inhibition of these neurons lowers epochs of wakefulness. These effects of the paraventricular thalamic neurons are mediated by the nucleus accumbens pathway, not the prefrontal cortex pathway (Ren et al., 2018). The ventromedial thalamus, in particular the calbindin-positive layer 1 projecting matrix neurons, increase their firing before behavioral arousal and before changes in the EEG. The activity of these neurons can trigger arousal from slow-wave sleep or anesthesia (Honjoh et al., 2018). In the whisker system, POm and VPM neurons increase firing during active states, and firing specifically in the VPM neurons (compared with POm) is more correlated with whisking behavior (Urbain et al., 2015). Recently, similar findings have been extended to sleeping mice, where spikes of VPM neurons are correlated with a stronger impact on the cortical LFP than spikes of POm neurons during NREM sleep (Urbain et al., 2019). In non-human primates, stimulation of the central lateral thalamus, a higher-order thalamic nucleus, has recently been shown to drive lamina specific (in infragranular cortical layers) changes in cortical states (Redinbaugh et al., 2020).

Taken together, these heterogeneous thalamic nuclei have a role in changing cortical activity. However, for this activity to be manifested as a change in brain state the corticothalamic loop must be engaged to broadcast any effect globally to the rest of the brain (Aru et al., 2019). These effects of the thalamus on arousal, on changing cortical state also has to occur in concert with inputs from the hypothalamus (hypocretin/orexinergic) and from the locus coeruleus (noradrenergic), both of which can



(Continued)

**FIGURE 1 |** Continued

connectivity, and the line thickness does not indicate connection strength. BF, basal forebrain; Cb, cerebellum; Hip, hippocampus; CL, central lateral thalamus; IO, inferior olive; LC, locus coeruleus; LH, lateral hypothalamus; OB, olfactory bulb; PFC, prefrontal cortex; pFRG, parafacial respiratory group; Po, posterior medial thalamic nucleus; PPT, pedunculopontine nucleus; PV, paraventricular thalamic nucleus; Rt, reticular thalamic nucleus; STn, subthalamic nucleus, Str, striatum; Thl, thalamus; VM, ventromedial thalamic nucleus; VPM, ventroposterior medial thalamic nucleus; vRT, vibrissal reticular formation; ZI, zona incerta.

trigger arousal (Adamantidis et al., 2007; Carter et al., 2010; Eban-Rothschild et al., 2018).

## NEUROMODULATION AND THE HYPOTHALAMUS

By the middle of the last century, it was increasingly clear that structures in the brainstem have a substantial role in the modulation of cortical states (Jones, 2008, 2020). Early experiments of Moruzzi and Magoun (1949) showed that electrical stimulation of the brainstem i.e., the reticular activating system, and other subcortical structures could abolish cortical synchrony in the EEG, replacing it with low amplitude fast activity. Lesioning some of the same regions abolished fast EEG activity (Lindsley et al., 1949). We now understand that many of the deep structures in the brainstem, midbrain, and hypothalamus are part of a large, interconnected neuromodulatory network that acts both directly and indirectly on the cortex to change brain states (Figure 1B). Collectively, these structures, which include the locus coeruleus, pedunculopontine nucleus, tuberomammillary nucleus, and others constitute the reticular activating system (reviewed in Jones, 2008, 2020).

Each structure is associated with a particular neuromodulatory transmitter, and stimulation of each of these structures can induce cortical state changes. The ascending cholinergic projections from the brainstem, for example, when stimulated optogenetically, produce powerful changes in brain state, increasing gamma and reducing slow wave rhythms in the cortex; and optogenetic stimulation of the locus coeruleus causes transitions from sleep to wake (Carter et al., 2010; Scammell et al., 2017; Cissé et al., 2018). Many of these subcortical and brainstem nuclei contain diverse neuronal types and transmit multiple transmitters, and therefore the roles that each structure play are still being updated. For example, the basal forebrain, known for its cholinergic neurons, also contains cortical projecting GABAergic neurons that powerfully affect brain states and initiate cortical gamma-band oscillations (Anacleit et al., 2015; Kim et al., 2015; Xu et al., 2015). In the locus coeruleus phasic but not tonic activation differentially encodes saliency in the cortex (Vazey et al., 2018), further highlighting the complexity of these neuromodulatory systems.

A more recent addition to the neuromodulatory subcortical network is the orexin/hypocretin neurons, which were discovered in 1998 by two groups at the same time (Sakurai et al.,

1998; de Lecea et al., 1998). Orexin/hypocretin neurons reside in the lateral hypothalamus and, like other neuromodulatory structures, send projections far and wide across the brain, including to most of the other neuromodulatory centers in the brain, with the densest innervation to the locus coeruleus, which strongly promotes wakefulness (Peyron et al., 1998; Scammell et al., 2017).

Many recent studies suggest that the orexin/hypocretin system functions to coordinate and potentiate other neuromodulatory systems that stabilize arousal (Alexandre et al., 2013). Orexin/hypocretin neurons were found to be the underlying cause of the sleep/wake disorder, narcolepsy (Chemelli et al., 1999; Peyron et al., 2000; Thannickal et al., 2000), strongly implicating orexin/hypocretin neurons in cortical state changes and arousal (de Lecea, 2015). Optogenetic and chemogenetic stimulation of these neurons promotes the transition to wakefulness (Adamantidis et al., 2007; Sasaki et al., 2011), whereas their inhibition promotes sleep (Tsunematsu et al., 2013). Orexin/hypocretin neurons also innervate the thalamus and neocortex. But unlike other cortical projecting neuromodulatory centers that diffusely innervate and affect a wide range of neurons across many layers of the neocortex, orexin/hypocretin neurons appear to exclusively target deep layer 6 (layer 6b). Application of orexin/hypocretin to cortical brain slices specifically depolarizes these deep-layer 6b neurons (Bayer et al., 2004; Wenger Combremont et al., 2016).

Although the indirect but powerful effects of orexin/hypocretin on arousal are clear, the direct pathway linking orexinergic activation to cortical activity has not been studied. Since deep-layer 6 neurons are strongly and specifically activated by orexin/hypocretin, these cortical neurons might help induce the arousal effects of orexinergic neurons. Deep-layer 6 neurons were recently shown to project to cortical and thalamic regions commonly associated with cortical arousal, including neocortical layer 5 and the non-specific thalamic nuclei (Viswanathan et al., 2017; Hoerder-Suabedissen et al., 2018). However, unlike layer 6a which is a key component of the cortico-thalamocortical loop, the deeper sublayer 6b is best driven by cortico-cortical projection pathways, suggesting two separate pathways that interact differently with the thalamus (Zolnik et al., 2020). These neurons are genetically diverse and the exact relationship between orexin/hypocretin sensitivity and their projections is not yet worked out. Nevertheless, it is interesting to note that many of the thalamic structures that deep layer 6 targets are also orexin/hypocretin sensitive (Bayer et al., 2002). For example, the thalamic neurons are targets of deep layer 6 and as noted above is a powerful driver and regulator of cortical state; and many of the other subcortical neuromodulatory centers, including the pedunculopontine nucleus, project to non-specific thalamic nuclei and exert their influence on cortex by thalamocortical projections. The role of the lateral hypothalamus in stabilizing cortical arousal might be assisted by deep layer 6 neurons that act both on the ascending thalamocortical projections and directly on layer 5 cortical pyramidal neurons that have numerous long-range intracortical targets.



Orexin/hypocretin neurons also make descending projections and target several key structures in the brainstem. Among these is the respiratory rhythm generator, pre-Bötzinger nucleus in the medulla, which contains neurons that are excited by orexin/hypocretin application (Young et al., 2005). Breathing is an active mechanism, that is in part modulated by the action of hypothalamic neurons. Breathing is often correlated with brain state (see below). Injection of orexin/hypocretin into the pre-Bötzinger nucleus increases inspiratory tidal volume, and knock-out of orexin/hypocretin receptor gene in mice reduces the CO<sub>2</sub> induced increases in breathing and increases the occurrence of sleep apnea (Young et al., 2005). Taken together, these findings show that neuromodulatory circuits can interdependently promote cortical state changes in a complex and diverse manner.

## BEHAVIOR, ACTIVE SENSATION AND BRAIN STATES

“The behaving organism embedded in a particular environment is what generates feelings” (Koch, 2004, p. 9). In the “enactive or sensorimotor” account of consciousness, the body (i.e., behavior) is closely linked to conscious perception. While this link is not explicit and can be easily argued against (see Koch, 2004), the link between movement and changes in brain state is clear. In quietly sitting, but awake mice, cortical activity is often dominated by low frequency 1–5 Hz (similar to respiratory rhythm, see below), fluctuations in the membrane potential (Wilson and Groves, 1981; Petersen et al., 2003; Ferezou et al., 2007), or fluctuations in multiunit activity (Ros et al., 2009). Intracranial EEG from human beings show similar low-frequency oscillations (Sachdev et al., 2015).

Movement (e.g., whisking or locomotion) abolishes the low frequencies and puts cortical circuits into a depolarized, persistently activated high gamma state (Petersen et al., 2003; Poulet et al., 2012; Polack et al., 2013; Zagha et al., 2013; Urbain et al., 2015). In the whisker system, even after the sensory nerve is cut, whisking initiates state changes (Poulet and Petersen, 2008; Poulet et al., 2012).

Even though movement can change brain state, two additional consequences of movement should be noted. One is that movement activates a variety of cortical circuits, and as mentioned above, the cortical pyramidal neuron links many sensorimotor brain systems together (Figure 1C). Thus, the interaction between these cortical circuits alone, without overt movement or the participation of the thalamus, can change brain state (Zagha et al., 2013). Second, movement is related to the phase and frequency of respiration. The latter point is important to consider because movement—at least orofacial movements in rats and mice—are modulated by respiration. Respiration can serve as a “common clock” for aligning diverse processes and behaviors (Cao et al., 2012; Moore et al., 2013; Ranade et al., 2013; Kleinfeld et al., 2014; McElvain et al., 2018).

While each breath does not reach conscious perception, it can be brought under voluntary control (Davenport et al., 2007). In this instance, sensorimotor processing is contingent upon a shift from autonomous respiratory rhythm to active sniffing

bouts (from 1–5 Hz basal rate to 5–14 Hz in rodents). Active sensing by sniffing is coordinated with attentional processes and rhythmic movements (e.g., head direction or fine orofacial movements like whisking) that enable selective sampling routines (Schroeder et al., 2010; Moore et al., 2013; Jordan et al., 2018). This view seems paradigmatic in rodents as they are macrosomatic animals that sniff to sample their environment. Humans on the other hand favor vision and audition as a major source of information about the outside world and are less reliant on sniffing. However, studies in human beings have shown that cognitive processing of non-olfactory tasks is closely tied to the respiratory cycle (Nakamura et al., 2018; Perl et al., 2019), and that artificial puffs of air directed at the nares can change the subjective sensation of an altered state of consciousness (Piarulli et al., 2018).

Global brain oscillations, coupled to breathing, are coming into light as a general neural principle that may provide a configuration for temporally organizing brain activity across distant brain regions. Often these respiration-related brain rhythms are discussed in the context of the olfactory system—that they are driven by entrainment to nasal airflow. Though this is perfectly reasonable, as discussed below, there could also be other pathways that carry respiration-related information to link breathing behavior to brain-state.

## RESPIRATION AFFECTS RHYTHMIC BRAIN ACTIVITY

Lord Adrian was the first to report that “Normal breathing (without odor stimulation) produces a regular series of large potential waves in the pyriform area at each inspiration” (Adrian, 1942, p. 472). Years later, Fontanini et al. (2003) demonstrated that slow oscillations (<1.5 Hz) in the olfactory bulb and piriform cortex correlated with respiration in anesthetized rats (Fontanini and Bower, 2006). Subsequent studies in anesthetized and awake rodents have shown that respiratory rhythms modulate the oscillations (<5 Hz during anesthesia or quiet awake; >5 Hz during active sniffing) in brain regions downstream to the olfactory system (e.g., orbitofrontal cortex, the prelimbic cortex, and the hippocampus; Yanovsky et al., 2014; Lockmann et al., 2016; Biskamp et al., 2017; Liu et al., 2017; Kszeghy et al., 2018; Moberly et al., 2018; Rojas-Libano et al., 2018), and also structures that are not associated with the olfactory system (e.g., somatosensory, primary motor, and primary visual cortices; Ito et al., 2014; Rojas-Libano et al., 2018; Tort et al., 2018b). The respiration-related oscillations selectively modulate the 80–120 Hz gamma frequency band, the power of which attenuates as it propagates away from the olfactory bulb to cortical structures (Zhong et al., 2017). These studies have been extended to the human brain showing that respiratory rhythms are locked to nasal, but not oral inhalation, suggesting that respiration modulates brain activity even though humans rely less on olfactory information compared with rodents (Zelano et al., 2016; Herrero et al., 2018; Perl et al., 2019). The respiration-related modulation of low frequency and gamma-band oscillations are often overlooked because respiration frequency overlaps with both low frequencies

(1–4 Hz) and with theta frequency bands (Tort et al., 2018a). When respiration is monitored, it is possible to differentiate between the respiration and theta coupling of spiking activity of cortical and hippocampal neurons (Ito et al., 2014; Yanovsky et al., 2014; Nguyen Chi et al., 2016; Biskamp et al., 2017; Zhong et al., 2017). Together, these findings suggest that olfaction-related respiratory rhythms are present globally across the brain, that is, respiration modulates rhythmic activity in multiple cortical areas.

One pathway that can generate respiration-related rhythms involves sensory signals generated in response to external nasal airflow that propagates to the cortex *via* olfactory pathways (Figure 1D) and then reverberates within the recurrent connections of the cortical network itself (Bagur and Benchenane, 2018; Tort et al., 2018a). The nasal airflow drawn by breathing is necessary for these rhythms because these oscillations disappear once the nasal airway is bypassed using tracheotomy (Fontanini et al., 2003; Phillips et al., 2012; Ito et al., 2014; Yanovsky et al., 2014; Lockmann et al., 2016), and can be reinstated with rhythmic artificial nasal air puffs (Phillips et al., 2012; Lockmann et al., 2016). Furthermore, eliminating bulbar activity, surgically or chemically, strongly reduces respiration-related oscillations in brain regions downstream to the olfactory bulb while sparing other frequency bands (i.e., theta, gamma; Phillips et al., 2012; Biskamp et al., 2017; Liu et al., 2017; Tantirigama et al., 2017).

A different pathway that is often overlooked in the discussion of respiratory-related rhythms originates from the brainstem. The act of respiration itself generates sensory signals not only from the nasal epithelium as discussed earlier, but also from mechanoreceptors in the chest, skin, and muscles that are continually moved by respiration. Brain stem nuclei that manage breathing connect with the thalamus, where thalamic neurons fire in synchrony with respiration (Chen et al., 1992; Pattinson et al., 2009; Yang and Feldman, 2018). From the thalamus, these respiration-related signals echo to many brain regions that are involved in respiratory proprioception and the qualitative evaluation of respiration (e.g., diencephalon, limbic structures, and neocortex; Chen et al., 1992; Davenport and Vovk, 2009). Recently, a loss-of-function study showed that silencing nucleus reuniens of the thalamus reduces the 2–5 Hz coherence between the prefrontal cortex and hippocampus without significantly affecting coherence for theta oscillation (Roy et al., 2017). Additionally, recent causal evidence shows that a set of neurons in the Pre-Bötzinger complex can influence the brain-state. Manipulation of these neurons does not affect resting respiratory activity, however, it can dramatically change arousal or vigilance *via* direct connections to noradrenergic neurons in the locus coeruleus (Yackle et al., 2017). As mentioned earlier, orexin/hypocretin neurons also target the pre-Bötzinger complex (Young et al., 2005), and are likely to coordinate with the noradrenergic system in the locus coeruleus. Given that the brainstem circuits managing respiration are ultimately setting the rhythm of respiration (Ramirez and Baertsch, 2018), and thus the rhythm of chest movement and entry of air into the nares, these circuits are likely to play a role in modulating brain state in collaboration with neuromodulatory pathways.

## THE CORTICAL PYRAMIDAL NEURON—THE KEY TO UNDERSTANDING BRAIN STATE TRANSITIONS

Here, we have tackled the various cortical/subcortical interactions that combine to influence brain rhythms (Figure 1). The thalamocortical loops, neuromodulatory signals, activity generated by movement and respiration all interact with each other in a manner that is not yet completely understood. However, their effect to produce a change in brain state is ultimately manifested across the pyramidal neuron population. The resultant signal associated with the system-wide firing of various interacting neurons and brain areas results in specific changes in the flow of extracellular cortical currents, which make up the bulk of the signal finally referred to as “brain state” (or EEG). This raises three interesting points that revolve around cortical pyramidal neurons.

The first thing to consider is that pyramidal neurons are not “point neurons” but rather large cells that span several cortical layers (Larkum et al., 2018). This means that the influence of inputs arriving at different lamina will engage different compartments of pyramidal neurons that depend exquisitely on both the precise locations of the input on the dendritic tree and on the intrinsic properties of the neuron. These synaptic interactions can be quite baroque under certain spatiotemporal input distributions (Larkum, 2013; Stuart and Spruston, 2015). A well-studied example of this is the generation of NMDA spikes and calcium spikes in the distal dendrites, which can couple with input to the soma and trigger a burst output (Palmer et al., 2012, 2014; Stuart and Spruston, 2015). Importantly, this input/output transformation enabled by calcium spikes is strongly correlated to brain state (Phillips et al., 2016; Suzuki and Larkum, 2017; Seibt et al., 2017), and recent work shows that when the brain-state changes, the coupling between the soma and dendrite also changes (Suzuki and Larkum, 2020). This means that the influence of input from subcortical structures on the cortex must be gauged in the light of its specific effect on the active dendritic properties of pyramidal neurons, and how they can contribute to producing changes in the brain-state (Phillips et al., 2016; Aru et al., 2019).

Second point is that the output of pyramidal neurons and their subcortical targets interact in a complex way. In particular, the intrinsic properties of pyramidal neurons lead them to switch their firing mode (e.g., from sparse to regular firing, or bursting) as a function of the spatiotemporal pattern of inputs, which in turn is presumed to have a variable influence on their target neurons as a function of the short-term synaptic plasticity at their output synapses (Williams, 2005). To reiterate, during state transitions, it is these neurons that can modify and modulate the activity of many subcortical targets that are involved in brain-state, including the thalamus, brainstem/hypothalamic neuromodulatory systems, motor systems and respiratory systems.

Third, as discussed above, many long-range influences on brain-state are neuromodulatory in nature (e.g., cholinergic,

orexinergic, noradrenergic, etc.), and arrive into specific layers of the cortex (Jacob and Nienborg, 2018). This means that at any moment a stereotypical, spatiotemporal pattern of activity onto a pyramidal neuron can be modified under the influence of neuromodulation. We are still only at the beginning of understanding how neuromodulation can modify the somatodendritic properties of pyramidal neurons and influence the brain-state on a moment-to-moment basis (e.g., Williams and Fletcher, 2019; Suzuki and Larkum, 2020).

All in all, there is a level of complexity in understanding and explaining the precise rhythms encompassed by brain-states that are rarely addressed in descriptions of system-wide interactions usually involving simple “block diagrams.” We predict that the key to understanding brain state transitions will require experiments and models that include these intricacies about pyramidal neurons.

## CONCLUSION

Our understanding of how brain state changes, and the circuits involved in producing these changes is rapidly expanding. Many circuits related to changes in brain state—i.e., the pathways through the thalamus, neuromodulatory systems in the hypothalamus and brainstem, and the link between breathing and orofacial behaviors, among them—have been revealed.

## REFERENCES

- Adamantidis, A. R., Zhang, F., Aravanis, A. M., Deisseroth, K., and de Lecea, L. (2007). Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature* 450, 420–424. doi: 10.1038/nature06310
- Adrian, E. D. (1942). Olfactory reactions in the brain of the hedgehog. *J. Physiol.* 100, 459–473. doi: 10.1113/jphysiol.1942.sp003955
- Aguilar, J., Humanes-Valera, D., Alonso-Calviño, E., Yague, J. G., Moxon, K. A., Oliviero, A., et al. (2010). Spinal cord injury immediately changes the state of the brain. *J. Neurosci.* 30, 7528–7537. doi: 10.1523/JNEUROSCI.0379-10.2010
- Alexandre, C., Andermann, M. L., and Scammell, T. E. (2013). Control of arousal by the orexin neurons. *Curr. Opin. Neurobiol.* 23, 752–759. doi: 10.1016/j.conb.2013.04.008
- Anaclet, C., Pedersen, N. P., Ferrari, L. L., Venner, A., Bass, C. E., Arrigoni, E., et al. (2015). Basal forebrain control of wakefulness and cortical rhythms. *Nat. Commun.* 6:8744. doi: 10.1038/ncomms9744
- Aru, J., Suzuki, M., Rutiku, R., Larkum, M. E., and Bachmann, T. (2019). Coupling the state and contents of consciousness. *Front. Syst. Neurosci.* 13:43. doi: 10.3389/fnsys.2019.00043
- Bagur, S., and Benchenane, K. (2018). Taming the oscillatory zoo in the hippocampus and neo-cortex: a review of the commentary of Lockmann and Tort on Roy et al. *Brain Struct. Funct.* 223, 5–9. doi: 10.1007/s00429-017-1569-x
- Barthó, P., Slézia, A., Varga, V., Bokor, H., Pinault, D., Buzsáki, G., et al. (2007). Cortical control of zona incerta. *J. Neurosci.* 27, 1670–1681. doi: 10.1523/JNEUROSCI.3768-06.2007
- Bayer, L., Eggermann, E., Saint-Mleux, B., Machard, D., Jones, B. E., Mühlthaler, M., et al. (2002). Selective action of orexin (hypocretin) on nonspecific thalamocortical projection neurons. *J. Neurosci.* 22, 7835–7839. doi: 10.1523/JNEUROSCI.22-18-07835.2002
- Bayer, L., Serafin, M., Eggermann, E., Saint-Mleux, B., Machard, D., Jones, B. E., et al. (2004). Exclusive postsynaptic action of hypocretin-orexin on sublayer 6b cortical neurons. *J. Neurosci.* 24, 6760–6764. doi: 10.1523/JNEUROSCI.1783-04.2004
- Biskamp, J., Bartos, M., and Sauer, J.-F. (2017). Organization of prefrontal network activity by respiration-related oscillations. *Sci. Rep.* 7:45508. doi: 10.1038/srep45508
- Buzsáki, G., Anastassiou, C., and Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* 13, 407–420. doi: 10.1038/nrn3241
- Cao, Y., Roy, S., Sachdev, R. N. S., and Heck, D. (2012). Dynamic correlation between whisking and breathing rhythms in mice. *J. Neurosci.* 32, 1653–1659. doi: 10.1523/JNEUROSCI.4395-11.2012
- Carter, M. E., Yizhar, O., Chikahisa, S., Nguyen, H., Adamantidis, A., Nishino, S., et al. (2010). Tuning arousal with optogenetic modulation of locus coeruleus neurons. *Nat. Neurosci.* 13, 1526–1533. doi: 10.1038/nn.2682
- Chemelli, R. M., Willie, J. T., Sinton, C. M., Elmquist, J. K., Scammell, T., Lee, C., et al. (1999). Narcolepsy in orexin knockout mice: molecular genetics of sleep regulation. *Cell* 98, 437–451. doi: 10.1016/s0092-8674(00)81973-x
- Chen, Z., Eldridge, F. L., and Wagner, P. G. (1992). Respiratory-associated thalamic activity is related to level of respiratory drive. *Respir. Physiol.* 90, 99–113. doi: 10.1016/0034-5687(92)90137-I
- Cissé, Y., Toossi, H., Ishibashi, M., Mainville, L., Leonard, C. S., Adamantidis, A., et al. (2018). Discharge and role of acetylcholine pontomesencephalic neurons in cortical activity and sleep-wake states examined by optogenetics and juxtacellular recording in mice. *eNeuro* 5:ENEURO.0270-18.2018. doi: 10.1523/eneuro.0270-18.2018
- Contreras, D., Destexhe, A., Sejnowski, T., and Steriade, M. (1996). Control of spatiotemporal coherence of a thalamic oscillation by corticothalamic feedback. *Science* 274, 771–774. doi: 10.1126/science.274.5288.771
- Contreras, D., and Steriade, M. (1995). Cellular basis of EEG slow rhythms: a study of dynamic corticothalamic relationships. *J. Neurosci.* 15, 604–622. doi: 10.1523/JNEUROSCI.15-01-00604.1995
- Cowan, R. L., and Wilson, C. J. (1994). Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex. *J. Neurophysiol.* 71, 17–32. doi: 10.1152/jn.1994.71.1.17

Although each of these pathways is important for changing brain-state, perhaps each one alone is insufficient to trigger brain state changes, unless they collaborate to produce a system-wide change. The structural and functional interactions between them are still not clear, and our current challenge is to understand how these diverse set of cellular properties, circuits, and pathways interact to modify brain state at different timescales.

## AUTHOR CONTRIBUTIONS

MT, TZ, BJ, ML, and RS wrote the manuscript.

## FUNDING

The following funding sources have supported this project: (1) Deutsche Forschungsgemeinschaft (DFG), Grant Nos. LA 3442/6-1, LA 3442/3-1 and LA 3442/5-1 to ML; (2) DFG Project number 327654276—SFB 1315 to ML and BJ; (3) European Commission Horizon 2020 Research and Innovation Program and Euratom Research and Training Program 2014–2018 (under grant agreement No. 670118 to ML); (4) Human Brain Project (HBP FPA No. 650003) EU Commission Grant 720270, HBP SGA1 and SGA2 to ML; (5) Einstein Foundation Berlin EVF-2017-363 to ML; (6) European Research Council (ERC- 2016-StG-714560) to BJ.



- Davenport, P. W., and Vovk, A. (2009). Cortical and subcortical central neural pathways in respiratory sensations. *Respir. Physiol. Neurobiol.* 167, 72–86. doi: 10.1016/j.resp.2008.10.001
- Davenport, P. W., Chan, P.-Y. S., Zhang, W., and Chou, Y.-L. (2007). Detection threshold for inspiratory resistive loads and respiratory-related evoked potentials. *J. Appl. Physiol.* 102, 276–285. doi: 10.1152/japplphysiol.01436.2005
- de Lecea, L. (2015). Optogenetic control of hypocretin (orexin) neurons and arousal circuits. *Curr. Top. Behav. Neurosci.* 25, 367–378. doi: 10.1007/7854\_2014\_364
- de Lecea, L., Kilduff, T. S., Peyron, C., Gao, X.-B., Foye, P. E., Danielson, P. E., et al. (1998). The hypocretins: hypothalamus-specific peptides with neuroexcitatory activity. *Proc. Natl. Acad. Sci. U S A* 95, 322–327. doi: 10.1073/pnas.95.1.322
- DeFelipe, J., and Fariñas, I. (1992). The pyramidal neuron of the cerebral cortex: morphological and chemical characteristics of the synaptic inputs. *Prog. Neurobiol.* 39, 563–607. doi: 10.1016/0301-0082(92)90015-7
- Eban-Rothschild, A., Appelbaum, L., and de Lecea, L. (2018). Neuronal mechanisms for sleep/wake regulation and modulatory drive. *Neuropsychopharmacology* 43, 937–952. doi: 10.1038/npp.2017.294
- Ferezou, I., Haiss, F., Gentet, L. J., Aronoff, R., Weber, B., and Petersen, C. C. H. (2007). Spatiotemporal dynamics of cortical sensorimotor integration in behaving mice. *Neuron* 56, 907–923. doi: 10.1016/j.neuron.2007.10.007
- Fontanini, A., and Bower, J. M. (2006). Slow-waves in the olfactory system: an olfactory perspective on cortical rhythms. *Trends Neurosci.* 29, 429–437. doi: 10.1016/j.tins.2006.06.013
- Fontanini, A., Spano, P., and Bower, J. M. (2003). Ketamine-xylazine-induced slow (<1.5 Hz) oscillations in the rat piriform (olfactory) cortex are functionally correlated with respiration. *J. Neurosci.* 23, 7993–8001. doi: 10.1523/JNEUROSCI.23-22-07993.2003
- Gloor, P., Ball, G., and Schaul, N. (1977). Brain lesions that produce delta waves in the EEG. *Neurology* 27, 326–326. doi: 10.1212/wnl.27.4.326
- Harris, K. D., and Thiele, A. (2011). Cortical state and attention. *Nat. Rev. Neurosci.* 12, 509–523. doi: 10.1038/nrn3084
- Hasenstaub, A., Sachdev, R. N., and McCormick, D. A. (2007). State changes rapidly modulate cortical neuronal responsiveness. *J. Neurosci.* 27, 9607–9622. doi: 10.1523/JNEUROSCI.2184-07.2007
- Herrero, J. L., Khuvis, S., Yeagle, E., Cerf, M., and Mehta, A. D. (2018). Breathing above the brain stem: volitional control and attentional modulation in humans. *J. Neurophysiol.* 119, 145–159. doi: 10.1152/jn.00551.2017
- Hoerder-Suabedissen, A., Hayashi, S., Upton, L., Nolan, Z., Casas-Torremocha, D., Grant, E., et al. (2018). Subset of cortical layer 6b neurons selectively innervates higher order thalamic nuclei in mice. *Cereb. Cortex* 28, 1882–1897. doi: 10.1093/cercor/bhy036
- Honjoh, S., Sasai, S., Schierack, S. S., Nagai, H., Tonomi, G., and Cirelli, C. (2018). Regulation of cortical activity and arousal by the matrix cells of the ventromedial thalamic nucleus. *Nat. Commun.* 9:2100. doi: 10.1038/s41467-018-04497-x
- Ireland, J. D. (2007). *Udana and the Itivuttaka: Two Classics from the Pali Canon. Vol. 214.* (Kandy, Sri Lanka: Buddhist Publication Society).
- Ito, J., Roy, S., Liu, Y., Cao, Y., Fletcher, M., Lu, L., et al. (2014). Whisker barrel cortex delta oscillations and gamma power in the awake mouse are linked to respiration. *Nat. Commun.* 5:3572. doi: 10.1038/ncomms4572
- Jacob, S. N., and Nienborg, H. (2018). Monoaminergic neuromodulation of sensory processing. *Front. Neural Circuits* 12:51. doi: 10.3389/fncir.2018.00051
- Jones, B. E. (2008). Modulation of cortical activation and behavioral arousal by cholinergic and orexinergic systems. *Ann. N Y Acad. Sci.* 1129, 26–34. doi: 10.1196/annals.1417.026
- Jones, B. E. (2020). Arousal and sleep circuits. *Neuropsychopharmacology* 45, 6–20. doi: 10.1038/s41386-019-0444-2
- Jordan, R., Fukunaga, I., Kollo, M., and Schaefer, A. T. (2018). Active sampling state dynamically enhances olfactory bulb odor representation. *Neuron* 98, 1214.e5–1228.e5. doi: 10.1016/j.neuron.2018.05.016
- Kasanetz, F., Riquelme, L. A., and Murer, M. G. (2002). Disruption of the two-state membrane potential of striatal neurones during cortical desynchronisation in anaesthetised rats. *J. Physiol.* 543, 577–589. doi: 10.1113/jphysiol.2002.002476
- Kim, T., Thankachan, S., McKenna, J. T., McNally, J. M., Yang, C., Choi, J. H., et al. (2015). Cortically projecting basal forebrain parvalbumin neurons regulate cortical gamma band oscillations. *Proc. Natl. Acad. Sci. U S A* 112, 3535–3540. doi: 10.1073/pnas.1413625112
- Kleinfeld, D., Deschênes, M., Wang, F., and Moore, J. D. (2014). More than a rhythm of life: breathing as a binder of orofacial sensation. *Nat. Neurosci.* 17, 647–651. doi: 10.1038/nn.3693
- Koch, C. (2004). *The Quest for Consciousness a Neurobiological Approach.* Englewood, CO: Roberts & Company Publishers.
- Kszeghy, Á., Lasztóczy, B., Forro, T., and Klausberger, T. (2018). Spike-timing of orbitofrontal neurons is synchronized with breathing. *Front. Cell. Neurosci.* 12:105. doi: 10.3389/fncel.2018.00105
- Larkum, M. E. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* 36, 141–151. doi: 10.1016/j.tins.2012.11.006
- Larkum, M. E., Petro, L. S., Sachdev, R. N., and Muckli, L. (2018). A perspective on cortical layering and layer-spanning neuronal elements. *Front. Neuroanat.* 12:56. doi: 10.3389/fnana.2018.00056
- Lindsley, D. B., Bowden, J., and Magoun, H. (1949). Effect upon the EEG of acute injury to the brain stem activating system. *Electroencephalogr. Clin. Neurophysiol.* 1, 475–486. doi: 10.1016/0013-4694(49)90221-7
- Liu, Y., McAfee, S. S., and Heck, D. H. (2017). Hippocampal sharp-wave ripples in awake mice are entrained by respiration. *Sci. Rep.* 7:8950. doi: 10.1038/s41598-017-09511-8
- Lockmann, A. L. V., Laplagne, D. A., Leão, R. N., and Tort, A. B. L. (2016). A respiration-coupled rhythm in the rat hippocampus independent of theta and slow oscillations. *J. Neurosci.* 36, 5338–5352. doi: 10.1523/JNEUROSCI.3452-15.2016
- Magill, P. J., Bolam, J. P., and Bevan, M. D. (2000). Relationship of activity in the subthalamic nucleus-globus pallidus network to cortical electroencephalogram. *J. Neurosci.* 20, 820–833. doi: 10.1523/JNEUROSCI.20-02-00820.2000
- Magill, P. J., Sharott, A., Bevan, M. D., Brown, P., and Bolam, J. P. (2004). Synchronous unit activity and local field potentials evoked in the subthalamic nucleus by cortical stimulation. *J. Neurophysiol.* 92, 700–714. doi: 10.1152/jn.00134.2004
- McElvain, L. E., Friedman, B., Karten, H. J., Svoboda, K., Wang, F., Deschênes, M., et al. (2018). Circuits in the rodent brainstem that control whisking in concert with other orofacial motor actions. *Neuroscience* 368, 152–170. doi: 10.1016/j.neuroscience.2017.08.034
- Moberly, A. H., Schreck, M., Bhattarai, J. P., Zweifel, L. S., Luo, W., and Ma, M. (2018). Olfactory inputs modulate respiration-related rhythmic activity in the prefrontal cortex and freezing behavior. *Nat. Commun.* 9:1528. doi: 10.1038/s41467-018-03988-1
- Molyneux, B. J., Arlotta, P., Menezes, J. R. L., and Macklis, J. D. (2007). Neuronal subtype specification in the cerebral cortex. *Nat. Rev. Neurosci.* 8, 427–437. doi: 10.1038/nrn2151
- Moore, J. D., Deschênes, M., Furuta, T., Huber, D., Smear, M. C., Demers, M., et al. (2013). Hierarchy of orofacial rhythms revealed through whisking and breathing. *Nature* 497, 205–210. doi: 10.1038/nature12076
- Moruzzi, G., and Magoun, H. W. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalogr. Clin. Neurophysiol.* 1, 455–473. doi: 10.1016/0013-4694(49)90219-9
- Motelow, J. E., Li, W., Zhan, Q., Mishra, A. M., Sachdev, R. N. S., Liu, G., et al. (2015). Decreased subcortical cholinergic arousal in focal seizures. *Neuron* 85, 561–572. doi: 10.1016/j.neuron.2014.12.058
- Nakamura, N. H., Fukunaga, M., and Oku, Y. (2018). Respiratory modulation of cognitive performance during the retrieval process. *PLoS One* 13:e0204021. doi: 10.1371/journal.pone.0204021
- Nguyen Chi, V., Müller, C., Wolfenstetter, T., Yanovsky, Y., Draguhn, A., Tort, A. B. L., et al. (2016). Hippocampal respiration-driven rhythm distinct from theta oscillations in awake mice. *J. Neurosci.* 36, 162–177. doi: 10.1523/JNEUROSCI.2848-15.2016
- Ni, K.-M., Hou, X.-J., Yang, C.-H., Dong, P., Li, Y., Zhang, Y., et al. (2016). Selectively driving cholinergic fibers optically in the thalamic reticular nucleus promotes sleep. *Elife* 5:e10382. doi: 10.7554/eLife.10382
- Palmer, L. M., Schulz, J. M., Murphy, S. C., Ledergerber, D., Murayama, M., and Larkum, M. E. (2012). The cellular basis of GABAB-mediated interhemispheric inhibition. *Science* 335, 989–993. doi: 10.1126/science.1217276



- Palmer, L. M., Shai, A. S., Reeve, J. E., Anderson, H. L., Paulsen, O., and Larkum, M. E. (2014). NMDA spikes enhance action potential generation during sensory input. *Nat. Neurosci.* 17, 383–390. doi: 10.1038/nn.3646
- Pattinson, K., Mitsis, G., Harvey, A., Jbabdi, S., Dirckx, S., Mayhew, S., et al. (2009). Determination of the human brainstem respiratory control network and its cortical connections *in vivo* using functional and structural imaging. *NeuroImage* 44, 295–305. doi: 10.1016/j.neuroimage.2008.09.007
- Perl, O., Ravia, A., Robinson, M., Eisen, A., Soroka, T., Mor, N., et al. (2019). Human non-olfactory cognition phase-locked with inhalation. *Nat. Hum. Behav.* 3, 501–512. doi: 10.1038/s41562-019-0556-z
- Petersen, C. C., Grinvald, A., and Sakmann, B. (2003). Spatiotemporal dynamics of sensory responses in layer 2/3 of rat barrel cortex measured *in vivo* by voltage-sensitive dye imaging combined with whole-cell voltage recordings and neuron reconstructions. *J. Neurosci.* 23, 1298–1309. doi: 10.1523/JNEUROSCI.23-04-01298.2003
- Petzold, A., Valencia, M., Pál, B., and Mena-Segovia, J. (2015). Decoding brain state transitions in the pedunculopontine nucleus: cooperative phasic and tonic mechanisms. *Front. Neural Circuits* 9:68. doi: 10.3389/fncir.2015.00068
- Peyron, C., Faraco, J., Rogers, W., Ripley, B., Overeem, S., Charnay, Y., et al. (2000). A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. *Nat. Med.* 6, 991–997. doi: 10.1038/79690
- Peyron, C., Tighe, D. K., van den Pol, A. N., de Lecea, L., Heller, H. C., Sutcliffe, J. G., et al. (1998). Neurons containing hypocretin (orexin) project to multiple neuronal systems. *J. Neurosci.* 18, 9996–10015. doi: 10.1523/JNEUROSCI.18-23-09996.1998
- Phillips, W. A., Larkum, M. E., Harley, C. W., and Silverstein, S. M. (2016). The effects of arousal on apical amplification and conscious state. *Neurosci. Conscious.* 2016:niw015. doi: 10.1093/nc/niw015
- Phillips, M. E., Sachdev, R. N. S., Willhite, D. C., and Shepherd, G. M. (2012). Respiration drives network activity and modulates synaptic and circuit processing of lateral inhibition in the olfactory bulb. *J. Neurosci.* 32, 85–98. doi: 10.1523/JNEUROSCI.4278-11.2012
- Piarulli, A., Zaccaro, A., Laurino, M., Menicucci, D., De Vito, A., Bruschini, L., et al. (2018). Ultra-slow mechanical stimulation of olfactory epithelium modulates consciousness by slowing cerebral rhythms in humans. *Sci. Rep.* 8:6581. doi: 10.1038/s41598-018-24924-9
- Polack, P. O., Friedman, J., and Golshani, P. (2013). Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *Nat. Neurosci.* 16, 1331–1333.
- Poulet, J. F. A., and Crochet, S. (2019). The cortical states of wakefulness. *Front. Syst. Neurosci.* 12:64. doi: 10.3389/fnsys.2018.00064
- Poulet, J. F. A., Fernandez, L. M. J., Crochet, S., and Petersen, C. C. H. (2012). Thalamic control of cortical states. *Nat. Neurosci.* 15, 370–372. doi: 10.1038/nn.3035
- Poulet, J. F., and Petersen, C. C. (2008). Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* 454, 881–885.
- Ramirez, J.-M., and Baertsch, N. A. (2018). The dynamic basis of respiratory rhythm generation: one breath at a time. *Annu. Rev. Neurosci.* 41, 475–499. doi: 10.1146/annurev-neuro-080317-061756
- Ranade, S., Hangya, B., and Kepecs, A. (2013). Multiple modes of phase locking between sniffing and whisking during active exploration. *J. Neurosci.* 33, 8250–8256. doi: 10.1523/JNEUROSCI.3874-12.2013
- Redinbaugh, M. J., Phillips, J. M., Kambi, N. A., Mohanta, S., Andryk, S., Dooley, G. L., et al. (2020). Thalamus modulates consciousness via layer-specific control of cortex. *Neuron* 106, 66.e12–75.e12. doi: 10.1016/j.neuron.2020.01.005
- Ren, S., Wang, Y., Yue, F., Cheng, X., Dang, R., Qiao, Q., et al. (2018). The paraventricular thalamus is a critical thalamic area for wakefulness. *Science* 362, 429–434. doi: 10.1126/science.aat2512
- Rojas-Libano, D., Wimmer Del Solar, J., Aguilar-Rivera, M., Montefusco-Siegmund, R., and Maldonado, P. E. (2018). Local cortical activity of distant brain areas can phase-lock to the olfactory bulb's respiratory rhythm in the freely behaving rat. *J. Neurophysiol.* 120, 960–972. doi: 10.1152/jn.00088.2018
- Roš, H., Magill, P. J., Moss, J., Bolam, J. P., and Mena-Segovia, J. (2010). Distinct types of non-cholinergic pedunculopontine neurons are differentially modulated during global brain states. *Neuroscience* 170, 78–91. doi: 10.1016/j.neuroscience.2010.06.068
- Ros, H., Sachdev, R. N. S., Yu, Y., Sestan, N., and McCormick, D. A. (2009). Neocortical networks entrain neuronal circuits in cerebellar cortex. *J. Neurosci.* 29, 10309–10320. doi: 10.1523/JNEUROSCI.2327-09.2009
- Rowland, N. C., Goldberg, J. A., and Jaeger, D. (2010). Cortico-cerebellar coherence and causal connectivity during slow-wave activity. *Neuroscience* 166, 698–711. doi: 10.1016/j.neuroscience.2009.12.048
- Roy, A., Svensson, F. P., Mazeh, A., and Kocsis, B. (2017). Prefrontal-hippocampal coupling by theta rhythm and by 2–5 Hz oscillation in the delta band: the role of the nucleus reuniens of the thalamus. *Brain Struct. Funct.* 222, 2819–2830. doi: 10.1007/s00429-017-1374-6
- Saalmann, Y. B. (2014). Intralaminar and medial thalamic influence on cortical synchrony, information transmission and cognition. *Front. Syst. Neurosci.* 8:83. doi: 10.3389/fnsys.2014.00083
- Sachdev, R. N., Ebner, F. F., and Wilson, C. J. (2004). Effect of subthreshold up and down states on the whisker-evoked response in somatosensory cortex. *J. Neurophysiol.* 92, 3511–3521. doi: 10.1152/jn.00347.2004
- Sachdev, R. N. S., Gaspard, N., Gerrard, J. L., Hirsch, L. J., Spencer, D. D., and Zaveri, H. P. (2015). Delta rhythm in wakefulness: evidence from intracranial recordings in human beings. *J. Neurophysiol.* 114, 1248–1254. doi: 10.1152/jn.00249.2015
- Sakurai, T., Amemiya, A., Ishii, M., Matsuzaki, I., Chemelli, R. M., Tanaka, H., et al. (1998). Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* 92, 573–585. doi: 10.1016/s0092-8674(00)80949-6
- Sara, S. J., and Hervé-Minvielle, A. (1995). Inhibitory influence of frontal cortex on locus coeruleus neurons. *Proc. Natl. Acad. Sci. U S A* 92, 6032–6036. doi: 10.1073/pnas.92.13.6032
- Sasaki, K., Suzuki, M., Mieda, M., Tsujino, N., Roth, B., and Sakurai, T. (2011). Pharmacogenetic modulation of orexin neurons alters sleep/wakefulness states in mice. *PLoS One* 6:e20360. doi: 10.1371/journal.pone.0020360
- Scammell, T. E., Arrigoni, E., and Lipton, J. O. (2017). Neural circuitry of wakefulness and sleep. *Neuron* 93, 747–765. doi: 10.1016/j.neuron.2017.01.014
- Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Curr. Opin. Neurobiol.* 20, 172–176. doi: 10.1016/j.conb.2010.02.010
- Seibt, J., Richard, C. J., Sigl-Glöckner, J., Takahashi, N., Kaplan, D. I., Doron, G., et al. (2017). Cortical dendritic activity correlates with spindle-rich oscillations during sleep in rodents. *Nat. Comm.* 8, 1–13.
- Sherman, S. M., and Guillery, R. W. (2011). Distinct functions for direct and transthalamic corticocortical connections. *J. Neurophysiol.* 106, 1068–1077. doi: 10.1152/jn.00429.2011
- Steriade, M., and McCarley, R. W. (2005). *Brainstem Control of Wakefulness and Sleep*. New York, NY: Plenum Press.
- Steriade, M. (2006). Grouping of brain rhythms in the corticothalamic systems. *Neuroscience* 137, 1087–1106. doi: 10.1016/j.neuroscience.2005.10.029
- Steriade, M., Contreras, D., Curro Dossi, R., and Nunez, A. (1993a). The slow (< 1 Hz) oscillation in reticular thalamic and thalamocortical neurons: scenario of sleep rhythm generation in interacting thalamic and neocortical networks. *J. Neurosci.* 13, 3284–3299.
- Steriade, M., Nuñez, A., and Amzica, F. (1993b). A novel slow (< 1 Hz) oscillation of neocortical neurons *in vivo*: depolarizing and hyperpolarizing components. *J. Neurosci.* 13, 3252–3265.
- Steriade, M., Nuñez, A., and Amzica, F. (1993c). Intracellular analysis of relations between the slow (< 1 Hz) neocortical oscillation and other sleep rhythms of the electroencephalogram. *J. Neurosci.* 13, 3266–3283. doi: 10.1523/jneurosci.13-08-03266.1993
- Stuart, G. J., and Spruston, N. (2015). Dendritic integration: 60 years of progress. *Nat. Neurosci.* 18, 1713–1721. doi: 10.1038/nn.4157
- Suzuki, M., and Larkum, M. E. (2017). Dendritic calcium spikes are clearly detectable at the cortical surface. *Nat. Commun.* 8:276. doi: 10.1038/s41467-017-00282-4
- Suzuki, M., and Larkum, M. E. (2020). General anesthesia decouples cortical pyramidal neurons. *Cell* 180, 666.e13–676.e13. doi: 10.1016/j.cell.2020.01.024
- Tantirigama, M. L. S., Huang, H. H.-Y., and Bekkers, J. M. (2017). Spontaneous activity in the piriform cortex extends the dynamic range of cortical odor coding. *Proc. Natl. Acad. Sci. U S A* 114, 2407–2412. doi: 10.1073/pnas.1620939114

- Thannickal, T. C., Moore, R. Y., Nienhuis, R., Ramanathan, L., Gulyani, S., Aldrich, M., et al. (2000). Reduced number of hypocretin neurons in human narcolepsy. *Neuron* 27, 469–474. doi: 10.1016/s0896-6273(00)00058-1
- Timofeev, I., Grenier, F., Bazhenov, M., Sejnowski, T. J., and Steriade, M. (2000). Origin of slow cortical oscillations in deafferented cortical slabs. *Cereb. Cortex* 10, 1185–1199. doi: 10.1093/cercor/10.12.1185
- Tort, A. B. L., Brankač, J., and Draguhn, A. (2018a). Respiration-entrained brain rhythms are global but often overlooked. *Trends Neurosci.* 41, 186–197. doi: 10.1016/j.tins.2018.01.007
- Tort, A. B. L., Ponsel, S., Jessberger, J., Yanovsky, Y., Brankač, J., and Draguhn, A. (2018b). Parallel detection of theta and respiration-coupled oscillations throughout the mouse brain. *Sci. Rep.* 8:6432. doi: 10.1038/s41598-018-24629-z
- Tseng, K. Y., Kasanetz, F., Kargieman, L., Riquelme, L. A., and Murer, M. G. (2001). Cortical slow oscillatory activity is reflected in the membrane potential and spike trains of striatal neurons in rats with chronic nigrostriatal lesions. *J. Neurosci.* 21, 6430–6439. doi: 10.1523/jneurosci.21-16-06430.2001
- Tsunematsu, T., Tabuchi, S., Tanaka, K. F., Boyden, E. S., Tominaga, M., and Yamanaka, A. (2013). Long-lasting silencing of orexin/hypocretin neurons using archaerhodopsin induces slow-wave sleep in mice. *Behav. Brain Res.* 255, 64–74. doi: 10.1016/j.bbr.2013.05.021
- Urbain, N., Fourcaud-Trocmé, N., Laheux, S., Salin, P. A., and Gentet, L. J. (2019). Brain-state-dependent modulation of neuronal firing and membrane potential dynamics in the somatosensory thalamus during natural sleep. *Cell Rep.* 26, 1443–1457. doi: 10.1016/j.celrep.2019.01.038
- Urbain, N., Salin, P. A., Libourel, P. A., Comte, J. C., Gentet, L. J., and Petersen, C. C. H. (2015). Whisking-related changes in neuronal firing and membrane potential dynamics in the somatosensory thalamus of awake mice. *Cell Rep.* 13, 647–656. doi: 10.1016/j.celrep.2015.09.029
- Vazey, E. M., Moorman, D. E., and Aston-Jones, G. (2018). Phasic locus coeruleus activity regulates cortical encoding of salience information. *Proc. Natl. Acad. Sci. U S A* 115, E9439–E9448. doi: 10.1073/pnas.1803716115
- Viswanathan, S., Sheikh, A., Looger, L. L., and Kanold, P. O. (2017). Molecularly defined subplate neurons project both to thalamocortical recipient layers and thalamus. *Cereb. Cortex* 27, 4759–4768. doi: 10.1093/cercor/bhw271
- Wenger Combremont, A.-L., Bayer, L., Dupré, A., Mühlethaler, M., and Serafin, M. (2016). Slow bursting neurons of mouse cortical layer 6b are depolarized by hypocretin/orexin and major transmitters of arousal. *Front. Neurol.* 7:88. doi: 10.3389/fneur.2016.00088
- Williams, S. R. (2005). Encoding and decoding of dendritic excitation during active states in pyramidal neurons. *J. Neurosci.* 25, 5894–5902. doi: 10.1523/jneurosci.0502-05.2005
- Williams, S. R., and Fletcher, L. N. (2019). A dendritic substrate for the cholinergic control of neocortical output neurons. *Neuron* 101, 486.e4–499.e4. doi: 10.1016/j.neuron.2018.11.035
- Wilson, C. J., and Groves, P. M. (1981). Spontaneous firing patterns of identified spiny neurons in the rat neostriatum. *Brain Res.* 220, 67–80. doi: 10.1016/0006-8993(81)90211-0
- Wilson, C., and Kawaguchi, Y. (1996). The origins of two-state spontaneous membrane potential fluctuations of neostriatal spiny neurons. *J. Neurosci.* 16, 2397–2410. doi: 10.1523/jneurosci.16-07-02397.1996
- Xu, M., Chung, S., Zhang, S., Zhong, P., Ma, C., Chang, W.-C., et al. (2015). Basal forebrain circuit for sleep-wake control. *Nat. Neurosci.* 18, 1641–1647. doi: 10.1038/nn.4143
- Yackle, K., Schwarz, L. A., Kam, K., Sorokin, J. M., Huguenard, J. R., Feldman, J. L., et al. (2017). Breathing control center neurons that promote arousal in mice. *Science* 355, 1411–1415. doi: 10.1126/science.aai7984
- Yang, C. F., and Feldman, J. L. (2018). Efferent projections of excitatory and inhibitory preBötzinger Complex neurons. *J. Comp. Neurol.* 526, 1389–1402. doi: 10.1002/cne.24415
- Yanovsky, Y., Ciatipis, M., Draguhn, A., Tort, A. B. L., and Brankač, J. (2014). Slow oscillations in the mouse hippocampus entrained by nasal respiration. *J. Neurosci.* 34, 5949–5964. doi: 10.1523/jneurosci.5287-13.2014
- Young, J. K., Wu, M., Manaye, K. F., Kc, P., Allard, J. S., Mack, S. O., et al. (2005). Orexin stimulates breathing via medullary and spinal pathways. *J. Appl. Physiol.* 98, 1387–1395. doi: 10.1152/jappphysiol.00914.2004
- Zagha, E., Casale, A. E., Sachdev, R. N. S., McGinley, M. J., and McCormick, D. A. (2013). Motor cortex feedback influences sensory processing by modulating network state. *Neuron* 79, 567–578. doi: 10.1016/j.neuron.2013.06.008
- Zelano, C., Jiang, H., Zhou, G., Arora, N., Schuele, S., Rosenow, J., et al. (2016). Nasal respiration entrains human limbic oscillations and modulates cognitive function. *J. Neurosci.* 36, 12448–12467. doi: 10.1523/JNEUROSCI.2586-16.2016
- Zhong, W., Ciatipis, M., Wolfenstetter, T., Jessberger, J., Müller, C., Ponsel, S., et al. (2017). Selective entrainment of gamma subbands by different slow network oscillations. *Proc. Natl. Acad. Sci. U S A* 114, 4519–4524. doi: 10.1073/pnas.1617249114
- Zolnik, T. A., Ledderose, J., Toumazou, M., Trimbuch, T., Oram, T., Rosenmund, C., et al. (2020). Layer 6b is driven by intracortical long-range projection neurons. *Cell Rep.* 30, 3492.e5–3505.e5. doi: 10.1016/j.celrep.2020.02.044

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Tantirigama, Zolnik, Judkewitz, Larkum and Sachdev. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Up and Down States and Memory Consolidation Across Somatosensory, Entorhinal, and Hippocampal Cortices

John J. Tukker<sup>1,2,3\*</sup>†, Prateep Beed<sup>1,2,4†</sup>, Dietmar Schmitz<sup>1,2,3,4,5,6</sup>, Matthew E. Larkum<sup>5,6,7</sup> and Robert N. S. Sachdev<sup>7\*</sup>

<sup>1</sup> Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany, <sup>2</sup> Neuroscience Research Center, Berlin, Germany, <sup>3</sup> German Center for Neurodegenerative Diseases (DZNE), Berlin, Germany, <sup>4</sup> Berlin Institute of Health, Berlin, Germany, <sup>5</sup> Cluster of Excellence NeuroCure, Berlin, Germany, <sup>6</sup> Einstein Center for Neurosciences Berlin, Berlin, Germany, <sup>7</sup> Institut für Biologie, Humboldt Universität, Berlin, Germany

## OPEN ACCESS

### Edited by:

Alain Destexhe,  
FRE3693 Unité de Neuroscience,  
Information et Complexité (UNIC),  
France

### Reviewed by:

M. Gustavo Murer,  
University of Buenos Aires, Argentina  
Ehren Lee Newman,  
Indiana University Bloomington,  
United States

### \*Correspondence:

John J. Tukker  
john.tukker@charite.de  
Robert N. S. Sachdev  
bs387ster@gmail.com

† These authors have contributed  
equally to this work

**Received:** 31 January 2020

**Accepted:** 03 April 2020

**Published:** 08 May 2020

### Citation:

Tukker JJ, Beed P, Schmitz D,  
Larkum ME and Sachdev RNS (2020)  
Up and Down States and Memory  
Consolidation Across Somatosensory,  
Entorhinal, and Hippocampal  
Cortices.  
Front. Syst. Neurosci. 14:22.  
doi: 10.3389/fnsys.2020.00022

In the course of a day, brain states fluctuate, from conscious awake information-acquiring states to sleep states, during which previously acquired information is further processed and stored as memories. One hypothesis is that memories are consolidated and stored during “offline” states such as sleep, a process thought to involve transfer of information from the hippocampus to other cortical areas. Up and Down states (UDS), patterns of activity that occur under anesthesia and sleep states, are likely to play a role in this process, although the nature of this role remains unclear. Here we review what is currently known about these mechanisms in three anatomically distinct but interconnected cortical areas: somatosensory cortex, entorhinal cortex, and the hippocampus. In doing so, we consider the role of this activity in the coordination of “replay” during sleep states, particularly during hippocampal sharp-wave ripples. We conclude that understanding the generation and propagation of UDS may provide key insights into the cortico-hippocampal dialogue linking archi- and neocortical areas during memory formation.

**Keywords:** Up and Down states, memory consolidation, sleep, somatosensory cortex, entorhinal cortex, hippocampus, inhibition, neuromodulation

## INTRODUCTION

The nature of consciousness and its neural correlates are still not completely established. Does consciousness require the activity of a subset of neurons in the brainstem or thalamus, or is it a property reflected in the global activated states that can be observed in electroencephalogram (EEG) recordings during wakefulness? Even though the pattern of activity in EEG can be related to distinct brain states, these recordings have not elucidated the nature of consciousness. Nor have they established how or why brain states change from conscious states to sleep states. One element of awake – information acquiring – conscious brain states is that cortical neurons are depolarized, close to action potential threshold, i.e., in an activated state (Berger, 1929; Timofeev et al., 2001). While low frequency fluctuations in the EEG and membrane potential – i.e., delta rhythm (1–4 Hz) – are dominant during sleep, they can also occur in awake states (Wilson and Groves, 1981;

**BOX 1 | Models of Cortical Up and Down States**

A number of computational models have sought to elucidate mechanisms underlying Up and Down states. The older models highlighted the intrinsic single cell properties of cortical neurons and the interaction between these properties and circuit activity that could generate slow oscillations in cortex. The newer models go further and try to capture differences in the dynamics of Up and Down states during sleep and anesthesia.

In 2002, Bazhenov and colleagues suggested that a summation of PSPs, depolarization, and a persistent  $\text{Na}^+$  current were important in initiating transitions from Down to Up states; a  $\text{Ca}^{2+}$ -dependent  $\text{K}^+$  current as well as synaptic depression and inhibition were important in terminating Up states (Bazhenov et al., 2002).

A model developed by Compte and colleagues suggested that the Up state was maintained by strong recurrent excitation balanced by inhibition and the transition to a Down state was due to a slow adaptation ( $\text{Na}^+$  dependent  $\text{K}^+$ ) current (Compte et al., 2003; see also Mattia and Sanchez-Vives, 2012).

A recent study by Jercog and colleagues modeled Up and Down states of urethane-anesthetized rats (Jercog et al., 2017). This model used adapting, balanced recurrent excitatory and inhibitory activity to reproduce a bistable membrane potential regime. The model suggests that whereas the transition from the Up to the Down state can be explained by adaptation in combination with the stochastic spiking activity during the Up state, the Down to Up transitions require external inputs.

In Levenstein and colleagues' study, the dynamics of naturally occurring Up and Down-states in hippocampus and neocortex, as measured in the firing rates of neurons recorded from sleeping rats, could both be captured by a relatively simple model including only recurrent excitation, a slowly adapting negative feedback, and a noisy external drive (Levenstein et al., 2019). The model could best reproduce the recorded dynamics of neocortical circuits (long Up states of variable duration and shorter less variable Down states) in a regime where the Up state was stable, meaning external inputs or noise are needed to trigger a transition to a Down state. In contrast, the Down state was not stable, meaning the system spontaneously reverts back to the Up state. The recorded hippocampal dynamics were opposite (long Down states of variable duration and shorter less variable Up states, known as sharp waves), and could best be reproduced when the model was in a regime where a stable Down state needs external inputs or noise to trigger a transition to the Up state.

Finally, Nghiem and colleagues show that the dynamics of Up and Down states differ in anesthesia and slow wave sleep and that varying the adaptation strength using a cholinergic input could explain the different dynamics (Nghiem et al., 2018). Their model predicts that sensitivity of cortical dynamics to external inputs is stronger in sleep compared to anesthetics, possibly explaining why sleep is associated with memory consolidation, whereas anesthesia is generally associated with amnesia.

Petersen et al., 2003; Crochet and Petersen, 2006; Ferezou et al., 2007; Sachdev et al., 2015). Whether low frequency fluctuations are hallmarks of local sleep of a cortical area (Vyazovskiy et al., 2011) or reflect a state of inattention is not clear, but work in epilepsy patients suggests that low frequency epochs occur when sensory awareness is diminished (Blumenfeld, 2002; Motelow et al., 2015).

The Up state is a simpler, reduced model of activated, awake states; it has been called a “fragment of wakefulness” (Destexhe et al., 2007). Up states are activated states where the EEG is desynchronized, dominated by gamma frequencies, during which individual cortical neurons are depolarized close to threshold (Metherate and Ashe, 1993; Steriade et al., 1993c; Cowan and Wilson, 1994). When relating consciousness or wakefulness to Up states, there is a caveat to keep in mind: during slow wave sleep, or under anesthesia, even during these epochs of activity, the brain is not in an information acquiring state; in fact the Up

states can be epochs of low responsiveness (Petersen et al., 2003; Sachdev et al., 2004; Hasenstaub et al., 2007). One other issue to bear in mind is that a majority of studies of Up and Down states (UDS) have been with anesthetics, and the slow waves that occur under anesthesia and those that occur under natural sleep have different dynamics and structure (Jercog et al., 2017; Nghiem et al., 2018; Levenstein et al., 2019; **Box 1**). Nevertheless, here we compare how these simple epochs of persistent activity occur in somatosensory and entorhinal cortex and how they are linked to each other and to hippocampal activity. In doing so, we consider the role of laminar structure and connectivity, intrinsic properties, neuromodulators, and inhibition in supporting UDS.

Cortical states of wakefulness, sleep, attention, and UDS have been written about extensively over the decades, with books and many reviews about the organization, mechanisms and effects of slow wave sleep (Steriade et al., 1993b; Steriade, 2001; Steriade and McCarley, 2005; Destexhe et al., 2007; Harris and Thiele, 2011; Zagha and McCormick, 2014; Poulet and Crochet, 2018). These states have been reviewed in many different contexts: (1) circuit mechanisms that generate slow rhythms and sleep and wakefulness (Steriade et al., 1993b; Steriade and McCarley, 2005; Krueger et al., 2013; Crunelli et al., 2015); (2) relationship to mechanisms of conscious processing and attention (Koch, 2004; Castro-Alamancos, 2009; Harris and Thiele, 2011; Koch et al., 2016); (3) relationship between local sleep and wakefulness, and the loss of consciousness during sleep or seizures (Blumenfeld and Taylor, 2003; Siclari and Tononi, 2017); (4) the effect of different brain states, especially sleep, slow wave sleep, and REM sleep on learning and memory and consolidation of memories (Stickgold, 2005; Stickgold and Walker, 2005; Rasch and Born, 2007; Dudai et al., 2015); (5) the role of neuromodulatory circuits of the brain including norepinephrine, acetylcholine, serotonin, dopamine, orexin, and melanocorticotrophin in changing brain states (Steriade and McCarley, 2005; Scammell et al., 2017); (6) slow rhythms in comas (Chatelle et al., 2012).

While a lot is known about the various circuits that promote sleep or wakefulness, the circuits that are causal in flipping the brain from slow wave – where the cortical EEG and the membrane potential of the individual neurons can be dominated by less than 1Hz slow waves – to REM or awake states, and vice versa, have not been elucidated. It is not known how the variety of different awake-promoting neuromodulatory circuits interact to create awake conscious states, or which mechanisms determine sleep states. Whether recurrent activity in cortical circuits is necessary to generate Up states is not known. Whether all classes of inhibitory interneurons in all cortical layers fire in phase with slow waves and Up states (Valero and Buzsáki, 2019), or whether some class of interneurons trigger the end of each Up state is still not known.

Here we begin by examining what is known about UDS in general, with a focus on the initial studies in primary sensorimotor cortical areas, and then examine what is known about these states in very differently organized interconnected cortical areas, the entorhinal and hippocampal circuits. For each of these three areas, we first describe the properties of UDS and the structure – function relationship in light of this network activity and the anatomical properties of each



area. Then we review the effects of neuromodulators on UDS, particularly (but not exclusively) acetylcholine, and discuss the crucial issue of inhibition and participation of inhibitory neurons in UDS. Finally, we consider the role of UDS in cortico-hippocampal interactions and hippocampal processing during slow wave sleep, particularly focusing on sharp-wave ripples and memory consolidation.

Memory and conscious states are intimately entwined (Tulving, 1985). Conscious experience involves the formation of memories, the remembering of a past. The formation of long term memories, i.e. consolidation, is greatly benefited by sleep (for reviews see Diekelmann and Born, 2010; Chen and Wilson, 2017; Sara, 2017; Klinzing et al., 2019). During sleep hippocampal place cells have an above-chance probability to be co-active in similar sequences as during the awake state (Lee and Wilson, 2002). During slow wave sleep, neocortical activity in each Up state is associated with hippocampal sharp wave ripples (Siapas and Wilson, 1998; Sirota et al., 2003; Isomura et al., 2006; Mölle et al., 2006; Peyrache et al., 2009; Sullivan et al., 2011; Maingret et al., 2016; Headley et al., 2017). This interaction between neocortical slow rhythm of sleep, sharp wave ripples in hippocampus, and hippocampal replay is thought to be the basis of memory consolidation (for reviews, see Sirota and Buzsáki, 2005; Buzsáki, 2015; Klinzing et al., 2019; Todorova and Zugaro, 2020). For this reason, here we examine the neocortical-entorhinal-hippocampal dialogue observed through the lens of slow wave sleep and Up and Down states.

## UP AND DOWN STATES IN NEOCORTICAL CIRCUITS

While slow rhythms had been observed earlier, Steriade and colleagues were the first to recognize the relationship between the intracellular membrane potential and the cortical EEG (Steriade et al., 1993b,c,d). At about the same time intracellular recordings by Metherate and colleagues (Metherate et al., 1992; Metherate and Ashe, 1993; Cowan and Wilson, 1994) established the ubiquity of these rhythms in rodent sensory and motor cortices. This rhythm was apparent in cortical intracellular and whole cell recordings as spontaneous (less than 1 Hz) fluctuations in membrane potential, and could be observed in the EEG, local field potential (LFP), and in the multiunit activity (MUA) (**Figure 1A**). The terminology of Up and Down states, coined by Wilson (Wilson and Kawaguchi, 1996), arose out of the key observation that the distribution of membrane potentials for cortical neurons was bimodal: single cortical and medium spiny striatal neurons showed two stable membrane potential states, the depolarized Up state and the hyperpolarized Down state (Cowan and Wilson, 1994; Wilson and Kawaguchi, 1996). A bimodal distribution indicated that the mix of excitatory and inhibitory inputs to the neuron, and the intrinsic voltage-dependent conductances, together created two preferred membrane potentials for each neuron. From this early work, it was also clear that the Up state was a high conductance state and that in the Down state, the input resistance of neurons was higher and few neurons were actively providing synaptic

input to any cortical or striatal neuron (Cowan and Wilson, 1994; Paré et al., 1998; Destexhe and Paré, 1999; Waters and Helmchen, 2006). Finally, this work and work from the McCormick lab showed that the Up state was dominated by gamma band activity and that the fast post-synaptic potentials (PSPs) in each Up state could be inhibitory, that is Up states were a balanced mixture of excitation and inhibition (Cowan and Wilson, 1994; Steriade et al., 1996; Sanchez-Vives and McCormick, 2000; Shu et al., 2003; Haider et al., 2006).

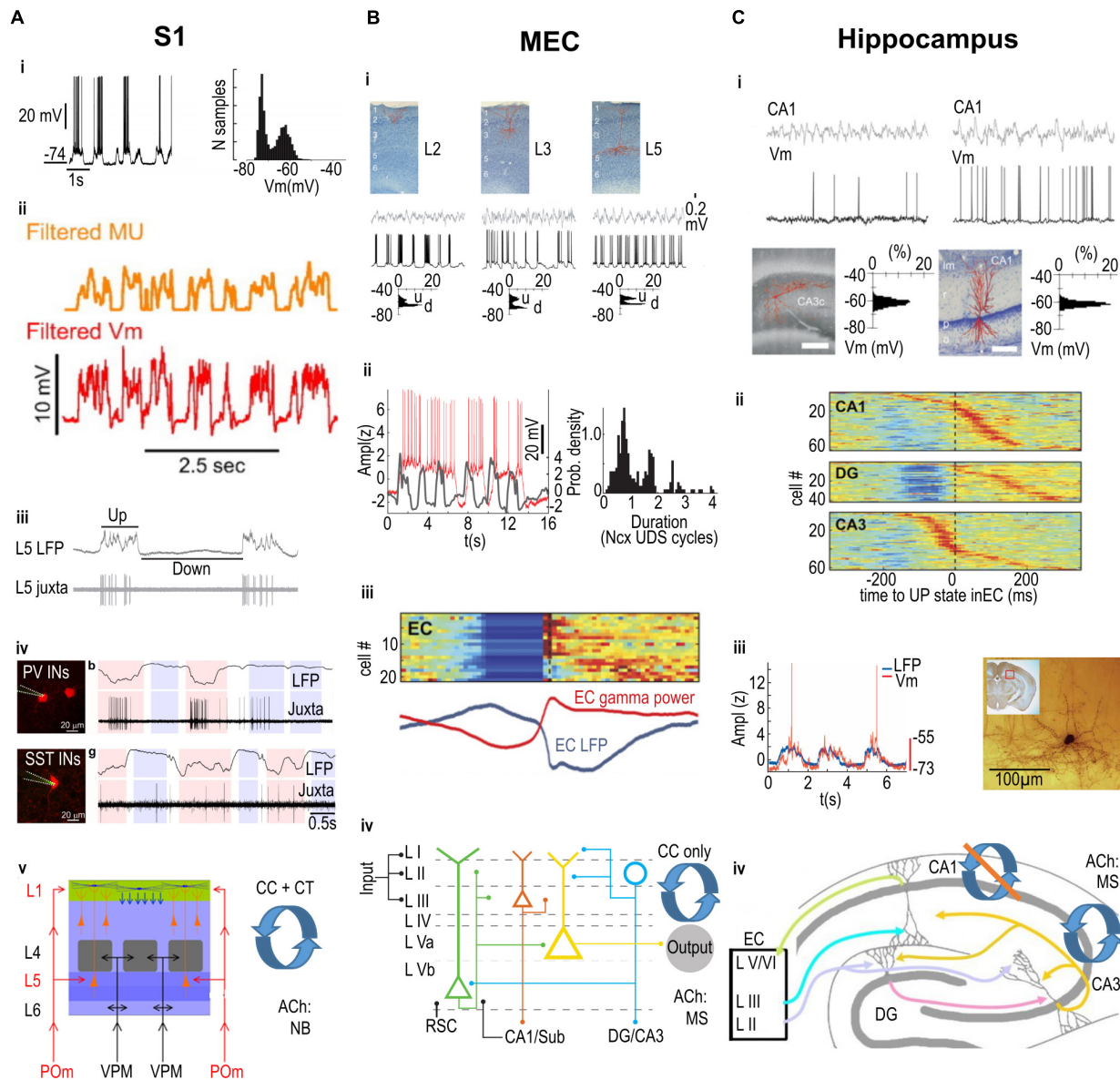
So within the slow wave sleep rhythm, two distinct brain states can be observed, one where the cortex, and ostensibly the brain are quiet, and another where cortex is in an activated, high gamma state (Destexhe et al., 2007). In the first papers describing this rhythm, the mechanisms of Up and Down states were explored in primary sensory cortical areas (Metherate et al., 1992; Metherate and Ashe, 1993; Steriade et al., 1993a,c,d; Cowan and Wilson, 1994). One idea was to examine whether these rhythms were generated intracortically, or thalamic inputs were necessary to generate the slow wave rhythms. This work indicated that even though thalamic neurons were in phase with cortical Up and Down states, large excitotoxic thalamic lesions, combined with transectioning of the corpus callosum had no effect on the generation of the slow rhythm in parietal cortex (Steriade et al., 1993d; but see Poulet et al., 2012; Zaghera et al., 2013). In fact, UDS can be elicited even in isolated brain slices (Sanchez-Vives and McCormick, 2000).

A second idea was that the brain stem arousal centers, cholinergic pedunculo-pontine tegmentum (PPT) and noradrenergic locus coeruleus, trigger Up states. Microstimulation in the PPT triggered persistent activity, an activated state that was evident as tonic depolarization in cortical neurons (Steriade et al., 1993a). Additionally, stimulation of the basal forebrain, another cholinergic site, elicited a desynchronization of the cortical EEG (Metherate et al., 1992; Metherate and Ashe, 1993).

Another hypothesis was that depolarizations reflected excitatory inputs, and the hyperpolarizations were a manifestation of inhibition. However, recordings from cortical pyramidal neurons *in vivo* and *in vitro* showed that the Up states were high conductance states, whereas down states were low conductance states; furthermore, the fast PSPs in the Up state membrane potential occurred in the gamma frequency, and reversed at the reversal potential of inhibition (Cowan and Wilson, 1994; Steriade et al., 1996; Shu et al., 2003; Hasenstaub et al., 2005; Haider et al., 2006; Waters and Helmchen, 2006). These experiments suggested that excitation and inhibition increased and decreased together, that the Up state was dominated by both excitation and inhibition, and the Down state was a state when cortical circuits were quiescent. In the last few years, these ideas have been investigated in greater detail in anesthesia, quiet wakefulness, and activated states.

## Up and Down States in the Somatosensory “Barrel” Cortex

The rodent barrel cortex is a model somatosensory system that has been used extensively for the study of Up and Down states. The organization, structure and function of this cortex have



**FIGURE 1 |** Summary of Up and Down state mechanisms across three cortical areas. **(A)** Primary somatosensory cortex (S1). i. An example intracellular recording from a neuron displaying depolarized Up states and hyperpolarized Down states (i, left panel), with a clear bimodal distribution of membrane potential values (i, right panel). ii. The multi-unit (MU) firing of neurons is correlated with the membrane potential (V<sub>m</sub>) fluctuations. iii. Juxtacellular recording illustrating that the firing of a pyramidal neuron (in this case in layer 5; L5) is linked to the Up state recorded in the local field potential (LFP). iv. Both Parvalbumin (PV, top row) and Somatostatin (SST, bottom row) expressing interneurons (INs) preferentially fire action potentials during Up states, but also fire a minority of spikes during Down states. v. Relevant anatomical properties of S1 barrel cortex include the laminar structure (left), with layer-specific inputs from specific thalamic nuclei (POm, VPM); recurrent corticocortical (CC) and corticothalamic (CT) connectivity; and cholinergic input from the Nucleus Basalis (NB). **(B)** Medial Entorhinal Cortex (MEC). i. Intracellularly recorded principal cells in layer 2 (L2), 3 (L3), and 5 (L5), shown in top row, display depolarized Up states and hyperpolarized Down states (middle row), with a clear bimodal distribution of membrane potential values (bottom row; u = Up State, d = Down state). ii. Left panel: Intracellularly recorded Up states in a L3 pyramidal cell (red trace) can extend over several simultaneously recorded Up states in the neocortex (gray trace). Right panel: The duration of MEC L3 pyramidal cell Up states is preferentially linked to a discrete number of neocortical Up and Down state (Ncx UDS) cycles. iii. Color-coded stacked cross-correlograms of Entorhinal Cortex (EC) neurons (one per row), recorded during slow wave sleep in naturally sleeping rats. Note the cross-correlograms are aligned to the transition from Down to Up states, and also correlate with the simultaneously recorded average LFP (blue) and gamma power (red) shown below. iv. Relevant anatomical properties of MEC include the laminar structure (left), with layer-specific inputs from other cortical areas including inputs from the retrosplenial cortex (RSC) and hippocampus (CA1/Sub) to the deep layers, as well as layer-specific outputs including outputs from L2 stellate cells (blue circle) to the hippocampal areas DG/CA3 and from L3 pyramidal cells (red triangle) to CA1/Sub; recurrent corticocortical (CC) connectivity only (i.e., no corticothalamic connectivity is known); and cholinergic input from the Medial Septum complex (MS). **(C)** Hippocampus. Intracellular recordings from pyramidal cells in areas CA3 (left) and CA1 (right); the filled cells are displayed on the bottom row. Note that neither the LFP (top row) nor the membrane potential (V<sub>m</sub>, second row) display clear Up and Down states, as shown in the unimodal

(Continued)

**FIGURE 1 | Continued**

distribution of Vm values (bottom row). ii. color-coded stacked cross-correlograms of DG, CA1, and CA3 neurons (one per row), recorded during slow wave sleep in a naturally sleeping rat (same rat as in panel **Biii.**), and aligned to the transition from Down to Up states. Note that there is a clear Down state only in DG (compare also to EC activity in **Biii.**). iii. Left: The intracellularly recorded membrane potential (red trace) of a CA1 interneuron fluctuates in phase with the simultaneously recorded neocortical LFP. Right: the filled interneuron from which the recording in the left panel was made, located in the “molecular layer” of CA1. iv. Relevant anatomical properties of the hippocampus, which has only a single layer of pyramidal cells extending from CA1 to CA3 (plus one layer in DG, with a strictly unidirectional connectivity), layer-specific inputs from superficial EC, and output from CA1 to deep layers of EC (other outputs not shown); recurrent connectivity is classically considered to be present amongst CA3, but not CA1 pyramidal cells; cholinergic input is from the Medial Septum complex (MS). (**Ai**) adapted by permission from Hasenstaub et al. (2007); (**Aii**) from Mease et al. (2016); (**Aiii**) from Zucca et al. (2017); (**Aiv**) from Castejon et al. (2016); (**Bi,iii,Ci,ii**) from Isomura et al. (2006); (**Bii**) from Hahn et al. (2012); (**Biv**) from Witter et al. (2017); (**Ciii**) from Hahn et al. (2006); (**Civ**) from Petrantonakis and Poirazi (2014).

been extensively reviewed (Kleinfeld et al., 2006; Petersen, 2007; Diamond et al., 2008; Sachdev et al., 2012; Feldmeyer et al., 2013). Briefly, the somatosensory cortex is a prototypical primary sensory cortex, it is 6 layered and receives thalamic inputs from lemniscal and paralemniscal thalamic nuclei. The lemniscal pathways carry touch, pressure, temperature, and proprioceptive input primarily targeting layer 4 and 6, and the paralemniscal pathways target layer 1 and 5. Thalamic input is excitatory, but the excitatory input can target inhibitory interneurons (Agmon and Connors, 1991). The canonical thalamocortical circuit (reviewed in Sachdev et al., 2012) consists of excitatory input from thalamus, that targets both excitatory and inhibitory neurons in cortex, with the excitatory input driving other cortical neurons in a feedforward manner, i.e., excitation is balanced by inhibition. When excitatory inputs are active, inhibitory inputs are activated within a few milliseconds. The somatostatin cortical interneuron in layer 2/3 and 5 is an exception to this rule that inhibitory neurons in cortex are co-activated with local excitatory neurons (Gentet et al., 2012). The principal cortical neuron, the pyramidal cell is the main output neuron of cortex, and it is the source of the majority of local recurrent and long-range excitation in cortex. Pyramidal cell bodies can be found in layer 2/3, 4, 5, and 6. It is worth noting that there is no known class of intrinsically active – pacemaker – excitatory neuron in somatosensory cortex, i.e., no neurons continue to fire action potentials in the absence of synaptic input, but there may be a class of inhibitory interneuron, a somatostatin positive neuron, that can fire intrinsically (Fanselow et al., 2008; Fanselow and Connors, 2010; see also Zucca et al., 2017).

## Up and Down States in the Medial Entorhinal Cortex

The entorhinal cortex (EC) connects the parietal, temporal and prefrontal cortices to the hippocampal formation; thus, it forms the interface between brain regions coding ongoing sensory information and storing long-term memory on the one hand, and those responsible for encoding and consolidating memories. The EC serves as the primary input and output structure of the hippocampus: there are two parallel input/output channels one in the lateral and the other in the medial entorhinal cortex each connected distinctly to the hippocampus. While primary sensory neocortical areas are all intimately linked to multiple thalamic nuclei, there is no evidence of strong monosynaptic thalamic inputs to the MEC. The persistence of up down state transitions

in the absence of thalamic inputs to the MEC indicates that there are several other motifs of recurrent connectivity that can provide the circuitry necessary for maintenance of UDS besides those found in the somatosensory cortex.

One consequence of the organization of the entorhinal cortex, specifically the fact that multiple neocortical areas provide excitatory long range input to this cortex, is that Up states in neocortex are likely to propagate into entorhinal cortex. Indeed, entorhinal neurons show UDS *in vivo* (Isomura et al., 2006; Hahn et al., 2012; **Figures 1Bi,ii**); and gamma oscillations are nested on up states (Isomura et al., 2006; **Figure 1Biii**). Surprisingly, even in the absence of long range input, *ex vivo* in brain slices, it is possible to generate UDS in both the lateral (Namiki et al., 2013) and medial (Mann et al., 2009; Tahvildari et al., 2012) subdivisions of the entorhinal cortices. *In vivo*, the Up states in MEC can last more than 2 s, almost double the duration of Up states in somatosensory and motor cortices (Cowan and Wilson, 1994; Sachdev et al., 2004; Hasenstaub et al., 2007; Hahn et al., 2012; **Figure 1Bii**; but also see Haider et al., 2006, 2013). One consequence of this long duration Up state in entorhinal cortex is that even though MEC activity is physically linked to neocortical Up states, activity in MEC can skip one or two cycles (Hahn et al., 2012). The longer Up states in MEC potentially arise from intrinsic membrane and synaptic properties of pyramidal neurons in MEC. Pyramidal cells in MEC have persistent firing and a Kainate receptor dependent slow glutamatergic current (Digby et al., 2017). This current is not found in pyramidal cells in primary somatosensory cortex (S1).

## Laminar Pattern of Up and Down States

There are two distinct possibilities for a laminar profile of activity during slow waves. First it is possible that a class of cortical neurons in a particular layer is more responsive to inputs or is intrinsically more active, or more excitable. A second possibility is that a class of inputs to a particular layer initiates each Up state. The studies that have examined this question come to very different conclusions. In ferret brain slices, Up states begin in layer 5 (L5) and propagate from there. *In vivo* in anesthetized mice, the picture is more complex. While L5 MUA can precede activity in other layers, activity in L4 can also precede the initiation of Up states in other cortical layers (Sanchez-Vives and McCormick, 2000).

In contrast, in MEC, L3 pyramidal neurons show the strongest correlation to Up and Down states. Layer 3 neurons maintain and propagate the UDS activity in the MEC; L3 pyramidal cells



are the most active population of excitatory cell types across all layers in the MEC. The strong role for L3 pyramidal neurons in maintaining and propagating the UDS activity in the MEC arises from: (1) the strong recurrent connections that target other layers (Dhillon and Jones, 2000; Winterer et al., 2017); (2) the slow decay kinetics of EPSPs that facilitate synaptic integration and spiking at lower frequencies (Gloveli et al., 1997).

Stellate cells in MEC L2 can also fire in phase with UP states, but fire much less than the L3 pyramidal neurons. This is in part due to strong inhibition of the L2 stellate cells (Beed et al., 2013), which reduces the subthreshold depolarization on individual cells during each Up state. Inhibition is also likely to bring about the Up to Down state transition which we discuss in more detail in the next section.

## ROLE OF INHIBITION AND PARTICIPATION OF INHIBITORY NEURONS DURING UDS

### Inhibition and Up and Down States in Neocortex

One simple hypothesis regarding Down states is that they are a reflection of increased inhibition. This hypothesis implies that the balance between excitation and inhibition shifts during each Up state, and that classes of inhibitory interneurons are active at the end of the Up state and terminate the Up state. Furthermore, these neurons could continue to fire in the Down state, maintaining the hyperpolarized Down state. A related hypothesis would be that a class of inhibitory neurons is intrinsically active requiring no excitatory input, or that the class of interneurons that trigger Down states are part of a distinct cortical network, that is partly out of phase with the global EEG and LFP recorded in cortex. These ideas have spurred a large number of additional investigations tracking whether interneurons all fire in phase with cortical Up states (Zucca et al., 2017, 2019; Knoblich et al., 2019; Senzai et al., 2019; Valero and Buzsáki, 2019).

As mentioned above, a large variety of cortical inhibitory interneurons, distributed throughout all cortical layers, have been described (Markram et al., 2004; Tremblay et al., 2016; Feldmeyer et al., 2018). The neurogliaform cells are the most common interneuron type found in layer 1 in S1; they are also found in layers 2–6 (Kawaguchi and Kubota, 1997; Ascoli et al., 2008; Kubota, 2014). Parvalbumin positive neurons are found in layers 2–6; they are the classical fast spiking, soma targeting (basket cells), and axon targeting (chandelier cells) neurons (Kawaguchi and Kubota, 1997; Ascoli et al., 2008). The somatostatin positive (SST) interneurons are a class of mostly non-fast spiking, dendrite targeting neurons, including the Martinotti cell found in layer 2/3 and 5/6. Non-Martinotti SST neurons can be found in layer 4 (Kawaguchi and Kubota, 1997; Wang et al., 2004; Ma et al., 2006; Xu et al., 2013; Kubota, 2014). Another important class of interneurons, characterized by the expression of vasointestinal peptide (VIP), is part of a

disinhibitory circuit with the SST neurons and are found in layers 1–6 (Caputi et al., 2009; Cauli et al., 2014; Prönnke et al., 2015).

These classes of interneurons have been studied extensively especially in quietly sitting awake animals and in anesthetized animals. A large body of work shows that low frequency oscillations are often evident in the quietly sitting awake state, and that these low frequency oscillations can be abolished by movement. In rodents, whisking activates the cortex, and the cortical EEG and membrane potential of pyramidal neurons reflects this change in state (Petersen et al., 2003; Crochet and Petersen, 2006; Poulet and Petersen, 2008; Gentet et al., 2010; Pala and Petersen, 2018). During quiet wakefulness, pyramidal neurons, fast spiking interneurons, and non-fast spiking SST neurons fire in phase with the slow rhythm. Activation of the cortex by whisker motion triggers a decrease in spiking of pyramidal and parvalbumin-positive interneurons. In contrast during activated states, the somatostatin positive interneurons are persistently depolarized and increase their firing rate suggesting that in the awake state when the mice whisk, this class of interneurons is linked to a distinct class of inputs that do not drive the pyramidal neuron or the fast spiking parvalbumin neurons.

In anesthetized animals, juxtacellular recordings also show that parvalbumin and SST neurons fire mostly in phase with Up states. However, even though the activity of both PV and SST interneurons decreases during Down states, it is not completely abolished (Zucca et al., 2017, 2019). Something very similar is observed in the visual cortex of anesthetized mice (Knoblich et al., 2019). In this study, while the membrane potential of pyramidal and parvalbumin positive neurons was bimodally distributed, SST neurons could be divided into two classes, one that fired in phase with Up and Down states – its membrane potential was bimodally distributed – and the other class of SST neurons that continued firing even in Down states and its membrane potential was unimodally distributed (Knoblich et al., 2019). Another recent study shows that a population of neurons in L6 of visual cortex fired specifically in Down states (Senzai et al., 2019). The anatomical identity of these extracellularly recorded neurons is not known. Finally, in prefrontal cortex of naturally sleeping mice SST positive cells fired in phase with the slow wave sleep rhythm (Funk et al., 2017). The spontaneous firing patterns of most other classes of interneurons have not been studied in the context of Up and Down states and have not been linked to slow oscillations.

### Inhibition and Up and Down States in the Medial Entorhinal Cortex

Superficial layers (layers 2 and 3) of the MEC are embedded in a dense inhibitory network. Parvalbumin positive, perisomatic targeting inhibitory neurons strongly modulate the output of both L2 stellate and L3 pyramidal cells, thereby organizing the circuitry predominantly at oscillatory frequencies in the gamma range. This pattern of inhibition is in stark contrast to the pattern seen in somatosensory cortex, where there is as yet no evidence of differential levels of inhibition in the different cortical layers. In the MEC on the other hand, there is a strong



gradient of inhibition along the dorsoventral axis. There is almost twofold stronger inhibition in the dorsal MEC compared to the inhibition at the ventral levels. This differential gradient is related to input from parvalbumin-expressing (PV) interneurons (Beed et al., 2013).

In *ex vivo* slice preparations of entorhinal cortex, it is possible to obtain Up and Down oscillatory activity in the superficial layers of the MEC (Mann et al., 2009; Tahvildari et al., 2012; Salkoff et al., 2015). Excitatory and inhibitory neurons in the layers 2 and 3 of the MEC take part during UDS (Tahvildari et al., 2012; Neske et al., 2015). Parvalbumin interneurons in layers 2 and 3 also participate in the slow rhythm and are strongly depolarized, firing action potentials in the Up state in gamma frequency range (Salkoff et al., 2015). In contrast, one subset of SST interneurons is only weakly driven by each Up state and fires few spikes (Tahvildari et al., 2012), whereas the fast spiking SST interneurons fire in phase with each Up state (Neske et al., 2015). Other work has shown that inhibition, in particular GABA-B receptor dependent inhibition, can play a role in terminating Up states (Mann et al., 2009; Craig et al., 2013). While a buildup of slow inhibitory effects is possible for terminating each Up state, no class of interneurons has shown a pattern where their activity increases over the course of an Up state, or where the neurons fire out of phase with each Up state.

There are multiple other classes of interneurons in MEC including those identified by expression of serotonin receptor type 3a (5HT3a), cholecystokinin (CCK) and VIP, but in the *ex vivo* slice model at least, these classes of interneurons are negligibly driven by each Up state (Tahvildari et al., 2012).

## NEUROMODULATION OF UDS

### Neuromodulation of Up and Down States in the Somatosensory Cortex

As mentioned above, brainstem and hypothalamic circuits have been linked to an ascending activating system. This system is the source of many neuromodulators (reviewed in Scammell et al., 2017). Here we will focus on two classical neuromodulators, acetylcholine and norepinephrine.

#### Acetylcholine

One source of acetylcholine in the brain is the brainstem pedunculopontine nucleus that has neurons that project to thalamus, hypothalamus and basal forebrain. But the cholinergic fibers that target the somatosensory cortex arise from the nucleus basalis (McKinney et al., 1983; Mesulam et al., 1983; Eckenstein et al., 1988). Fibers immunoreactive for the acetylcholine synthetic enzyme, choline acetyltransferase or stained for the cholinergic degradative enzyme, acetylcholine esterase can be detected in all layers of the rat parietal cortex, but the innervation to laminae I–III and V is particularly dense (Kristt, 1979; Houser et al., 1985; Lysakowski et al., 1989; Umbriaco et al., 1994). In barrel cortex, the acetylcholinesterase fibers are dense in all layers except for layer 4 (Sachdev et al., 1998).

Microsimulation in the brainstem, in the pedunculopontine nuclei can trigger persistent activity, an activated state, that is

evident as a tonic depolarization in cortical neurons (Steriade et al., 1993a). The effect of the stimulation in changing the cortical state, can take two pathways, one where the cholinergic PPT neurons activate thalamus (Crunelli et al., 2015), a second where the cholinergic PPT neurons activate the basal forebrain neurons. In line with the second mechanism, stimulation of the basal forebrain can also elicit a acetylcholine dependent desynchronization of the cortical EEG (Metherate et al., 1992; Metherate and Ashe, 1993).

The cholinergic PPT neurons and the basal forebrain neurons increase their activity during wakefulness (McCarley and Hobson, 1975; Lee et al., 2004; Hassani et al., 2009). One difference between the brainstem cholinergic neurons and the basal forebrain cholinergic neurons is that while 80% of the pedunculopontine cholinergic neurons fire in phase with cortical Up and Down states (Mena-Segovia et al., 2008; Ros et al., 2010; Mena-Segovia and Bolam, 2011) the basal forebrain neurons show no clear phasic activity with cortical Up and Down states (Lee et al., 2004; Hassani et al., 2009). Indeed, even though the activity of basal forebrain axons in S1 increases and is related to the generation of activated states, the effect of cholinergic activity is to oppose the movement-triggered activated and persistently depolarized state observed in barrel cortex (Eggermann et al., 2014).

#### Norepinephrine

The locus coeruleus is the exclusive source of norepinephrine in the central nervous system, with the individual axons of these neurons traversing widespread cortical areas (Dahlstroem and Fuxe, 1964; Swanson and Hartman, 1975; Sara, 2009). The noradrenergic fibers traverse all cortical layers in somatosensory barrel cortex, including layer 4 (Simpson et al., 2006).

Just like the brainstem pedunculopontine neurons, the locus coeruleus neurons fire in phase with cortical Up and Down states (Lestienne et al., 1997; Eschenko et al., 2012) and they are more active during wakefulness than during slow wave sleep (Aston-Jones and Bloom, 1981). Optogenetic stimulation of these neurons can cause an immediate transition from sleep to wakefulness (Dringenberg and Vanderwolf, 1997; Carter et al., 2010). Furthermore, in the awake and activated state, activity of the locus coeruleus neurons promotes persistent activity and a desynchronized state (Constantinople and Bruno, 2011).

One caveat with most of this work, is that even though the study of the brainstem activating system has a long history, the study of how these systems interact with each other and with the hypothalamic hypocretin/orexin and neurons is still in its infancy (Scammell et al., 2017).

### Neuromodulation of Up and Down States in the Medial Entorhinal Cortex

Acetylcholine is a key neuromodulator of the MEC. In contrast to the somatosensory cortex, which receives cholinergic inputs from the nucleus basalis, the bulk of the cholinergic input to the MEC arises from the medial septum (Lewis and Shute, 1967; Köhler et al., 1984; Gonzalez-Sulser et al., 2014; Desikan et al., 2018).

The medial septum cholinergic neurons are key players in generating theta oscillations, a marker of attention and

arousal. This rhythm is absent during sleep suggesting that the cholinergic tone in the entorhinal cortex is low during slow wave sleep. Medial septum cholinergic fibers strongly modulate and innervate both excitatory and inhibitory neurons (Desikan et al., 2018). Fast and slow cholinergic currents depolarize both the SST and 5HT3a interneuron populations whereas the slow cholinergic currents mostly hyperpolarize the excitatory cells and the PV interneurons (Desikan et al., 2018). The bimodal effect of cholinergic modulation on the different cell types coupled with their respective roles during Up and Down states makes it an important neuromodulatory candidate for brain states in the MEC.

In addition, acetylcholine released from cholinergic fibers has a profound effect both on subthreshold membrane potential oscillations (Heys et al., 2010) as well as spiking activity in the MEC (Heys et al., 2012; Tsuno et al., 2013). The application of cholinergic agonists on *ex vivo* slice preparation can induce persistent firing (i.e., Up states) in pyramidal neurons in the MEC (Tahvildari et al., 2012; Jochems et al., 2013). This effect of acetylcholine is unique for neurons in the entorhinal cortices and is not observed in other cortical pyramidal neurons. Although a lot is known about the *ex vivo* effects of acetylcholine, very little is known about the spatio-temporal organization of cholinergic inputs during changes in brain state in the MEC.

## CORTICO – HIPPOCAMPAL INTERACTION AND HIPPOCAMPAL PROCESSING DURING UDS

### Memory Consolidation, Hippocampal Replay, and Sharp-Wave Ripples in UDS

Memory consolidation is a key function associated with slow wave sleep. In the hippocampus, memory consolidation has been associated with sharp wave ripples (SWRs), consisting of a high-frequency (~150 Hz) “ripple” riding on top of a depolarizing “sharp-wave” event (see Buzsáki, 2015, for a review). SWRs appear to be generated in area CA3 and propagate via CA1 and the subiculum into the cortex (Buzsáki et al., 1983; Chrobak and Buzsáki, 1996; Csicsvari et al., 2000; Khodagholy et al., 2017; Nitzan et al., 2020). A process called “replay” has also been linked to SWRs; during replay when the animal is at rest or asleep, cell assemblies active during waking information-acquiring behaviors are reactivated (for a review, see Foster, 2017). One hypothesis is that replay strengthens synaptic connections in the neocortex, thereby mediating a “transfer” of memory from the short-term hippocampal system to the long-term neocortical system (Marr, 1971; Buzsáki, 1989; Chen and Wilson, 2017).

Evidence in support of this idea has shown that: (1) replay in hippocampus and other cortical areas is significantly correlated (Ji and Wilson, 2007; Ólafsdóttir et al., 2016; Rothschild et al., 2017); (2) interrupting hippocampal SWRs (and thus the replay events associated with them) during post-learning sleep reduces memory consolidation (Girardeau et al., 2009; Ego-Stengel and Wilson, 2010); (3) ripple activity in hippocampus is coupled to cortical activity in many brain areas (Chrobak and Buzsáki,

1994, 1996; Logothetis et al., 2012; Wang and Ikemoto, 2016; Headley et al., 2017; Khodagholy et al., 2017), including cortical replay events in the prefrontal cortex (PFC; Peyrache et al., 2009); and (4) strengthening this coupling by precisely timed SWR-locked stimulation of the PFC improves memory consolidation (Maingret et al., 2016).

During slow wave sleep, SWRs in the hippocampus appear remarkably closely linked to neocortical Up and Down states. Some studies report that the SWRs in CA1 are more prevalent during Down states (Battaglia et al., 2004), their occurrence peaking just before the Down-Up transition (see also Khodagholy et al., 2017). Other studies report that SWRs are more prevalent during Up states as defined by: (1) EEG, local field potential or multiunit activity in prefrontal cortex (PFC; Siapas and Wilson, 1998; Mölle et al., 2006; Isomura et al., 2006; Peyrache et al., 2009; Maingret et al., 2016); (2) MUA and LFP in somatosensory cortex (Sirota et al., 2003); (3) MUA in entorhinal cortex (Isomura et al., 2006; Sullivan et al., 2011) or (4) pooled MUA from several cortical sites (Headley et al., 2017). Taken together SWRs seem to be coupled to cortical Up and Down states.

This coordination between cortex and hippocampus is likely to be important for memory consolidation. A number of studies have looked into the coordination of cortical and hippocampal activity in the context of SWR propagation, memory consolidation, and replay. These studies show that both fast oscillations and unit activity in MEC and other retrohippocampal areas are coupled to hippocampal ripples (Chrobak and Buzsáki, 1994, 1996). The functional consequences of this feedback to the MEC are not known, but the inputs from CA1 and subiculum, impinging mostly on MEC L5, could support the coordinated replay between CA1 and the deep layers of MEC during sleep (Ólafsdóttir et al., 2016). In contrast, recordings from the superficial layers of MEC in awake animals reveal that replay in the MEC can also be independent of replay in CA1 (O'Neill et al., 2017). Thus, although it is possible that replay in deep MEC layers reflects replay of information from hippocampus, replay in the superficial layers of MEC may serve an as yet unknown function (Trimper et al., 2017).

In the context of memory consolidation, the direct output from CA1 to the prefrontal cortex (Jay and Witter, 1991) plays an important role (Takehara-Nishiuchi et al., 2006): (1) PFC engram cells can be re-activated optogenetically, and reactivation induces recall (Kitamura et al., 2017). (2) Performance of a spatial memory task induces increased hippocampal-PFC coupling in which optogenetic activation time-locked to hippocampal ripples in the post-task sleep epoch improves memory consolidation (Maingret et al., 2016). Importantly, such consolidation is not limited to areas monosynaptically connected to the hippocampus (auditory cortex; Rothschild et al., 2017).

Besides projecting to retrohippocampal areas and PFC, CA1 axons also extend to other brain areas (Groen and Wyss, 1990; Cenquizca and Swanson, 2006, 2007). But the bulk of the hippocampal outputs to cortical and diencephalic targets arises from the subiculum (Witter and Groenewegen, 1990; Kim and Spruston, 2012). The axons of individual subicular neurons extend to several target areas (Cembrowski et al., 2018), consistent with the idea that the subicular activity transforms

the sparse spatial code in area CA1 (Thompson and Best, 1989) into a dense, distributed, redundant representation in the target areas (Kim et al., 2012). This organization makes it possible for any subset of output neurons to convey the same information to different downstream targets. While this has important implications for memory consolidation during slow wave sleep, very little is known about replay or slow wave sleep rhythms in the subiculum (Isomura et al., 2006). Nevertheless, the prime target of the subiculum, the retrosplenial cortex, can show learning-dependent, and memory-retrieval-predicting memory engrams, similar to those described above for PFC (Cowansage et al., 2014; Milczarek et al., 2018; see also Maviel et al., 2004).

## Up and Down States in the Hippocampus

As mentioned above, the generation of SWRs in CA3 (and downstream area CA1) is important for replay and memory consolidation. The circuit mechanisms that link SWRs to cortical Up and Down states are only dimly understood. Do cortical Up and Down states, particularly in the entorhinal cortex, affect activity in the hippocampus?

In contrast to both entorhinal cortex and somatosensory cortex, the hippocampus does not display clear epochs of Up and Down states; instead, hippocampal LFP activity during slow wave sleep has traditionally been characterized as “irregular” (for an early review of hippocampal activity, see O’Keefe and Nadel, 1978). Intracellular recordings in urethane-anesthetized mice (Hahn et al., 2007) and rats (Isomura et al., 2006) revealed unimodal, not bimodal, distributions of the membrane potential of hippocampal pyramidal neurons in CA1 and CA3 (**Figure 1Ci**). There was no clear depolarization related to Up states. The activity of granule cells in the dentate gyrus (DG), the target of entorhinal inputs via the perforant pathway, is skewed and only somewhat bimodal (Isomura et al., 2006; Hahn et al., 2007), likely reflecting a more direct and powerful cortical input to this structure. Finally, consistent with the extensive neocortical, paleocortical, and thalamic connectivity of the subiculum (Ding, 2013), the membrane potential of pyramidal cells in the subiculum, the principal output structure of the hippocampus, is clearly bimodal (Isomura et al., 2006).

Here for the sake of brevity, we focus on hippocampal area CA1. Although, as mentioned above, CA1 does not show the typical bimodal membrane potential distribution described for other cortical areas, neurons in CA1 do receive input from the MEC (**Figure 1Civ**), and MEC neurons display Up and Down states in their membrane potential. Bimodality in the neocortical membrane potential arises from some as yet undefined combination of recurrent connectivity, short term plasticity mechanisms, thalamic inputs, and intrinsic voltage gated conductances. Note that even if the membrane potential of single neurons is not bimodally distributed, inputs during each Up state could still increase membrane potential noise and trigger action potentials. Indeed, the firing of individual pyramidal neurons in CA1 is coupled to Up states recorded in S1 in both mice and rats (Sirota et al., 2003). Spikes in CA1 neurons occur just after the trough of S1 delta waves, just a bit after spikes in S1 units. This suggests that the synchronous cortical activity in

an Up state can drive CA1. The firing of CA1 pyramidal neurons and the membrane potential of these neurons is also coupled to delta waves and Up-states in other cortical areas in the rat, including entorhinal cortex (Isomura et al., 2006; **Figure 1Cii**).

But the coupling of CA1 multi-unit activity reflecting the activity of CA1 pyramids, was more strongly modulated by the “persistent” Up states in MEC layer 3 than by Up states recorded in other neocortical areas (Hahn et al., 2012; **Figure 1Bii**). This coupling is consistent with the fact that axons from layer 3 pyramidal cells directly innervate CA1 pyramidal cells (**Figure 1Civ**). The depth of UDS modulation of the firing rates of CA1 pyramidal cells also appears to be subtype-specific: modulation strength differed between two types of CA1 pyramidal cells, identified by their sublayer-location within the pyramidal cell layer (Mizuseki et al., 2011). Finally, the LFP in CA1 stratum lacunosum moleculare (SLM) also reflects the influence of EC inputs, which impinge in this layer in CA1 and drive gamma oscillations locked to local slow oscillations and to the MEC Up state (Isomura et al., 2006; Wolansky et al., 2006; Sullivan et al., 2011).

So why is the cortical input to CA1 pyramidal cells not sufficient to induce large depolarizations in each Up state? The absence of a strong thalamic inputs could play a role. However, this is a feature the hippocampus shares with the entorhinal cortex (**Figure 1Biv**) and, as pointed out above, entorhinal pyramidal cells display bimodality in their membrane potential (**Figure 1Bi**) and can locally generate Up-Down states. The difference between CA1 and entorhinal or neocortical circuits, could be related to the stratified organization of inputs in CA1, where entorhinal axons are restricted to the SLM and contact the distal apical tufts of CA1 pyramidal cells. It could also be due to the paucity of recurrent excitatory connections between CA1 pyramidal cells (<1% of recorded pairs; Knowles and Schwartzkroin, 1981; Deuchars and Thomson, 1996; **Figure 1Civ**). This is in contrast to the neocortex, where one organizing feature of neocortical pyramidal neurons is that all pyramidal neurons have a local recurrent excitatory axon, providing excitation to pyramidal neurons and to the local inhibitory interneurons. Note, however, that recent work suggests extensive recurrent connections along the longitudinal axis of the hippocampus (Yang et al., 2014). Finally, the difference between CA1, entorhinal and neocortical circuits could also be related to intrinsic properties, for example persistent activity of CA1 pyramidal cells appears to be supported by intrinsic mechanisms (Knauer et al., 2013).

## Inhibition and Up and Down States in the Hippocampus

The axons from entorhinal L3 pyramidal cells contact CA1 pyramidal cells and interneurons, including those located at the border of stratum radiatum (SR) and SLM. These interneurons fire in phase with neocortical UP states (Hahn et al., 2006; **Figure 1Ciii**) and could be a source of feedforward inhibition to CA1 pyramidal cells (Vida et al., 1998). Interneurons at the SR-SLM border also receive inputs from pyramidal neurons in MEC layer 2 in mice (Kitamura et al., 2014) and express either nNOS, suggesting they include neurogliaform cells (Price



et al., 2005), or calbindin, suggesting dendritic-layer-innervating or periformant-path-associated CCK interneurons (Klausberger et al., 2005; Lasztóczy et al., 2011). CCK interneurons in turn can also innervate DG. Whether all classes of interneurons in CA1 fire in phase with Up and Down states has not been established (Isomura et al., 2006).

In general, global brain states modulate the level and type of activity of many interneuron types: (1) juxtacellular recordings show that SST-expressing O-LM cells decrease their firing rates during sleep compared to their activity during awake behaviors (Katona et al., 2014). The sleep-related activity of the also SST-expressing bistratified cells was more heterogeneous. (2) parvalbumin-expressing basket cells on the other hand increase their firing rate during slow wave sleep, whereas Ivy cells, the most common interneuron type in CA1 (Fuentelba et al., 2008), were not modulated by global brain state (Lapray et al., 2012); (3) Calcium imaging of activity shows that somatostatin and parvalbumin cell activity decreases during slow wave sleep compared to the awake state (Niethard et al., 2016). One limitation of the calcium imaging approach is that the different SST-expressing (e.g., bistratified versus O-LM) and PV-expressing (e.g., basket vs. axo-axonic) interneuron types cannot be distinguished when Ca-sensor expression is driven by just a single genetic marker (Klausberger and Somogyi, 2008). Taken together, these studies suggest that during slow wave sleep the amount of dendritic inhibition is reduced when compared to the awake state (Niethard et al., 2017). It is possible that this reduction in inhibition is an important component enabling memory-specific upregulation of synapse formation.

To gain a deeper understanding of how the activity of different interneuron types is modulated a more sophisticated genetic tagging approach (e.g., Fenno et al., 2014) or single-cell methods like juxtacellular or whole-cell recording are necessary (Tukker, 2019), preferably from naturally sleeping animals (Averkin et al., 2016).

## Neuromodulation of Up and Down States in the Hippocampus

As in other cortices, the hippocampal slow oscillations and ripple initiation are modulated by the action of a variety of neuromodulators including acetylcholine (Wolansky et al., 2006; Vandecasteele et al., 2014), dopamine (McNamara and Dupret, 2017), serotonin (Ponomarenko et al., 2003; Ul Haq et al., 2016b), noradrenalin (Ul Haq et al., 2016a; Swift et al., 2018), and others (see Atherton et al., 2015; Buzsáki, 2015).

Here we focus on the role of acetylcholine because it has been shown to have a strong effect on SWR initiation. Optogenetic stimulation of ChAT-expressing neurons in the medial septum reduces the incidence of SWRs, and suppresses the power of slow oscillations in CA1 (Vandecasteele et al., 2014). Systemically administered cholinergic agonists abolish hippocampal slow oscillations, whereas antagonists enhance them (Wolansky et al., 2006). This is consistent with a general role for ACh from the basal forebrain in shifting overall sleep-state in the neocortex (Han et al., 2014; Xu et al., 2015). The precise circuit mechanisms underlying direct effects of acetylcholine in the hippocampus

### BOX 2 | Novel Methods and Considerations Related to Replay and Memory

Novel methods are likely to increase our understanding of the role of hippocampal outputs in memory consolidation during sleep (Chen and Wilson, 2017; Maboudi et al., 2018). Currently, the mechanisms that drive replay, and the role of the different inhibitory interneurons and neuromodulators in this process are only dimly understood. Beyond the relatively short SWR events associated with replay and memory consolidation, the possible role of hippocampal outputs during slow wave sleep also remains unclear. In this respect, it might be relevant to make a distinction between *deep* sleep stages, where homeostatic mechanisms are more important, and *light* sleep stages, where synaptic potentiation and thus memory consolidation is more likely to occur (Genzel et al., 2014). Up and Down states differ during the different stages of sleep, in part due to differences in neuromodulator levels (Genzel et al., 2014; Krishnan et al., 2016), although particularly in rodents these differences have not been thoroughly investigated.

The levels of most neuromodulators also differ markedly between sleep and awake states. This may help explain why hippocampal replay during wakefulness differs from that during sleep: awake replay can also occur in reverse order, and is likely to have additional roles besides memory consolidation (Foster and Wilson, 2006; O'Neill et al., 2006; Diba and Buzsáki, 2007; Jadhav et al., 2012; Roux et al., 2017). It is important to emphasize that even within awake periods there are large differences between the active and the resting state (McGinley et al., 2015; Poulet and Crochet, 2018; Kay and Frank, 2019). SWRs and replay have been observed during sleep, and during active and resting awake states, but are likely to involve different mechanisms and play different roles, at least in part related to acetylcholine and dopamine levels (Atherton et al., 2015).

Although it is increasingly clear that Up and Down states are also present in awake animals (e.g., Petersen et al., 2003; Ferezou et al., 2007; Vyazovskiy et al., 2011; Senzai et al., 2019), the precise differences of UDS generation and propagation between various awake and sleep states remain to be elucidated. During sleep, different subclasses of UDS were recently shown to be relevant for the consolidation of neuroprosthetic skill learning: optogenetically driving slow wave oscillations in rat primary motor area M1 improved consolidation, whereas delta waves supported forgetting of the task (Kim et al., 2019).

Whether this is also the case for episodic memory remains to be seen. There are important differences in the consolidation of procedural versus declarative memory (Headley and Paré, 2017), including the role of the hippocampus (Squire and Zola-Morgan, 2015). It is also helpful to distinguish between different types of declarative memory: semantic and episodic memory are likely to involve dopaminergic and noradrenergic systems, respectively (Duszkiewicz et al., 2019).

are only dimly understood. The cell-type-specific synaptic and non-synaptic effects of acetylcholine on pyramidal cells and interneurons are not clear (Teles-Grilo Ruivo and Mellor, 2013; Dannenberg et al., 2017). It should also be noted that the systemic effects of cholinergic agonists and antagonists, as well as the effects of optogenetic or chemogenetic manipulations of cholinergic neurons in the medial septum, could be indirectly mediated via local network effects in the medial septum, which also sends GABAergic and glutamatergic projections to the hippocampal formation, with which it is reciprocally connected (for recent reviews, see Tsanov, 2015; Müller and Remy, 2018).

## SUMMARY AND OUTLOOK

Spontaneous activity in cortical circuits is generated internally during sleep and in quiet wakefulness, and propagates through the brain linking a variety of cortical and subcortical structures.



While the links between neocortex, entorhinal cortex and hippocampus persist during the different brain states, the functional consequences of the activity change from state to state. In the awake state sensory information propagates to entorhinal cortex, and hippocampus, and during slow wave sleep and sharp wave ripples, the memories of the day are consolidated in neocortex with a replay of the day's activity. To establish how this actually occurs, experiments imaging and manipulating the activity of axons that provide cortical and subcortical inputs to hippocampus during sensory information acquisition states and during replay will be necessary (Suh et al., 2011; Kaifosh et al., 2013; Kitamura et al., 2017; Yamamoto and Tonegawa, 2017; Qin et al., 2019), ideally in combination with novel methods of analysis and new approaches differentiating between different types of memory and awake/sleep states (**Box 2**). It would be useful to also image the dendritic activity of classes of neurons during these different states, to establish whether the active properties of dendrites, or the occurrence of NMDA/calcium spikes (Palmer et al., 2012, 2014; Larkum, 2013; Bittner et al., 2015, 2017; Takahashi et al., 2016; Schmidt-Hieber et al., 2017; Sheffield et al., 2017) have a role in replay, or memory consolidation and memory formation. It is still too early to tell whether changing the coupling between dendrites and soma has a role in learning or memory (Suzuki and Larkum, 2020), and whether the hypothalamic sleep and wake promoting neurons that are linked to other neuromodulatory systems and to neocortex have a role in the dialogue between hippocampus and neocortex (Zolnik et al., 2020).

In this review, we examined how the simple slow rhythm of the Up and Down states propagates from neocortex to hippocampus and back. A deeper understanding of the hippocampo-cortical dialogue and the relation between slow oscillations and memory consolidation is likely to be important for translational approaches targeting several brain disorders, including schizophrenia (Phillips et al., 2012), Alzheimer's disease (Mander et al., 2015), and epilepsy (Gelinas et al., 2016). However, as is obvious

from this review, the complete story of the interaction between these areas during sleep (and wakefulness) and the mechanisms that trigger memory consolidation is still to be told.

## AUTHOR CONTRIBUTIONS

The initial concept of this review manuscript was conceived by RS and PB and discussed by all authors. JT, PB, and RS wrote the initial manuscript. JT prepared the figure. All authors edited and discussed.

## FUNDING

The following funding sources have supported this project: (1) Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Grant No. LA 3442/6-1 to ML; Grant Nos. LA 3442/3-1 and LA 3442/5-1 to ML. (2) DFG - Project number 327654276 - SFB 1315 to ML and DS. (3) European Commission Horizon 2020 Research and Innovation Program and Euratom Research and Training Program 2014-2018 (under grant agreement no. 670118 to ML). (4) Human Brain Project HBP FPA No. 650003 (EU-Commission Grant 720270, HBP SGA1 and SGA2, Context-sensitive Multisensory Object Recognition: A Deep Network Model Constrained by Multi-Level, Multi-Species Data to ML). (5) Einstein Foundation Berlin to DS, Grant Nos. EVF-2017-363 and A-2016-311 to ML. (6) Stiftung Charité to PB. (7) ERC BrainPlay-Synergy grant to DS. (8) DFG NeuroCure Cluster of Excellence EXC-2049-390688087 to DS and ML.

## ACKNOWLEDGMENTS

We would like to thank Larkum and Schmitz lab members for useful discussions, and Linda Hahn-Tukker for help in making the figure.

## REFERENCES

- Agmon, A., and Connors, B. W. (1991). Thalamocortical responses of mouse somatosensory (barrel) cortex in vitro. *Neuroscience* 41, 365–379. doi: 10.1016/0306-4522(91)90333-j
- Ascoli, G. A., Alonso-Nanclares, L., Anderson, S. A., Barrionuevo, G., Benavides-Piccion, R., Burkhalter, A., et al. (2008). Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* 9, 557–568. doi: 10.1038/nrn2402
- Aston-Jones, G., and Bloom, F. E. (1981). Activity of norepinephrine-containing locus coeruleus neurons in behaving rats anticipates fluctuations in the sleep-waking cycle. *J. Neurosci.* 1, 876–886. doi: 10.1523/jneurosci.01-08-00876.1981
- Atherton, L. A., Dupret, D., and Mellor, J. R. (2015). Memory trace replay: the shaping of memory consolidation by neuromodulation. *Trends Neurosci* 38, 560–570. doi: 10.1016/j.tins.2015.07.004
- Averkin, R. G., Szemenyei, V., Bordé, S., and Tamás, G. (2016). Identified Cellular Correlates of Neocortical Ripple and High-Gamma Oscillations during Spindles of Natural Sleep. *Neuron* 92, 916–928. doi: 10.1016/j.neuron.2016.09.032
- Battaglia, F. P., Sutherland, G. R., and McNaughton, B. L. (2004). Hippocampal sharp wave bursts coincide with neocortical “up-state” transitions. *Learn. Mem.* 11, 697–704. doi: 10.1101/lm.73504
- Bazhenov, M., Timofeev, I., Steriade, M., and Sejnowski, T. J. (2002). Model of thalamocortical slow-wave sleep oscillations and transitions to activated States. *J. Neurosci.* 22, 8691–8704. doi: 10.1523/jneurosci.22-19-08691.2002
- Beed, P., Gundlfinger, A., Schneiderbauer, S., Song, J., Böhm, C., Burgalossi, A., et al. (2013). Inhibitory gradient along the dorsoventral axis in the medial entorhinal cortex. *Neuron* 79, 1197–1207. doi: 10.1016/j.neuron.2013.06.038
- Berger, H. (1929). Ueber das Elektroenkephalogramm des Menschen. *Arch. Psychiatr. Nervenkr.* 87, 527–570.
- Bittner, K. C., Grienberger, C., Vaidya, S. P., Milstein, A. D., Macklin, J. J., Suh, J., et al. (2015). Conjunctive input processing drives feature selectivity in hippocampal CA1 neurons. *Nat. Neurosci.* 18, 1133–1142. doi: 10.1038/nn.4062
- Bittner, K. C., Milstein, A. D., Grienberger, C., Romani, S., and Magee, J. C. (2017). Behavioral time scale synaptic plasticity underlies CA1 place fields. *Science* 357, 1033–1036. doi: 10.1126/science.aan3846
- Blumenfeld, H. (2002). The thalamus and seizures. *Arch. Neurol.* 59, 135–137.

- Blumenfeld, H., and Taylor, J. (2003). Why do seizures cause loss of consciousness? *Neuroscientist* 9, 301–310. doi: 10.1177/1073858403255624
- Buzsáki, G. (1989). Two-stage model of memory trace formation: a role for “noisy” brain states. *Neuroscience* 31, 551–570. doi: 10.1016/0306-4522(89)90423-5
- Buzsáki, G. (2015). Hippocampal sharp wave-ripple: a cognitive biomarker for episodic memory and planning. *Hippocampus* 25, 1073–1188. doi: 10.1002/hipo.22488
- Buzsáki, G., Leung, L. W., and Vanderwolf, C. H. (1983). Cellular bases of hippocampal EEG in the behaving rat. *Brain Res.* 287, 139–171. doi: 10.1016/0165-0173(83)90037-1
- Caputi, A., Rozov, A., Blatow, M., and Monyer, H. (2009). Two calretinin-positive GABAergic cell types in layer 2/3 of the mouse neocortex provide different forms of inhibition. *Cereb. Cortex* 19, 1345–1359. doi: 10.1093/cercor/bhn175
- Carter, M. E., Yizhar, O., Chikahisa, S., Nguyen, H., Adamantidis, A., Nishino, S., et al. (2010). Tuning arousal with optogenetic modulation of locus coeruleus neurons. *Nat. Neurosci.* 13, 1526–1533. doi: 10.1038/nn.2682
- Castejon, C., Barros-Zulaica, N., and Nuñez, A. (2016). Control of somatosensory cortical processing by thalamic posterior medial nucleus: a new role of thalamus in cortical function. *PLoS ONE* 11:e0148169. doi: 10.1371/journal.pone.0148169
- Castro-Alamancos, M. A. (2009). Cortical up and activated states: implications for sensory information processing. *Neuroscientist* 15, 625–634. doi: 10.1177/1073858409333074
- Cauli, B., Zhou, X., Tricoire, L., Toussay, X., and Staiger, J. F. (2014). Revisiting enigmatic cortical calretinin-expressing interneurons. *Front. Neuroanat.* 8:52. doi: 10.3389/fnana.2014.00052
- Cembrowski, M. S., Phillips, M. G., DiLisio, S. F., Shields, B. C., Winnubst, J., Chandrashekar, J., et al. (2018). Dissociable structural and functional hippocampal outputs via distinct subiculum cell classes. *Cell* 173, 1280–1292.e18. doi: 10.1016/j.cell.2018.03.031
- Cenquizca, L. A., and Swanson, L. W. (2006). An analysis of direct hippocampal cortical field CA1 axonal projections to diencephalon in the rat. *J. Comp. Neurol.* 497, 101–114. doi: 10.1002/cne.20985
- Cenquizca, L. A., and Swanson, L. W. (2007). Spatial organization of direct hippocampal field CA1 axonal projections to the rest of the cerebral cortex. *Brain Res. Rev.* 56, 1–26. doi: 10.1016/j.brainresrev.2007.05.002
- Chatelle, C., Chennu, S., Noirhomme, Q., Cruse, D., Owen, A. M., and Laureys, S. (2012). Brain-computer interfacing in disorders of consciousness. *Brain Inj.* 26, 1510–1522. doi: 10.3109/02699052.2012.698362
- Chen, Z., and Wilson, M. A. (2017). Deciphering neural codes of memory during sleep. *Trends Neurosci* 40, 260–275. doi: 10.1016/j.tins.2017.03.005
- Chrobak, J. J., and Buzsáki, G. (1994). Selective activation of deep layer (V-VI) retrohippocampal cortical neurons during hippocampal sharp waves in the behaving rat. *J. Neurosci.* 14, 6160–6170. doi: 10.1523/jneurosci.14-10-06160.1994
- Chrobak, J. J., and Buzsáki, G. (1996). High-frequency oscillations in the output networks of the hippocampal-entorhinal axis of the freely behaving rat. *J. Neurosci.* 16, 3056–3066. doi: 10.1523/jneurosci.16-09-03056.1996
- Compte, A., Sanchez-Vives, M. V., McCormick, D. A., and Wang, X.-J. (2003). Cellular and network mechanisms of slow oscillatory activity (<1 Hz) and wave propagations in a cortical network model. *J. Neurophysiol.* 89, 2707–2725. doi: 10.1152/jn.00845.2002
- Constantinople, C. M., and Bruno, R. M. (2011). Effects and mechanisms of wakefulness on local cortical networks. *Neuron* 69, 1061–1068. doi: 10.1016/j.neuron.2011.02.040
- Cowan, R. L., and Wilson, C. J. (1994). Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex. *J. Neurophysiol.* 71, 17–32. doi: 10.1152/jn.1994.71.1.17
- Cowsang, K. K., Shuman, T., Dillingham, B. C., Chang, A., Golshani, P., and Mayford, M. (2014). Direct reactivation of a coherent neocortical memory of context. *Neuron* 84, 432–441. doi: 10.1016/j.neuron.2014.09.022
- Craig, M. T., Mayne, E. W., Bettler, B., Paulsen, O., and McBain, C. J. (2013). Distinct roles of GABAB1a- and GABAB1b-containing GABAB receptors in spontaneous and evoked termination of persistent cortical activity. *J. Physiol. (Lond.)* 591, 835–843. doi: 10.1113/jphysiol.2012.248088
- Crochet, S., and Petersen, C. C. H. (2006). Correlating whisker behavior with membrane potential in barrel cortex of awake mice. *Nat. Neurosci.* 9, 608–610. doi: 10.1038/nn1690
- Crunelli, V., David, F., Lörincz, M. L., and Hughes, S. W. (2015). The thalamocortical network as a single slow wave-generating unit. *Curr. Opin. Neurobiol.* 31, 72–80. doi: 10.1016/j.conb.2014.09.001
- Csicsvari, J., Hirase, H., Mamiya, A., and Buzsáki, G. (2000). Ensemble patterns of hippocampal CA3-CA1 neurons during sharp wave-associated population events. *Neuron* 28, 585–594. doi: 10.1016/S0896-6273(00)00135-5
- Dahlstroem, A., and Fuxe, K. (1964). Evidence for the existence of monoamine-containing neurons in the central nervous system. I. Demonstration of monoamines in the cell bodies of brain stem neurons. *Acta Physiol. Scand. Suppl.* 232, 1–55.
- Dannenberg, H., Young, K., and Hasselmo, M. (2017). Modulation of hippocampal circuits by muscarinic and nicotinic receptors. *Front. Neural Circuits* 11:102. doi: 10.3389/fncir.2017.00102
- Desikan, S., Koser, D. E., Neitz, A., and Monyer, H. (2018). Target selectivity of septal cholinergic neurons in the medial and lateral entorhinal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 115, E2644–E2652.
- Destexhe, A., Hughes, S. W., Rudolph, M., and Crunelli, V. (2007). Are corticothalamic “up” states fragments of wakefulness? *Trends Neurosci.* 30, 334–342. doi: 10.1016/j.tins.2007.04.006
- Destexhe, A., and Paré, D. (1999). Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *J. Neurophysiol.* 81, 1531–1547. doi: 10.1152/jn.1999.81.4.1531
- Deuchars, J., and Thomson, A. M. (1996). CA1 pyramid-pyramid connections in rat hippocampus in vitro: dual intracellular recordings with biocytin filling. *Neuroscience* 74, 1009–1018. doi: 10.1016/S0306-4522(96)00251-5
- Dhillon, A., and Jones, R. S. (2000). Laminar differences in recurrent excitatory transmission in the rat entorhinal cortex in vitro. *Neuroscience* 99, 413–422. doi: 10.1016/S0306-4522(00)00225-6
- Diamond, M. E., von Heimendahl, M., Knutsen, P. M., Kleinfeld, D., and Ahissar, E. (2008). “Where” and “what” in the whisker sensorimotor system. *Nat. Rev. Neurosci.* 9, 601–612. doi: 10.1038/nrn2411
- Diba, K., and Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nat. Neurosci.* 10, 1241–1242. doi: 10.1038/nn1961
- Diekelmann, S., and Born, J. (2010). The memory function of sleep. *Nat. Rev. Neurosci.* 11, 114–126.
- Digby, R. J., Bravo, D. S., Paulsen, O., and Magloire, V. (2017). Distinct mechanisms of Up state maintenance in the medial entorhinal cortex and neocortex. *Neuropharmacology* 113, 543–555. doi: 10.1016/j.neuropharm.2016.11.009
- Ding, S.-L. (2013). Comparative anatomy of the prosubiculum, subiculum, presubiculum, postsubiculum, and parasubiculum in human, monkey, and rodent. *J. Comp. Neurol.* 521, 4145–4162. doi: 10.1002/cne.23416
- Dringenberg, H. C., and Vanderwolf, C. H. (1997). Neocortical activation: modulation by multiple pathways acting on central cholinergic and serotonergic systems. *Exp. Brain Res.* 116, 160–174. doi: 10.1007/PL00005736
- Dudai, Y., Karni, A., and Born, J. (2015). The consolidation and transformation of memory. *Neuron* 88, 20–32. doi: 10.1016/j.neuron.2015.09.004
- Duszkiewicz, A. J., McNamara, C. G., Takeuchi, T., and Genzel, L. (2019). Novelty and dopaminergic modulation of memory persistence: a tale of two systems. *Trends Neurosci.* 42, 102–114. doi: 10.1016/j.tins.2018.10.002
- Eckstein, F. P., Baughman, R. W., and Quinn, J. (1988). An anatomical study of cholinergic innervation in rat cerebral cortex. *Neuroscience* 25, 457–474. doi: 10.1016/0306-4522(88)90251-5
- Eggermann, E., Kremer, Y., Crochet, S., and Petersen, C. C. H. (2014). Cholinergic signals in mouse barrel cortex during active whisker sensing. *Cell Rep.* 9, 1654–1660. doi: 10.1016/j.celrep.2014.11.005
- Egorov, A. V., Hamam, B. N., Fransén, E., Hasselmo, M. E., and Alonso, A. A. (2002). Graded persistent activity in entorhinal cortex neurons. *Nature* 420, 173–178. doi: 10.1038/nature01171
- Ego-Stengel, V., and Wilson, M. A. (2010). Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus* 20, 1–10.
- Eschenko, O., Magri, C., Panzeri, S., and Sara, S. J. (2012). Noradrenergic neurons of the locus coeruleus are phase locked to cortical up-down states during sleep. *Cereb. Cortex* 22, 426–435. doi: 10.1093/cercor/bhr121
- Fanselow, E. E., and Connors, B. W. (2010). The roles of somatostatin-expressing (GIN) and fast-spiking inhibitory interneurons in UP-DOWN states of mouse neocortex. *J. Neurophysiol.* 104, 596–606. doi: 10.1152/jn.00206.2010

- Fanselow, E. E., Richardson, K. A., and Connors, B. W. (2008). Selective, state-dependent activation of somatostatin-expressing inhibitory interneurons in mouse neocortex. *J. Neurophysiol.* 100, 2640–2652. doi: 10.1152/jn.90691.2008
- Feldmeyer, D., Brecht, M., Helmchen, F., Petersen, C. C. H., Poulet, J. F. A., Staiger, J. F., et al. (2013). Barrel cortex function. *Prog. Neurobiol.* 103, 3–27.
- Feldmeyer, D., Qi, G., Emmenegger, V., and Staiger, J. F. (2018). Inhibitory interneurons and their circuit motifs in the many layers of the barrel cortex. *Neuroscience* 368, 132–151. doi: 10.1016/j.neuroscience.2017.05.027
- Fenno, L. E., Mattis, J., Ramakrishnan, C., Hyun, M., Lee, S. Y., He, M., et al. (2014). Targeting cells with single vectors using multiple-feature Boolean logic. *Nat. Methods* 11, 763–772. doi: 10.1038/nmeth.2996
- Ferezou, I., Haiss, F., Gentet, L. J., Aronoff, R., Weber, B., and Petersen, C. C. H. (2007). Spatiotemporal dynamics of cortical sensorimotor integration in behaving mice. *Neuron* 56, 907–923. doi: 10.1016/j.neuron.2007.10.007
- Foster, D. J. (2017). Replay comes of age. *Annu. Rev. Neurosci.* 40, 581–602. doi: 10.1146/annurev-neuro-072116-031538
- Foster, D. J., and Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* 440, 680–683. doi: 10.1038/nature04587
- Fuentealba, P., Begum, R., Capogna, M., Jinno, S., Márton, L. F., Csicsvari, J., et al. (2008). Ivy cells: a population of nitric-oxide-producing, slow-spiking GABAergic neurons and their involvement in hippocampal network activity. *Neuron* 57, 917–929. doi: 10.1016/j.neuron.2008.01.034
- Funk, C. M., Peelman, K., Bellesi, M., Marshall, W., Cirelli, C., and Tononi, G. (2017). Role of somatostatin-positive cortical interneurons in the generation of sleep slow waves. *J. Neurosci.* 37, 9132–9148. doi: 10.1523/jneurosci.1303-17.2017
- Gelinas, J. N., Khodagholy, D., Thesen, T., Devinsky, O., and Buzsáki, G. (2016). Interictal epileptiform discharges induce hippocampal-cortical coupling in temporal lobe epilepsy. *Nat. Med.* 22, 641–648. doi: 10.1038/nm.4084
- Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F., and Petersen, C. C. H. (2010). Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. *Neuron* 65, 422–435. doi: 10.1016/j.neuron.2010.01.006
- Gentet, L. J., Kremer, Y., Taniguchi, H., Huang, Z. J., Staiger, J. F., and Petersen, C. C. H. (2012). Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex. *Nat. Neurosci.* 15, 607–612. doi: 10.1038/nn.3051
- Genzel, L., Kroes, M. C. W., Dresler, M., and Battaglia, F. P. (2014). Light sleep versus slow wave sleep in memory consolidation: a question of global versus local processes? *Trends Neurosci.* 37, 10–19. doi: 10.1016/j.tins.2013.10.002
- Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G., and Zugaro, M. B. (2009). Selective suppression of hippocampal ripples impairs spatial memory. *Nat. Neurosci.* 12, 1222–1223. doi: 10.1038/nn.2384
- Gloveli, T., Schmitz, D., Empson, R. M., and Heinemann, U. (1997). Frequency-dependent information flow from the entorhinal cortex to the hippocampus. *J. Neurophysiol.* 78, 3444–3449. doi: 10.1152/jn.1997.78.6.3444
- Gonzalez-Sulser, A., Parthier, D., Candela, A., McClure, C., Pastoll, H., Garden, D., et al. (2014). GABAergic projections from the medial septum selectively inhibit interneurons in the medial entorhinal cortex. *J. Neurosci.* 34, 16739–16743. doi: 10.1523/jneurosci.1612-14.2014
- Groen, T. V., and Wyss, J. M. (1990). Extrinsic projections from area CA1 of the rat hippocampus: olfactory, cortical, subcortical, and bilateral hippocampal formation projections. *J. Compar. Neurol.* 302, 515–528. doi: 10.1002/cne.903020308
- Hahn, T. T., Sakmann, B., and Mehta, M. R. (2006). Phase-locking of hippocampal interneurons' membrane potential to neocortical up-down states. *Nat. Neurosci.* 9, 1359–1361. doi: 10.1038/nn1788
- Hahn, T. T., Sakmann, B., and Mehta, M. R. (2007). Differential responses of hippocampal subfields to cortical up-down states. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5169–5174. doi: 10.1073/pnas.0700222104
- Hahn, T. T. G., McFarland, J. M., Berberich, S., Sakmann, B., and Mehta, M. R. (2012). Spontaneous persistent activity in entorhinal cortex modulates cortico-hippocampal interaction in vivo. *Nat. Neurosci.* 15, 1531–1538. doi: 10.1038/nn.3236
- Haider, B., Duque, A., Hasenstaub, A. R., and McCormick, D. A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.* 26, 4535–4545. doi: 10.1523/jneurosci.5297-05.2006
- Haider, B., Häusser, M., and Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. *Nature* 493, 97–100. doi: 10.1038/nature11665
- Han, Y., Shi, Y., Xi, W., Zhou, R., Tan, Z., Wang, H., et al. (2014). Selective activation of cholinergic basal forebrain neurons induces immediate sleep-wake transitions. *Curr. Biol.* 24, 693–698. doi: 10.1016/j.cub.2014.02.011
- Harris, K. D., and Thiele, A. (2011). Cortical state and attention. *Nat. Rev. Neurosci.* 12, 509–523.
- Hasenstaub, A., Sachdev, R. N. S., and McCormick, D. A. (2007). State changes rapidly modulate cortical neuronal responsiveness. *J. Neurosci.* 27, 9607–9622. doi: 10.1523/jneurosci.2184-07.2007
- Hasenstaub, A., Shu, Y., Haider, B., Kraushaar, U., Duque, A., and McCormick, D. A. (2005). Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks. *Neuron* 47, 423–435. doi: 10.1016/j.neuron.2005.06.016
- Hassani, O. K., Lee, M. G., Henny, P., and Jones, B. E. (2009). Discharge profiles of identified GABAergic in comparison to cholinergic and putative glutamatergic basal forebrain neurons across the sleep-wake cycle. *J. Neurosci.* 29, 11828–11840. doi: 10.1523/jneurosci.1259-09.2009
- Headley, D. B., Kanta, V., and Paré, D. (2017). Intra- and interregional cortical interactions related to sharp-wave ripples and dentate spikes. *J. Neurophysiol.* 117, 556–565. doi: 10.1152/jn.00644.2016
- Headley, D. B., and Paré, D. (2017). Common oscillatory mechanisms across multiple memory systems. *NPJ Sci. Learn.* 2, 1–8. doi: 10.1038/s41539-016-0001-2
- Heys, J. G., Giocomo, L. M., and Hasselmo, M. E. (2010). Cholinergic modulation of the resonance properties of stellate cells in layer II of medial entorhinal cortex. *J. Neurophysiol.* 104, 258–270. doi: 10.1152/jn.00492.2009
- Heys, J. G., Schultheiss, N. W., Shay, C. F., Tsuno, Y., and Hasselmo, M. E. (2012). Effects of acetylcholine on neuronal properties in entorhinal cortex. *Front. Behav. Neurosci.* 6:32. doi: 10.3389/fnbeh.2012.00032
- Houser, C. R., Crawford, G. D., Salvaterra, P. M., and Vaughn, J. E. (1985). Immunocytochemical localization of choline acetyltransferase in rat cerebral cortex: a study of cholinergic neurons and synapses. *J. Comp. Neurol.* 234, 17–34. doi: 10.1002/cne.902340103
- Isomura, Y., Sirota, A., Özen, S., Montgomery, S., Mizuseki, K., Henze, D. A., et al. (2006). Integration and segregation of activity in entorhinal-hippocampal subregions by neocortical slow oscillations. *Neuron* 52, 871–882. doi: 10.1016/j.neuron.2006.10.023
- Jadhav, S. P., Kemere, C., German, P. W., and Frank, L. M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. *Science* 336, 1454–1458. doi: 10.1126/science.1217230
- Jay, T. M., and Witter, M. P. (1991). Distribution of hippocampal CA1 and subicular efferents in the prefrontal cortex of the rat studied by means of anterograde transport of Phaseolus vulgaris-leucoagglutinin. *J. Comp. Neurol.* 313, 574–586. doi: 10.1002/cne.903130404
- Jercog, D., Roxin, A., Barthó, P., Luczak, A., Compte, A., and de la Rocha, J. (2017). UP-DOWN cortical dynamics reflect state transitions in a bistable network. *Elife* 6, 1–33. doi: 10.7554/eLife.22425
- Ji, D., and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107. doi: 10.1038/nn1825
- Jochims, A., Reboreda, A., Hasselmo, M. E., and Yoshida, M. (2013). Cholinergic receptor activation supports persistent firing in layer III neurons in the medial entorhinal cortex. *Behav. Brain Res.* 254, 108–115. doi: 10.1016/j.bbr.2013.06.027
- Kaifosh, P., Lovett-Barron, M., Turi, G. F., Reardon, T. R., and Losonczy, A. (2013). Septo-hippocampal GABAergic signaling across multiple modalities in awake mice. *Nat. Neurosci.* 16, 1182–1184. doi: 10.1038/nn.3482
- Katona, L., Lapray, D., Viney, T. J., Oulhaj, A., Borhegyi, Z., Micklem, B. R., et al. (2014). Sleep and movement differentiates actions of two types of somatostatin-expressing GABAergic interneuron in rat hippocampus. *Neuron* 82, 872–886. doi: 10.1016/j.neuron.2014.04.007
- Kawaguchi, Y., and Kubota, Y. (1997). GABAergic cell subtypes and their synaptic connections in rat frontal cortex. *Cereb. Cortex* 7, 476–486. doi: 10.1093/cercor/7.6.476



- Kay, K., and Frank, L. M. (2019). Three brain states in the hippocampus and cortex. *Hippocampus* 29, 184–238. doi: 10.1002/hipo.22956
- Khodagholy, D., Gelinas, J. N., and Buzsáki, G. (2017). Learning-enhanced coupling between ripple oscillations in association cortices and hippocampus. *Science* 358, 369–372. doi: 10.1126/science.aan6203
- Kim, J., Gulati, T., and Ganguly, K. (2019). Competing roles of slow oscillations and delta waves in memory consolidation versus forgetting. *Cell* 179, 514–526.e13. doi: 10.1016/j.cell.2019.08.040
- Kim, S. M., Ganguli, S., and Frank, L. M. (2012). Spatial information outflow from the hippocampal circuit: distributed spatial coding and phase precession in the subiculum. *J. Neurosci.* 32, 11539–11558. doi: 10.1523/jneurosci.5942-11.2012
- Kim, Y., and Spruston, N. (2012). Target-specific output patterns are predicted by the distribution of regular-spiking and bursting pyramidal neurons in the subiculum. *Hippocampus* 22, 693–706. doi: 10.1002/hipo.20931
- Kitamura, T., Ogawa, S. K., Roy, D. S., Okuyama, T., Morrissey, M. D., Smith, L. M., et al. (2017). Engrams and circuits crucial for systems consolidation of a memory. *Science* 356, 73–78. doi: 10.1126/science.aam6808
- Kitamura, T., Pignatelli, M., Suh, J., Kohara, K., Yoshiki, A., Abe, K., et al. (2014). Island cells control temporal association memory. *Science* 343, 896–901. doi: 10.1126/science.1244634
- Klausberger, T., Marton, L. F., O'Neill, J., Huck, J. H. J., Dalezios, Y., Fuentealba, P., et al. (2005). Complementary roles of cholecystokinin- and parvalbumin-expressing GABAergic neurons in hippocampal network oscillations. *J. Neurosci.* 25, 9782–9793. doi: 10.1523/jneurosci.3269-05.2005
- Klausberger, T., and Somogyi, P. (2008). Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science* 321, 53–57. doi: 10.1126/science.1149381
- Kleinfeld, D., Ahissar, E., and Diamond, M. E. (2006). Active sensation: insights from the rodent vibrissa sensorimotor system. *Curr. Opin. Neurobiol.* 16, 435–444. doi: 10.1016/j.conb.2006.06.009
- Klinzing, J. G., Niethard, N., and Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nat. Neurosci.* 22, 1598–1610. doi: 10.1038/s41593-019-0467-3
- Knauer, B., Jochems, A., Valero-Aracama, M. J., and Yoshida, M. (2013). Long-lasting intrinsic persistent firing in rat CA1 pyramidal cells: a possible mechanism for active maintenance of memory. *Hippocampus* 23, 820–831. doi: 10.1002/hipo.22136
- Knoblich, U., Huang, L., Zeng, H., and Li, L. (2019). Neuronal cell-subtype specificity of neural synchronization in mouse primary visual cortex. *Nat. Commun.* 10:2533.
- Knowles, W. D., and Schwartzkroin, P. A. (1981). Local circuit synaptic interactions in hippocampal brain slices. *J. Neurosci.* 1, 318–322. doi: 10.1523/jneurosci.01-03-00318.1981
- Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts and Company Publishers.
- Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321.
- Köhler, C., Chan-Palay, V., and Wu, J. Y. (1984). Septal neurons containing glutamic acid decarboxylase immunoreactivity project to the hippocampal region in the rat brain. *Anat. Embryol.* 169, 41–44. doi: 10.1007/bf00300585
- Krishnan, G. P., Chauvette, S., Shamie, I., Soltani, S., Timofeev, I., Cash, S. S., et al. (2016). Cellular and neurochemical basis of sleep stages in the thalamocortical network. *Elife* 5:e18607.
- Kristt, D. A. (1979). Development of neocortical circuitry: histochemical localization of acetylcholinesterase in relation to the cell layers of rat somatosensory cortex. *J. Comp. Neurol.* 186, 1–15. doi: 10.1002/cne.901860102
- Krueger, J. M., Huang, Y. H., Rector, D. M., and Buysse, D. J. (2013). Sleep: a synchrony of cell activity-driven small network states. *Eur. J. Neurosci.* 38, 2199–2209. doi: 10.1111/ejn.12238
- Kubota, Y. (2014). Untangling GABAergic wiring in the cortical microcircuit. *Curr. Opin. Neurobiol.* 26, 7–14. doi: 10.1016/j.conb.2013.10.003
- Lapray, D., Lasztóczy, B., Lagler, M., Viney, T. J., Katona, L., Valenti, O., et al. (2012). Behavior-dependent specialization of identified hippocampal interneurons. *Nat. Neurosci.* 15, 1265–1271. doi: 10.1038/nn.3176
- Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* 36, 141–151. doi: 10.1016/j.tins.2012.11.006
- Lasztóczy, B., Tukker, J. J., Somogyi, P., and Klausberger, T. (2011). Terminal field and firing selectivity of cholecystokinin-expressing interneurons in the hippocampal CA3 area. *J. Neurosci.* 31, 18073–18093. doi: 10.1523/jneurosci.3573-11.2011
- Lee, A. K., and Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* 36, 1183–1194. doi: 10.1016/s0896-6273(02)01096-6
- Lee, M. G., Manns, I. D., Alonso, A., and Jones, B. E. (2004). Sleep-wake related discharge properties of basal forebrain neurons recorded with micropipettes in head-fixed rats. *J. Neurophysiol.* 92, 1182–1198. doi: 10.1152/jn.01003.2003
- Lestienne, R., Hervé-Minvielle, A., Robinson, D., Briois, L., and Sara, S. J. (1997). Slow oscillations as a probe of the dynamics of the locus coeruleus-frontal cortex interaction in anesthetized rats. *J. Physiol. Paris* 91, 273–284. doi: 10.1016/s0928-4257(97)82407-2
- Levenstein, D., Buzsáki, G., and Rinzel, J. (2019). NREM sleep in the rodent neocortex and hippocampus reflects excitable dynamics. *Nat. Commun.* 10:2478.
- Lewis, P. R., and Shute, C. C. (1967). The cholinergic limbic system: projections to hippocampal formation, medial cortex, nuclei of the ascending cholinergic reticular system, and the subfornical organ and supra-optic crest. *Brain* 90, 521–540. doi: 10.1093/brain/90.3.521
- Logothetis, N. K., Eschenko, O., Murayama, Y., Augath, M., Steudel, T., Evrard, H. C., et al. (2012). Hippocampal-cortical interaction during periods of subcortical silence. *Nature* 491, 547–553. doi: 10.1038/nature11618
- Lysakowski, A., Wainer, B. H., Bruce, G., and Hersch, L. B. (1989). An atlas of the regional and laminar distribution of choline acetyltransferase immunoreactivity in rat cerebral cortex. *Neuroscience* 28, 291–336. doi: 10.1016/0306-4522(89)90180-2
- Ma, Y., Hu, H., Berrebi, A. S., Mathers, P. H., and Agmon, A. (2006). Distinct subtypes of somatostatin-containing neocortical interneurons revealed in transgenic mice. *J. Neurosci.* 26, 5069–5082. doi: 10.1523/jneurosci.0661-06.2006
- Maboudi, K., Ackermann, E., de Jong, L. W., Pfeiffer, B. E., Foster, D., Diba, K., et al. (2018). Uncovering temporal structure in hippocampal output patterns. *Elife* 7, 1–24. doi: 10.7554/eLife.34467
- Maingret, N., Girardeau, G., Todorova, R., Goutierre, M., and Zugaro, M. (2016). Hippocampo-cortical coupling mediates memory consolidation during sleep. *Nat. Neurosci.* 19, 959–964. doi: 10.1038/nn.4304
- Mander, B. A., Marks, S. M., Vogel, J. W., Rao, V., Lu, B., Saletin, J. M., et al. (2015).  $\beta$ -amyloid disrupts human NREM slow waves and related hippocampus-dependent memory consolidation. *Nat. Neurosci.* 18, 1051–1057. doi: 10.1038/nn.4035
- Mann, E. O., Kohl, M. M., and Paulsen, O. (2009). Distinct roles of GABA(A) and GABA(B) receptors in balancing and terminating persistent cortical activity. *J. Neurosci.* 29, 7513–7518. doi: 10.1523/jneurosci.6162-08.2009
- Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.* 5, 793–807. doi: 10.1038/nrn1519
- Marr, D. (1971). Simple memory: a theory for archicortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 262, 23–81. doi: 10.1098/rstb.1971.0078
- Mattia, M., and Sanchez-Vives, M. V. (2012). Exploring the spectrum of dynamical regimes and timescales in spontaneous cortical activity. *Cogn. Neurodyn.* 6, 239–250. doi: 10.1007/s11571-011-9179-4
- Maviel, T., Durkin, T. P., Menzaghi, F., and Bontempi, B. (2004). Sites of neocortical reorganization critical for remote spatial memory. *Science* 305, 96–99. doi: 10.1126/science.1098180
- McCarley, R. W., and Hobson, J. A. (1975). Discharge patterns of cat pontine brain stem neurons during desynchronized sleep. *J. Neurophysiol.* 38, 751–766. doi: 10.1152/jn.1975.38.4.751
- McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C. R., et al. (2015). Waking state: rapid variations modulate neural and behavioral responses. *Neuron* 87, 1143–1161. doi: 10.1016/j.neuron.2015.09.012
- McKinney, M., Coyle, J. T., and Hedreen, J. C. (1983). Topographic analysis of the innervation of the rat neocortex and hippocampus by the basal forebrain cholinergic system. *J. Comp. Neurol.* 217, 103–121. doi: 10.1002/cne.902170109
- McNamara, C. G., and Dupret, D. (2017). Two sources of dopamine for the hippocampus. *Trends Neurosci.* 40, 383–384. doi: 10.1016/j.tins.2017.05.005



- Mease, R. A., Sumser, A., Sakmann, B., and Groh, A. (2016). Corticothalamic spike transfer via the L5B-POm Pathway in vivo. *Cereb. Cortex* 26, 3461–3475. doi: 10.1093/cercor/bhw123
- Mena-Segovia, J., and Bolam, J. P. (2011). Phasic modulation of cortical high-frequency oscillations by pedunculo-pontine neurons. *Prog. Brain Res.* 193, 85–92. doi: 10.1016/b978-0-444-53839-0.00006-5
- Mena-Segovia, J., Sims, H. M., Magill, P. J., and Bolam, J. P. (2008). Cholinergic brainstem neurons modulate cortical gamma activity during slow oscillations. *J. Physiol. (Lond.)* 586, 2947–2960. doi: 10.1113/jphysiol.2008.153874
- Mesulam, M. M., Mufson, E. J., Wainer, B. H., and Levey, A. I. (1983). Central cholinergic pathways in the rat: an overview based on an alternative nomenclature (Ch1-Ch6). *Neuroscience* 10, 1185–1201. doi: 10.1016/0306-4522(83)90108-2
- Metherate, R., and Ashe, J. H. (1993). Ionic flux contributions to neocortical slow waves and nucleus basalis-mediated activation: whole-cell recordings in vivo. *J. Neurosci.* 13, 5312–5323. doi: 10.1523/jneurosci.13-12-05312.1993
- Metherate, R., Cox, C. L., and Ashe, J. H. (1992). Cellular bases of neocortical activation: modulation of neural oscillations by the nucleus basalis and endogenous acetylcholine. *J. Neurosci.* 12, 4701–4711. doi: 10.1523/jneurosci.12-12-04701.1992
- Milczarek, M. M., Vann, S. D., and Sengpiel, F. (2018). Spatial memory engram in the mouse retrosplenial cortex. *Curr. Biol.* 28, 1975–1980.e6. doi: 10.1016/j.cub.2018.05.002
- Mizuseki, K., Diba, K., Pastalkova, E., and Buzsáki, G. (2011). Hippocampal CA1 pyramidal cells form functionally distinct sublayers. *Nat. Neurosci.* 14, 1174–1181. doi: 10.1038/nn.2894
- Mölle, M., Yeshenko, O., Marshall, L., Sara, S. J., and Born, J. (2006). Hippocampal sharp wave-ripples linked to slow oscillations in rat slow-wave sleep. *J. Neurophysiol.* 96, 62–70. doi: 10.1152/jn.00014.2006
- Motelow, J. E., Li, W., Zhan, Q., Mishra, A. M., Sachdev, R. N. S., Liu, G., et al. (2015). Decreased subcortical cholinergic arousal in focal seizures. *Neuron* 85, 561–572. doi: 10.1016/j.neuron.2014.12.058
- Müller, C., and Remy, S. (2018). Septo-hippocampal interaction. *Cell Tissue Res.* 373, 565–575. doi: 10.1007/s00441-017-2745-2
- Namiki, S., Norimoto, H., Kobayashi, C., Nakatani, K., Matsuki, N., and Ikegaya, Y. (2013). Layer III neurons control synchronized waves in the immature cerebral cortex. *J. Neurosci.* 33, 987–1001. doi: 10.1523/jneurosci.2522-12.2013
- Neske, G. T., Patrick, S. L., and Connors, B. W. (2015). Contributions of diverse excitatory and inhibitory neurons to recurrent network activity in cerebral cortex. *J. Neurosci.* 35, 1089–1105. doi: 10.1523/jneurosci.2279-14.2015
- Nghiem, T.-A. E., Tort-Colet, N., Górski, T., Ferrari, U., Moghimi-firoozabad, S., Goldman, J. S., et al. (2018). Cholinergic switch between two different types of slow waves in cerebral cortex. *bioRxiv* [Preprint]. doi: 10.1093/cercor/bhz320
- Niethard, N., Burgalossi, A., and Born, J. (2017). Plasticity during sleep is linked to specific regulation of cortical circuit activity. *Front. Neural Circuits* 11:65. doi: 10.3389/fncir.2017.00065
- Niethard, N., Hasegawa, M., Itokazu, T., Oyanedel, C. N., Born, J., and Sato, T. R. (2016). Sleep-stage-specific regulation of cortical excitation and inhibition. *Curr. Biol.* 26, 2739–2749. doi: 10.1016/j.cub.2016.08.035
- Nitzan, N., McKenzie, S., Beed, P., English, D. F., Oldani, S., Tukker, J. J., et al. (2020). Propagation of hippocampal ripples to the neocortex by way of a subiculum-retrosplenial pathway. *bioRxiv* [Preprint]. doi: 10.1101/2020.02.27.966770
- O'Keefe, J., and Nadel, L. (1978). *The Hippocampus As a Cognitive Map*. Oxford: Oxford University Press.
- Ólafsdóttir, H. F., Carpenter, F., and Barry, C. (2016). Coordinated grid and place cell replay during rest. *Nat. Neurosci.* 19, 792–794. doi: 10.1038/nn.4291
- O'Neill, J., Boccarda, C. N., Stella, F., Schoenenberger, P., and Csicsvari, J. (2017). Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. *Science* 355, 184–188. doi: 10.1126/science.aag2787
- O'Neill, J., Senior, T., and Csicsvari, J. (2006). Place-selective firing of CA1 pyramidal cells during sharp wave/ripple network patterns in exploratory behavior. *Neuron* 49, 143–155. doi: 10.1016/j.neuron.2005.10.037
- Pala, A., and Petersen, C. C. (2018). State-dependent cell-type-specific membrane potential dynamics and unitary synaptic inputs in awake mice. *Elife* 7, 1–13. doi: 10.7554/eLife.35869
- Palmer, L. M., Schulz, J. M., Murphy, S. C., Ledergerber, D., Murayama, M., and Larkum, M. E. (2012). The cellular basis of GABA(B)-mediated interhemispheric inhibition. *Science* 335, 989–993. doi: 10.1126/science.1217276
- Palmer, L. M., Shai, A. S., Reeve, J. E., Anderson, H. L., Paulsen, O., and Larkum, M. E. (2014). NMDA spikes enhance action potential generation during sensory input. *Nat. Neurosci.* 17, 383–390. doi: 10.1038/nn.3646
- Paré, D., Shink, E., Gaudreau, H., Destexhe, A., and Lang, E. J. (1998). Impact of spontaneous synaptic activity on the resting properties of cat neocortical pyramidal neurons In vivo. *J. Neurophysiol.* 79, 1450–1460. doi: 10.1152/jn.1998.79.3.1450
- Petersen, C. C., Hahn, T. T., Mehta, M., Grinvald, A., and Sakmann, B. (2003). Interaction of sensory responses with spontaneous depolarization in layer? barrel cortex. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13638–13643. doi: 10.1073/pnas.2235811100
- Petersen, C. C. H. (2007). The functional organization of the barrel cortex. *Neuron* 56, 339–355. doi: 10.1016/j.neuron.2007.09.017
- Petronantonakis, P. C., and Poirazi, P. (2014). A compressed sensing perspective of hippocampal function. *Front. Syst. Neurosci.* 8:141. doi: 10.3389/fnsys.2014.00141
- Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., and Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat. Neurosci.* 12, 919–926. doi: 10.1038/nn.2337
- Phillips, K. G., Bartsch, U., McCarthy, A. P., Edgar, D. M., Tricklebank, M. D., Wafford, K. A., et al. (2012). Decoupling of sleep-dependent cortical and hippocampal interactions in a neurodevelopmental model of schizophrenia. *Neuron* 76, 526–533. doi: 10.1016/j.neuron.2012.09.016
- Ponomarenko, A. A., Knoche, A., Korotkova, T. M., and Haas, H. L. (2003). Aminergic control of high-frequency (approximately 200 Hz) network oscillations in the hippocampus of the behaving rat. *Neurosci. Lett.* 348, 101–104. doi: 10.1016/s0304-3940(03)00742-0
- Poulet, J. F. A., and Crochet, S. (2018). The cortical states of wakefulness. *Front. Syst. Neurosci.* 12:64. doi: 10.3389/fnsys.2018.00064
- Poulet, J. F. A., Fernandez, L. M. J., Crochet, S., and Petersen, C. C. H. (2012). Thalamic control of cortical states. *Nat. Neurosci.* 15, 370–372. doi: 10.1038/nn.3035
- Poulet, J. F. A., and Petersen, C. C. H. (2008). Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* 454, 881–885. doi: 10.1038/nature07150
- Price, C. J., Cauli, B., Kovacs, E. R., Kulik, A., Lambolez, B., Shigemoto, R., et al. (2005). Neurogliaform neurons form a novel inhibitory network in the hippocampal CA1 area. *J. Neurosci.* 25, 6775–6786. doi: 10.1523/jneurosci.1135-05.2005
- Prönneke, A., Scheuer, B., Wagener, R. J., Möck, M., Witte, M., and Staiger, J. F. (2015). Characterizing VIP neurons in the barrel cortex of VIPcre/tdTomato mice reveals layer-specific differences. *Cereb. Cortex* 25, 4854–4868. doi: 10.1093/cercor/bhv202
- Qin, H., Lu, J., Jin, W., Chen, X., and Fu, L. (2019). Multichannel fiber photometry for mapping axonal terminal activity in a restricted brain region in freely moving mice. *Neurophotonics* 6:35011.
- Rasch, B., and Born, J. (2007). Maintaining memories by reactivation. *Curr. Opin. Neurobiol.* 17, 698–703. doi: 10.1016/j.conb.2007.11.007
- Ros, H., Magill, P. J., Moss, J., Bolam, J. P., and Mena-Segovia, J. (2010). Distinct types of non-cholinergic pedunculo-pontine neurons are differentially modulated during global brain states. *Neuroscience* 170, 78–91. doi: 10.1016/j.neuroscience.2010.06.068
- Rothschild, G., Eban, E., and Frank, L. M. (2017). A cortical-hippocampal-cortical loop of information processing during memory consolidation. *Nat. Neurosci.* 20, 251–259. doi: 10.1038/nn.4457
- Roux, L., Eichler, R., Stark, E., and Buzsáki, G. (2017). Sharp wave ripples during learning stabilize hippocampal spatial map. *Nat. Neurosci.* 20, 845–853. doi: 10.1038/nn.4543
- Sachdev, R. N., Lu, S. M., Wiley, R. G., and Ebner, F. F. (1998). Role of the basal forebrain cholinergic projection in somatosensory cortical plasticity. *J. Neurophysiol.* 79, 3216–3228. doi: 10.1152/jn.1998.79.6.3216
- Sachdev, R. N. S., Ebner, F. F., and Wilson, C. J. (2004). Effect of subthreshold up and down states on the whisker-evoked response in somatosensory cortex. *J. Neurophysiol.* 92, 3511–3521. doi: 10.1152/jn.00347.2004

- Sachdev, R. N. S., Gaspard, N., Gerrard, J. L., Hirsch, L. J., Spencer, D. D., and Zaveri, H. P. (2015). Delta rhythm in wakefulness: evidence from intracranial recordings in human beings. *J. Neurophysiol.* 114, 1248–1254. doi: 10.1152/jn.00249.2015
- Sachdev, R. N. S., Krause, M. R., and Mazer, J. A. (2012). Surround suppression and sparse coding in visual and barrel cortices. *Front. Neural Circuits* 6:43. doi: 10.3389/fncir.2012.00043
- Salkoff, D. B., Zagha, E., Yüzgeç, Ö., and McCormick, D. A. (2015). Synaptic mechanisms of tight spike synchrony at gamma frequency in cerebral cortex. *J. Neurosci.* 35, 10236–10251. doi: 10.1523/jneurosci.0828-15.2015
- Sanchez-Vives, M. V., and McCormick, D. A. (2000). Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nat. Neurosci.* 3, 1027–1034. doi: 10.1038/79848
- Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nat. Rev. Neurosci.* 10, 211–223. doi: 10.1038/nrn2573
- Sara, S. J. (2017). Sleep to remember. *J. Neurosci.* 37, 457–463.
- Scammell, T. E., Arrigoni, E., and Lipton, J. O. (2017). Neural circuitry of wakefulness and sleep. *Neuron* 93, 747–765. doi: 10.1016/j.neuron.2017.01.014
- Schmidt-Hieber, C., Toleikyte, G., Aitchison, L., Roth, A., Clark, B. A., Branco, T., et al. (2017). Active dendritic integration as a mechanism for robust and precise grid cell firing. *Nat. Neurosci.* 20, 1114–1121. doi: 10.1038/nn.4582
- Senzai, Y., Fernandez-Ruiz, A., and Buzsáki, G. (2019). Layer-specific physiological features and interlaminar interactions in the primary visual cortex of the mouse. *Neuron* 101, 500–513.e5. doi: 10.1016/j.neuron.2018.12.009
- Sheffield, M. E. J., Adoff, M. D., and Dombeck, D. A. (2017). Increased prevalence of calcium transients across the dendritic arbor during place field formation. *Neuron* 96, 490–504.e5. doi: 10.1016/j.neuron.2017.09.029
- Shu, Y., Hasenstaub, A., and McCormick, D. A. (2003). Turning on and off recurrent balanced cortical activity. *Nature* 423, 288–293. doi: 10.1038/nature01616
- Siapas, A. G., and Wilson, M. A. (1998). Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron* 21, 1123–1128. doi: 10.1016/s0896-6273(00)80629-7
- Siclari, F., and Tononi, G. (2017). Local aspects of sleep and wakefulness. *Curr. Opin. Neurobiol.* 44, 222–227. doi: 10.1016/j.conb.2017.05.008
- Simpson, K. L., Waterhouse, B. D., and Lin, R. C. S. (2006). Characterization of neurochemically specific projections from the locus coeruleus with respect to somatosensory-related barrels. *Anat. Rec. A Discov. Mol. Cell Evol. Biol.* 288, 166–173. doi: 10.1002/ar.a.20287
- Sirota, A., and Buzsáki, G. (2005). Interaction between neocortical and hippocampal networks via slow oscillations. *Thalamus Relat. Syst.* 3, 245–259.
- Sirota, A., Csicsvari, J., Buhl, D., and Buzsáki, G. (2003). Communication between neocortex and hippocampus during sleep in rodents. *Proc. Natl. Acad. Sci. U.S.A.* 100, 2065–2069. doi: 10.1073/pnas.0437938100
- Squire, L. R., and Zola-Morgan, M. (1991). The memory system. *Cold Spring. Harb. Perspect. Biol.* 7:a021667.
- Steriade, M. (2001). Impact of network activities on neuronal properties in corticothalamic systems. *J. Neurophysiol.* 86, 1–39. doi: 10.1152/jn.2001.86.1.1
- Steriade, M., Amzica, F., and Nuñez, A. (1993a). Cholinergic and noradrenergic modulation of the slow (approximately 0.3 Hz) oscillation in neocortical cells. *J. Neurophysiol.* 70, 1385–1400. doi: 10.1152/jn.1993.70.4.1385
- Steriade, M., McCormick, D. A., and Sejnowski, T. J. (1993b). Thalamocortical oscillations in the sleeping and aroused brain. *Science* 262, 679–685. doi: 10.1126/science.8235588
- Steriade, M., Nuñez, A., and Amzica, F. (1993c). A novel slow (< 1 Hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *J. Neurosci.* 13, 3252–3265. doi: 10.1523/jneurosci.13-08-03252.1993
- Steriade, M., Nuñez, A., and Amzica, F. (1993d). Intracellular analysis of relations between the slow (< 1 Hz) neocortical oscillation and other sleep rhythms of the electroencephalogram. *J. Neurosci.* 13, 3266–3283. doi: 10.1523/jneurosci.13-08-03266.1993
- Steriade, M., Contreras, D., Amzica, F., and Timofeev, I. (1996). Synchronization of fast (30–40 Hz) spontaneous oscillations in intrathalamic and thalamocortical networks. *J. Neurosci.* 16, 2788–2808. doi: 10.1523/jneurosci.16-08-02788.1996
- Steriade, M., and McCarley, R. W. (2005). *Brain Control of Wakefulness and Sleep*. New York, NY: Kluwer Academic.
- Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature* 437, 1272–1278. doi: 10.1038/nature04286
- Stickgold, R., and Walker, M. P. (2005). Memory consolidation and reconsolidation: what is the role of sleep? *Trends Neurosci.* 28, 408–415. doi: 10.1016/j.tins.2005.06.004
- Suh, J., Rivest, A. J., Nakashiba, T., Tominaga, T., and Tonegawa, S. (2011). Entorhinal cortex layer III input to the hippocampus is crucial for temporal association memory. *Science* 334, 1415–1420. doi: 10.1126/science.1210125
- Sullivan, D., Csicsvari, J., Mizuseki, K., Montgomery, S., Diba, K., and Buzsáki, G. (2011). Relationships between hippocampal sharp waves, ripples, and fast gamma oscillation: influence of dentate and entorhinal cortical activity. *J. Neurosci.* 31, 8605–8616. doi: 10.1523/JNEUROSCI.0294-11.2011
- Suzuki, M., and Larkum, M. E. (2020). General Anesthesia Decouples Cortical Pyramidal Neurons. *Cell* 180, 666–676.e3. doi: 10.1016/j.cell.2020.01.024
- Swanson, L. W., and Hartman, B. K. (1975). The central adrenergic system. An immunofluorescence study of the location of cell bodies and their efferent connections in the rat utilizing dopamine-beta-hydroxylase as a marker. *J. Comp. Neurol.* 163, 467–505. doi: 10.1002/cne.901630406
- Swift, K. M., Gross, B. A., Frazer, M. A., Bauer, D. S., Clark, K. J. D., Vazey, E. M., et al. (2018). Abnormal locus coeruleus sleep activity alters sleep signatures of memory consolidation and impairs place cell stability and spatial memory. *Curr. Biol.* 28, 3599–3609.e4. doi: 10.1016/j.cub.2018.09.054
- Tahvildari, B., Wölfel, M., Duque, A., and McCormick, D. A. (2012). Selective functional interactions between excitatory and inhibitory cortical neurons and differential contribution to persistent activity of the slow oscillation. *J. Neurosci.* 32, 12165–12179. doi: 10.1523/JNEUROSCI.1181-12.2012
- Takahashi, N., Oertner, T. G., Hegemann, P., and Larkum, M. E. (2016). Active cortical dendrites modulate perception. *Science* 354, 1587–1590. doi: 10.1126/science.aah6066
- Takehara-Nishiuchi, K., Nakao, K., Kawahara, S., Matsuki, N., and Kirino, Y. (2006). Systems consolidation requires postlearning activation of NMDA receptors in the medial prefrontal cortex in trace eyeblink conditioning. *J. Neurosci.* 26, 5049–5058. doi: 10.1523/JNEUROSCI.4381-05.2006
- Teles-Grilo Ruivo, L. M., and Mellor, J. R. (2013). Cholinergic modulation of hippocampal network function. *Front. Synaptic Neurosci.* 5:2. doi: 10.3389/fnsyn.2013.00002
- Thompson, L. T., and Best, P. J. (1989). Place cells and silent cells in the hippocampus of freely-behaving rats. *J. Neurosci.* 9, 2382–2390. doi: 10.1523/JNEUROSCI.09-07-02382.1989
- Timofeev, I., Grenier, F., and Steriade, M. (2001). Disfacilitation and active inhibition in the neocortex during the natural sleep-wake cycle: an intracellular study. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1924–1929. doi: 10.1073/pnas.041430398
- Todorova, R., and Zugaro, M. (2020). Hippocampal ripples as a mode of communication with cortical and subcortical areas. *Hippocampus* 30, 39–49. doi: 10.1002/hipo.22997
- Tremblay, R., Lee, S., and Rudy, B. (2016). GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* 91, 260–292. doi: 10.1016/j.neuron.2016.06.033
- Trimper, J. B., Trettel, S. G., Hwaun, E., and Colgin, L. L. (2017). Methodological caveats in the detection of coordinated replay between place cells and grid cells. *Front. Syst. Neurosci.* 11:57. doi: 10.3389/fnsys.2017.00057
- Tsanov, M. (2015). Septo-hippocampal signal processing: breaking the code. *Prog. Brain Res.* 219, 103–120. doi: 10.1016/bs.pbr.2015.04.002
- Tsuno, Y., Schultheiss, N. W., and Hasselmo, M. E. (2013). In vivo cholinergic modulation of the cellular properties of medial entorhinal cortex neurons. *J. Physiol. (Lond.)* 591, 2611–2627. doi: 10.1113/jphysiol.2012.250431
- Tukker, J. J. (2019). “Recording identified neurons in awake and anesthetized rodents,” in *Hippocampal Microcircuits: A Computational Modeler's Resource Book*, 2nd Edn, eds V. Cutsuridis, B. Graham, S. Cobb, and I. Vida (Cham: Springer), 365–409.
- Tulving, E. (1985). Memory and consciousness. *Can. Psychol.* 26, 1–12. doi: 10.1037/h0080017
- Ul-Haq, R., Anderson, M., Liotta, A., Shafiq, M., Sherkheli, M. A., and Heinemann, U. (2016a). Pretreatment with  $\beta$ -adrenergic receptor agonists facilitates induction of LTP and sharp wave ripple complexes in rodent hippocampus. *Hippocampus* 26, 1486–1492. doi: 10.1002/hipo.22665

- Ul Haq, R., Anderson, M. L., Hollnagel, J.-O., Worschech, F., Sherkheli, M. A., Behrens, C. J., et al. (2016b). Serotonin dependent masking of hippocampal sharp wave ripples. *Neuropharmacology* 101, 188–203. doi: 10.1016/j.neuropharm.2015.09.026
- Umbriaco, D., Watkins, K. C., Descarries, L., Cozzari, C., and Hartman, B. K. (1994). Ultrastructural and morphometric features of the acetylcholine innervation in adult rat parietal cortex: an electron microscopic study in serial sections. *J. Comp. Neurol.* 348, 351–373. doi: 10.1002/cne.903480304
- Valero, M., and Buzsáki, G. (2019). *Physiological and Network Properties of Down State-Active Neurons in Neocortex*. Available online at: <https://www.abstractsonline.com/pp8/#!/7883/presentation/71933> (accessed October 29, 2019)
- Vandecasteele, M., Varga, V., Berényi, A., Papp, E., Barthó, P., Venance, L., et al. (2014). Optogenetic activation of septal cholinergic neurons suppresses sharp wave ripples and enhances theta oscillations in the hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* 111, 13535–13540. doi: 10.1073/pnas.1411233111
- Vida, I., Halasy, K., Szinyei, C., Somogyi, P., and Buhl, E. H. (1998). Unitary IPSPs evoked by interneurons at the stratum radiatum-stratum lacunosum-moleculare border in the CA1 area of the rat hippocampus in vitro. *J. Physiol.* 506, 755–773. doi: 10.1111/j.1469-7793.1998.755bv.x
- Vyazovskiy, V. V., Olcese, U., Hanlon, E. C., Nir, Y., Cirelli, C., and Tononi, G. (2011). Local sleep in awake rats. *Nature* 472, 443–447. doi: 10.1038/nature10009
- Wang, D. V., and Ikemoto, S. (2016). Coordinated interaction between hippocampal sharp-wave ripples and anterior cingulate unit activity. *J. Neurosci.* 36, 10663–10672. doi: 10.1523/JNEUROSCI.1042-16.2016
- Wang, Y., Toledo-Rodriguez, M., Gupta, A., Wu, C., Silberberg, G., Luo, J., et al. (2004). Anatomical, physiological and molecular properties of Martinotti cells in the somatosensory cortex of the juvenile rat. *J. Physiol. (Lond.)* 561, 65–90. doi: 10.1113/jphysiol.2004.073353
- Waters, J., and Helmchen, F. (2006). Background synaptic activity is sparse in neocortex. *J. Neurosci.* 26, 8267–8277. doi: 10.1523/JNEUROSCI.2152-06.2006
- Wilson, C. J., and Groves, P. M. (1981). Spontaneous firing patterns of identified spiny neurons in the rat neostriatum. *Brain Res.* 220, 67–80. doi: 10.1016/0006-8993(81)90211-0
- Wilson, C. J., and Kawaguchi, Y. (1996). The origins of two-state spontaneous membrane potential fluctuations of neostriatal spiny neurons. *J. Neurosci.* 16, 2397–2410. doi: 10.1523/JNEUROSCI.16-07-02397.1996
- Winterer, J., Maier, N., Wozny, C., Beed, P., Breustedt, J., Evangelista, R., et al. (2017). Excitatory microcircuits within superficial layers of the medial entorhinal cortex. *Cell Rep.* 19, 1110–1116. doi: 10.1016/j.celrep.2017.04.041
- Witter, M. P., Doan, T. P., Jacobsen, B., Nilssen, E. S., and Ohara, S. (2017). Architecture of the entorhinal cortex A review of entorhinal anatomy in rodents with some comparative notes. *Front. Syst. Neurosci.* 11:46. doi: 10.3389/fnsys.2017.00046
- Witter, M. P., and Groenewegen, H. J. (1990). The subiculum: cytoarchitectonically a simple structure, but hodologically complex. *Prog. Brain Res.* 83, 47–58. doi: 10.1016/s0079-6123(08)61240-6
- Wolansky, T., Clement, E. A., Peters, S. R., Palczak, M. A., and Dickson, C. T. (2006). Hippocampal slow oscillation: a novel EEG state and its coordination with ongoing neocortical activity. *J. Neurosci.* 26, 6213–6229. doi: 10.1523/JNEUROSCI.5594-05.2006
- Xu, H., Jeong, H.-Y., Tremblay, R., and Rudy, B. (2013). Neocortical somatostatin-expressing GABAergic interneurons disinhibit the thalamorecipient layer 4. *Neuron* 77, 155–167. doi: 10.1016/j.neuron.2012.11.004
- Xu, M., Chung, S., Zhang, S., Zhong, P., Ma, C., Chang, W.-C., et al. (2015). Basal forebrain circuit for sleep-wake control. *Nat. Neurosci.* 18, 1641–1647. doi: 10.1038/nn.4143
- Yamamoto, J., and Tonegawa, S. (2017). Direct medial entorhinal cortex input to hippocampal CA1 is crucial for extended quiet awake replay. *Neuron* 96, 217–227.e4. doi: 10.1016/j.neuron.2017.09.017
- Yang, S., Yang, S., Moreira, T., Hoffman, G., Carlson, G. C., Bender, K. J., et al. (2014). Interlamellar CA1 network in the hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12919–12924. doi: 10.1073/pnas.1405468111
- Zagha, E., Casale, A. E., Sachdev, R. N. S., McGinley, M. J., and McCormick, D. A. (2013). Motor cortex feedback influences sensory processing by modulating network state. *Neuron* 79, 567–578. doi: 10.1016/j.neuron.2013.06.008
- Zagha, E., and McCormick, D. A. (2014). Neural control of brain state. *Curr. Opin. Neurobiol.* 29, 178–186. doi: 10.1016/j.conb.2014.09.010
- Zolnik, T. A., Ledderose, J., Toumazou, M., Trimbuch, T., Oram, T., Rosenmund, C., et al. (2020). Layer 6b is driven by intracortical long-range projection neurons. *Cell Rep.* 30, 3492–3505.e5. doi: 10.1016/j.celrep.2020.02.044
- Zucca, S., D'Urso, G., Pasquale, V., Vecchia, D., Pica, G., Bovetti, S., et al. (2017). An inhibitory gate for state transition in cortex. *eLife* 6:e26177. doi: 10.7554/eLife.26177
- Zucca, S., Pasquale, V., Lagomarsino, de Leon Roig, P., Panzeri, S., and Fellin, T. (2019). Thalamic drive of cortical parvalbumin-positive interneurons during down states in anesthetized mice. *Curr. Biol.* 29, 1481–1490.e6. doi: 10.1016/j.cub.2019.04.007

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Tukker, Beed, Schmitz, Larkum and Sachdev. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The (Un)Conscious Mouse as a Model for Human Brain Functions: Key Principles of Anesthesia and Their Impact on Translational Neuroimaging

Henning M. Reimann<sup>1\*</sup> and Thoralf Niendorf<sup>1,2</sup>

<sup>1</sup>Berlin Ultrahigh Field Facility (B.U.F.F.), Max-Delbrück Center for Molecular Medicine, Helmholtz Association of German Research Centers (HZ), Berlin, Germany, <sup>2</sup>Experimental and Clinical Research Center, A Joint Cooperation Between the Charité Medical Faculty and the Max-Delbrück Center for Molecular Medicine, Berlin, Germany

## OPEN ACCESS

### Edited by:

Cyriel Pennartz,  
University of Amsterdam,  
Netherlands

### Reviewed by:

Emery N. Brown,  
Massachusetts Institute of  
Technology, United States  
Jorrit Steven Montijn,  
Netherlands Institute for  
Neuroscience (KNAW), Netherlands

### \*Correspondence:

Henning M. Reimann  
henning.reimann@mdc-berlin.de

**Received:** 02 July 2019

**Accepted:** 27 January 2020

**Published:** 19 May 2020

### Citation:

Reimann HM and Niendorf T  
(2020) The (Un)Conscious Mouse as  
a Model for Human Brain Functions:  
Key Principles of Anesthesia and  
Their Impact on Translational  
Neuroimaging.  
Front. Syst. Neurosci. 14:8.  
doi: 10.3389/fnsys.2020.00008

In recent years, technical and procedural advances have brought functional magnetic resonance imaging (fMRI) to the field of murine neuroscience. Due to its unique capacity to measure functional activity non-invasively, across the entire brain, fMRI allows for the direct comparison of large-scale murine and human brain functions. This opens an avenue for bidirectional translational strategies to address fundamental questions ranging from neurological disorders to the nature of consciousness. The key challenges of murine fMRI are: (1) to generate and maintain functional brain states that approximate those of calm and relaxed human volunteers, while (2) preserving neurovascular coupling and physiological baseline conditions. Low-dose anesthetic protocols are commonly applied in murine functional brain studies to prevent stress and facilitate a calm and relaxed condition among animals. Yet, current mono-anesthesia has been shown to impair neural transmission and hemodynamic integrity. By linking the current state of murine electrophysiology,  $\text{Ca}^{2+}$  imaging and fMRI of anesthetic effects to findings from human studies, this systematic review proposes general principles to design, apply and monitor anesthetic protocols in a more sophisticated way. The further development of balanced multimodal anesthesia, combining two or more drugs with complementary modes of action helps to shape and maintain specific brain states and relevant aspects of murine physiology. Functional connectivity and its dynamic repertoire as assessed by fMRI can be used to make inferences about cortical states and provide additional information about whole-brain functional dynamics. Based on this, a simple and comprehensive functional neurosignature pattern can be determined for use in defining brain states and anesthetic depth in rest and in response to stimuli. Such a signature can be evaluated and shared between labs to indicate the brain state of a mouse during experiments, an important step toward translating findings across species.

**Keywords:** mouse, fMRI-functional magnetic resonance imaging, anesthesia, translational mouse-human, brain functional connectivity, EEG,  $\text{Ca}^{2+}$  imaging, MIND signature



## INTRODUCTION

Much of our understanding of human brain functions comes from murine studies. The ease of genetic modification and other practical and financial issues have made the mouse the best-explored mammalian model organism in neuroscience. A multitude of murine protocols and repositories provide reliable benchmarks in today's brain research (including the *Allen Brain Project*<sup>1</sup>, the *Blue Brain Project*<sup>2</sup>, and the *Mouse ENCODE Project*<sup>3</sup>; Lein et al., 2007; Sunkin et al., 2013; Erö et al., 2018; Keller et al., 2018; Frankish et al., 2019). Still, it remains an open question to what extent, and under what conditions, findings from the mouse can be translated into an understanding of human brain functions. Although the mouse brain is not merely a miniaturized version of the human brain, its comparably small and flat neocortex exhibits striking functional and structural similarities, and the subcortical architecture is evolutionarily largely preserved (Ventura-Antunes et al., 2013; Hofman, 2014; Glickfeld and Olsen, 2017; Halley and Krubitzer, 2019).

In recent years, technical and procedural advances have brought functional magnetic resonance imaging (fMRI) to the field of murine neuroscience. The unique capacity of fMRI to measure functional brain activity non-invasively and across the entire brain relies on tight neurovascular coupling, in which increased neural activity triggers local elevations in cerebral blood flow (CBF), cerebral blood volume (CBV), and blood oxygenation (Hamilton et al., 2010; Hall et al., 2014). Each of these hemodynamic parameters can be assessed by fMRI as a surrogate for neural activity. Blood oxygenation level-dependent (BOLD) fMRI is the most popular approach due to its high sensitivity and comparably fast acquisition times (Ogawa et al., 1990; Kim, 2018). Modern MR scanners operating at ultrahigh magnetic field strengths ( $\geq 7$  T) have been tailored for use with small rodents, and can achieve a relative spatial resolution analogous to that commonly used in human fMRI (voxels per anatomical region; voxel size of  $\sim 200 \mu\text{m}^3$ ). This permits direct comparisons of large-scale murine and human brain functions, and opens up opportunities to use a plethora of genetically engineered models to clarify a wide range of clinical and basic neuroscience issues—from the pathogenesis of neurological disorders to fundamental questions about consciousness.

Although the first report on murine fMRI was published more than two decades ago (Huang et al., 1996), its application to mice continues to require extensive refinements (Mandino et al., 2020). Aside from technical issues including signal amplification from the small mouse brain or its vulnerability to physiological perturbations, a fundamental problem of interspecies translation involves the *non-voluntary task conundrum*. In comparison to (adult) humans, a mouse has no interest in participating in an fMRI study. Head fixation, body restraint, and habituation to the acoustic noise produced by the MR gradients (up to 115 dB sound pressure) cause enormous stress to this remarkably rousable

creature, which is far less amenable to training than rats (Jonckers et al., 2014, 2015; Low et al., 2016b; Dopfel and Zhang, 2018). Studying unbiased nociception in mice is almost impossible; even the application of aversive stimuli, like mild cooling of the paw, causes immediate withdrawal, and functional patterns across the brain reflect not only stress, fear, and anticipation, but also unrelated motor and sensory responses. This introduces tremendous complexity to fMRI of the awake mouse.

Low-dose anesthetic protocols are commonly applied in mouse fMRI to address these problems and to alleviate potential suffering due to stress, fear, and pain. The goal is to achieve high-quality translational data from sedated, i.e., calm, relaxed, and undistracted subjects. However, anesthetics have been shown to impair neural transmission (Baumgart et al., 2015; Hemmings et al., 2019) and also affect other aspects of murine physiology including hemodynamics (Franceschini et al., 2010; Masamoto and Kanno, 2012), thermoregulation (Reimann et al., 2016), respiration (van Alst et al., 2019), and cardiovascular control (Sinclair, 2003; Low et al., 2016a) in a dose-dependent manner. All these parameters can affect neurovascular coupling, which links the BOLD effect to the activity of neural populations. Therefore, anesthetic protocols in fMRI face two major challenges: first, to preserve functional *brain states* based on neural oscillation and transmission characteristics (see “Brain States, Anesthetic Depth, and Murine Consciousness” section), and second, to maintain hemodynamic integrity (see “Murine fMRI and Hemodynamic Integrity” section).

Recognizing the progress, opportunities, and challenges of murine functional brain mapping, this work provides a review of the literature on current anesthetic protocols and their meaning for murine functional neuroimaging. The aim is to introduce the basic principles of anesthesia to better understand and interpret the outcome of murine fMRI studies, and to develop novel anesthetic protocols and monitoring strategies dedicated to promoting reproducible and translational neuroimaging.

## BRAIN STATES, ANESTHETIC DEPTH, AND MURINE CONSCIOUSNESS

In the murine as in the human brain, neural information is processed against a background of spontaneous, ongoing activity generated by the promiscuous firing of neurons or structured patterns of neural populations. These oscillatory dynamics can be assessed *via* invasive and non-invasive neural readout techniques, such as electroencephalography (EEG), local field potentials (LFP), or calcium ( $\text{Ca}^{2+}$ ) imaging, yielding distinct spatiotemporal profiles that can be considered as signatures of distinct brain states.

### Cortical States of Wakefulness

Awake brain states undergo constant alterations in response to changing contexts of arousal, attention, and behavior (Steriade, 2000; Pfaff, 2005; Olcese et al., 2018; Poulet and Crochet, 2019). The spectrum of EEG patterns during these states ranges from synchronous low-frequency oscillations ( $< 8$  Hz) during quiet wakefulness to structured higher-frequency oscillations ( $\sim 8$ –100 Hz) during attentional tasks

<sup>1</sup>mouse.brain-map.org

<sup>2</sup>bbp.epfl.ch/nexus/cell-atlas

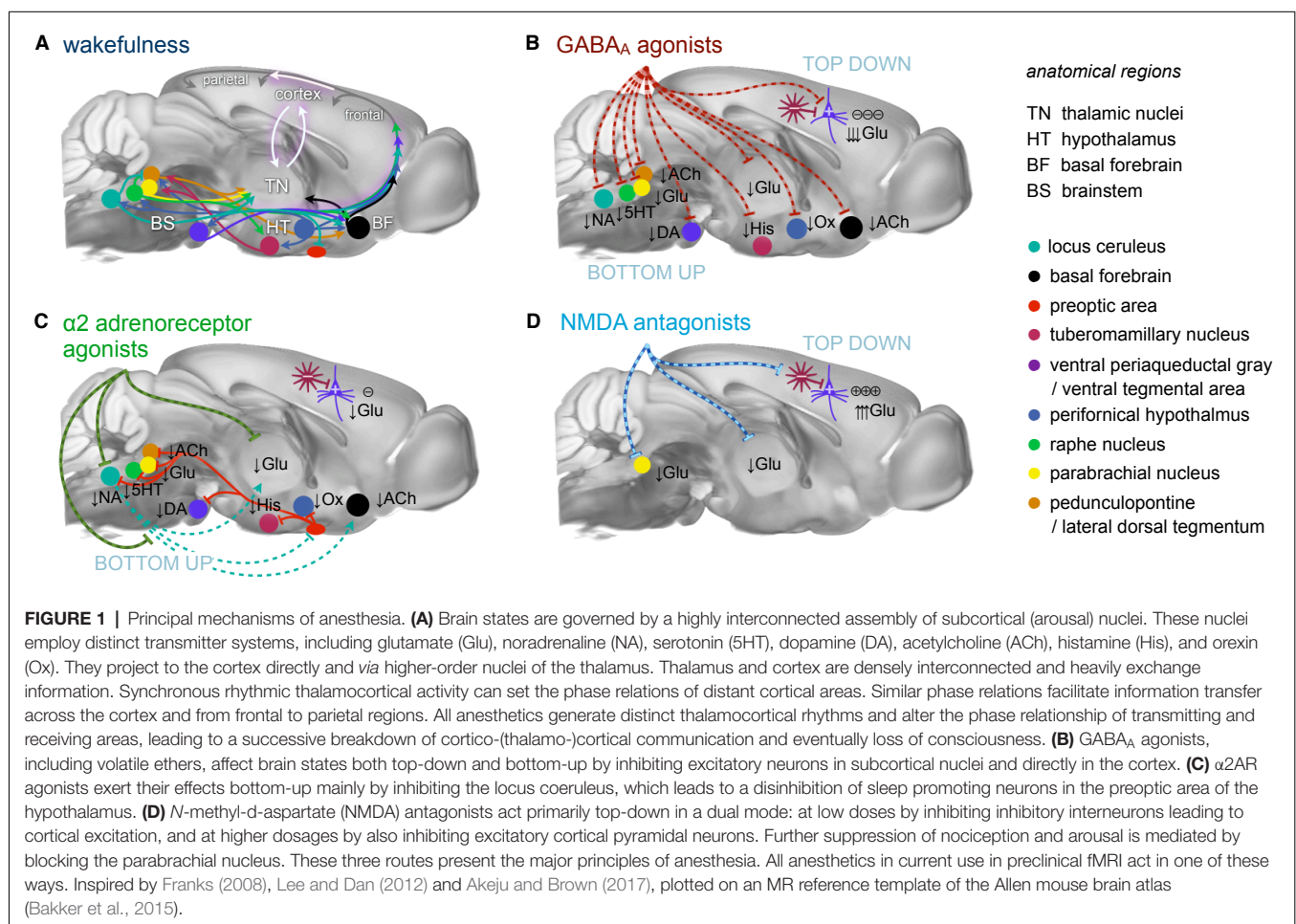
<sup>3</sup>encodeproject.org

and highly desynchronized high-frequency oscillations during aroused states (Petersen et al., 2003; Palva et al., 2005; Fries et al., 2007; Sachidhanandam et al., 2013; McGinley et al., 2015a; Olcese et al., 2018; Poulet and Crochet, 2019). The level of *cortical activation* (i.e., higher-frequency oscillations) is positively correlated to the level of general arousal, which is mediated by modulatory input from subcortical cholinergic and monoaminergic (dopaminergic, noradrenergic, serotonergic, and histaminergic) nuclei (**Figure 1A**; Aston-Jones et al., 1999; Pfaff, 2005; Harris and Thiele, 2011; Lee and Dan, 2012; McGinley et al., 2015b; Lecrux and Hamel, 2016; Ma et al., 2018). These arousal nuclei project into the cortex either directly or *via* the medial thalamus, and govern cortical states at both the global and local levels. Defined cortical regions have been shown to be remotely modulated, for instance, by inhibitory projections of the thalamic reticular nucleus (Lewis et al., 2015; Herrera et al., 2016; Fernandez et al., 2017) or excitatory projections of cholinergic neurons in a modality-selective manner (Kim et al., 2016; Záborszky et al., 2018).

Sensory stimuli are processed differently against the background of diverse cortical states (Shimaoka et al., 2018; Poulet and Crochet, 2019). Ensembles of neighboring neurons depend on a certain level of desynchronization to encode complex features of stimuli by concerting the firing rate, spike

timing, and the temporal order at which they fire (Hopfield, 1995; Montemurro et al., 2008; Kayser et al., 2009; Mohajerani et al., 2013; Luczak et al., 2015; Montijn et al., 2016). Increasing synchronization may lend structure to features, but can also obscure them (Fries et al., 2007; Pachitariu et al., 2015; Olcese et al., 2018). An intermediate level of arousal has been found to enhance the consistency and signal strength of encoded stimuli and is associated with optimal sensory processing (Polack et al., 2013; Schneider and Logan, 2014; Schölvinck et al., 2015; McGinley et al., 2015a; Olcese et al., 2018; Shimaoka et al., 2018). Cortical information processing is substantially shaped by the activity of diverse inhibitory interneurons and relies on a balanced interplay of excitatory and inhibitory inputs (Isaacson and Scanziani, 2011; Rubin et al., 2017). Cortical states can become altered regionally and globally upon “attention” to a stimulus (Olcese et al., 2018; Poulet and Crochet, 2019). This flexibility permits the mode of sensory processing to adapt to situational demands.

Such modes of attention, sensory processing, and integration have been described in terms of cortical oscillations of distinct wavelengths (delta, theta, alpha, beta, and gamma, as reviewed in Fries et al., 2007; Schroeder and Lakatos, 2009; Saalmann et al., 2012; Fries, 2015; McVea et al., 2016; Jensen et al., 2019; Sikkens et al., 2019). The underlying rhythmic synchronization of



neural assemblies is largely produced by a synchronized spiking of inhibitory interneurons at higher frequencies (reviewed in Fries et al., 2007; Jensen et al., 2019). The high magnitudes of slower rhythms are generated by other mechanisms (see “Brain States Under Anesthesia” section; reviewed in Neske, 2015; Sanchez-Vives et al., 2017). Neurons that act in phase are more likely to fire together, since the active and refractory periods of presynaptic and postsynaptic neurons are aligned (Saalmann, 2014). This temporal coding scheme can facilitate or inhibit information transfer between local and distant cell groups by varying their phase relations.

Cortical oscillations can propagate as traveling waves of neural depolarization and spiking activity across the cortical surface (reviewed in McVea et al., 2016; Kuroki et al., 2018; Muller et al., 2018). Higher frequencies are thereby nested within slower waves, leading to rich patterns of background information onto which local events are processed. The distinct and complex spatiotemporal dynamics of these traveling waves (e.g., radial, planar, spiraling, or rotating) are coordinated by recurrent cortical networks and distant projections from thalamic and subthalamic nuclei (Bhattacharya et al., 2019). Anesthesia can affect neural information processing at both the local level (by setting an imbalance of inhibitory and excitatory inputs) and globally (by spoiling phase relations of cell ensembles and disrupting information transfer across brain areas).

## Brain States Under Anesthesia

Selecting an appropriate anesthetic protocol is in first place a question of choosing the right anesthetic class and dosage. Anesthetics currently used for murine fMRI (Table 1) can be divided into three classes based on their main molecular target receptors:  $\gamma$ -aminobutyric acid subtype A (GABA<sub>A</sub>) receptor agonists including volatile ethers,  $\alpha$ 2 adrenoreceptor ( $\alpha$ 2AR) agonists, and *N*-methyl-D-aspartate (NMDA) receptor antagonists. Each class exerts its sedating effects based on one of three key principles: (1) top-down by inhibiting cortical neurons directly (NMDA receptor antagonists); (2) bottom-up by suppressing subcortical arousal nuclei that affect cortical states ( $\alpha$ 2AR agonists); or (3) both, by inhibiting neurons across the entire brain (GABA<sub>A</sub> agonists and volatile ethers; Figure 1).

Because their target sites differ, each anesthetic class produces distinct cortical oscillatory dynamics (Figures 2A,B; Ching and Brown, 2014; Purdon et al., 2015; Flores et al., 2017), which vary with the anesthetic dosage (Figures 2B,C). Most anesthetics within one class further exhibit variations in their cortical signatures as they act at multiple receptor types (Table 1), interfering with partially overlapping pathways. Scientific advances over the last decades have revealed an increasingly comprehensive, yet still incomplete picture of the underlying mechanisms (Steriade, 2000; Campagna et al., 2003; Alkire et al., 2008; Franks, 2008; Brown et al., 2010, 2018a; Akeju and Brown, 2017; Flores et al., 2017; Hemmings et al., 2019). Here, we briefly introduce current models on how these rhythms are generated by the effects of different anesthetics, and how the

brain states that are indicated by distinct rhythms affect neural transmission and processing.

## Thalamocortical Rhythmogenesis

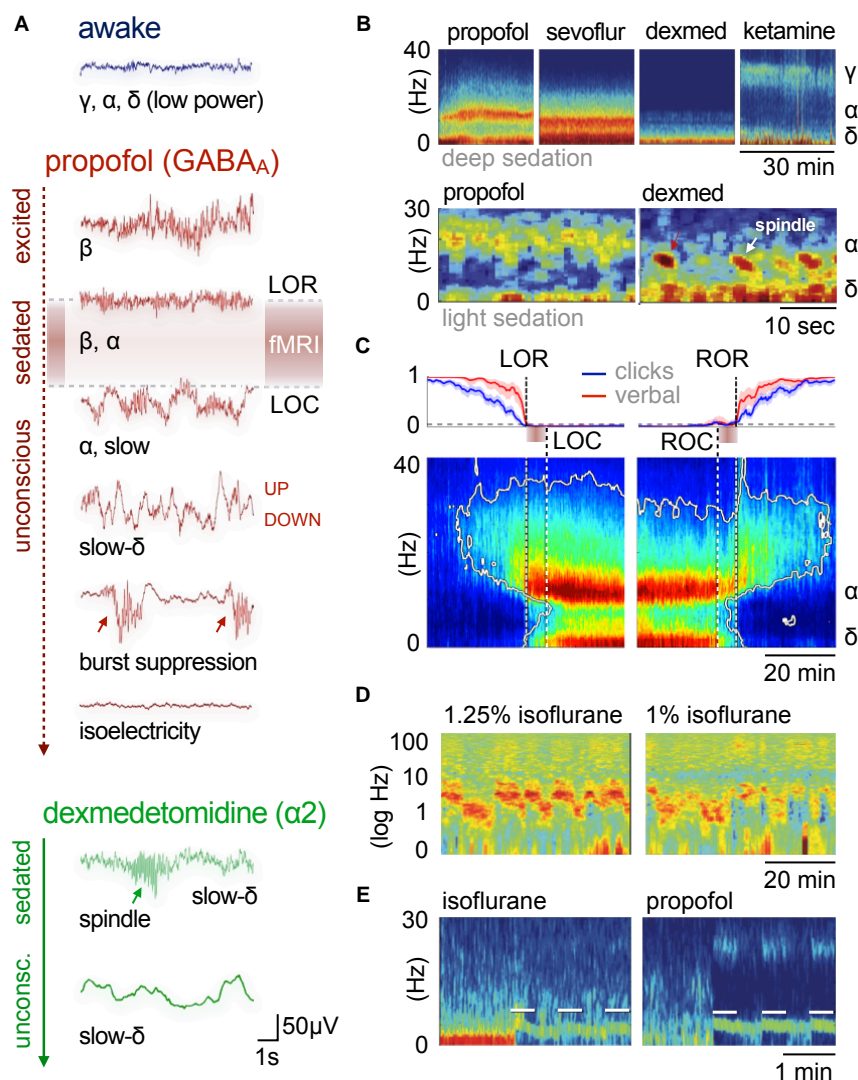
The generation of cortical rhythms by anesthesia follows a common pattern for all anesthetics (as described in more detail below). All anesthetics exert their sedating effects by acting on specific target receptor types (Table 1), leading to inhibition and thus hyperpolarization of the affected neurons. Many types of excitatory and inhibitory neurons comprise specific membrane channels that open at distinct levels of hyperpolarization and lead to inward cation currents that cause the cells to quickly depolarize. Hence, the affected cells respond to strong inhibition with excitation, leading to *post-inhibitory rebound spiking*. These bursts of neural activity recur periodically, paced by the sum of inhibitory (hyperpolarizing) and excitatory (depolarizing) inputs that are required to exceed the threshold for rebound spiking. These inputs can be differently modulated by distinct anesthetic types and dosages. The resulting spiking patterns reverberate in recurrent neuronal circuits and can synchronize large networks of cortical neurons by reciprocal coupling, which manifests as cortical oscillations on EEG. These oscillations are generated either within the cortex or the thalamus, or between these two densely interconnected regions (Figure 3A; reviewed in McCormick and Bal, 1997; Steriade, 2001; Fuentealba and Steriade, 2005; Franks, 2008; Ching and Brown, 2014; Akeju and Brown, 2017).

Thalamic key players of rhythm generation are excitatory thalamocortical (TC) cells in intralaminar and medial thalamic nuclei (Baker et al., 2014; Flores et al., 2017). These higher-order TC nuclei receive and process input from cortical areas and serve as crucial relays of information transfer across the cortex (Theyel et al., 2010; Sherman, 2017; Mo and Sherman, 2019). Effective information transfer is thereby facilitated by frequency and phase synchronization of projection, relay, and target areas (Saalmann, 2014). In contrast, first-order TC nuclei primarily relay incoming sensory information to the cortex. First and higher-order TC nuclei undergo a major inhibitory impact of GABAergic thalamic reticular (RE) neurons, which project from the thin, outer shell of the thalamus—the thalamic reticular nucleus (as reviewed in Varela, 2014). RE neurons control TC traffic, synchronicity of firing patterns, and the level of arousal by targeted inhibition of TC cells (Sun et al., 2012). The level of depolarization in both TC and RE neurons is further regulated by afferents from cholinergic and monoaminergic arousal nuclei (Saper et al., 2010), which also modulate the membrane potential of cortical neurons: GABAergic inhibitory interneurons and glutamatergic excitatory pyramidal cells (Figure 3A).

## GABA<sub>A</sub> Agonists Suppress Neural Activity Across the Central Nervous System

Volatile ethers and GABA<sub>A</sub> agonists such as isoflurane and propofol predominantly act on postsynaptic GABA<sub>A</sub> receptors of excitatory neurons by increasing the preference of those ligand-gated chloride channels for the open state (Bai et al., 1999; Lee and Maguire, 2014). This enhances the inward chloride current and facilitates hyperpolarization and thus





**FIGURE 2 |** Cortical oscillatory signatures of brain states under anesthesia. **(A)** Changes in oscillatory signatures for increasing doses of propofol and dexmedetomidine measured from the human anterior cortex. With increasing anesthetic depth, oscillation frequencies decrease and synchronize. **(B)** Distinct oscillatory signatures induced by different anesthetics (upper panel) and dosages (lower panel). The spectrogram represents the spectrum of frequencies in a time and frequency domain, which facilitates the identification of structured frequency bands. **(C)** Loss of responsiveness (LOR) and connected consciousness (LOC) induced by increasing doses of propofol. LOR is defined by suppression of responses to click and verbal commands (upper panel). Note the occurrence of alpha bands and the absence of slow-delta bands in the phase between LOR and LOC in the spectrogram. The same is true for return of consciousness (ROC) and responsiveness (ROR). In rodents, alpha waves coincide with the loss of righting reflex (LORR), which corresponds to LOR, and slow-delta waves are indicated by a complete loss of movement (LOM), which corresponds to LOC. **(D)** Brain states may not be fully stable when maintained *via* anesthesia. Transitions between two or more intermediate states occurred over longer periods of constant isoflurane concentration in rats. This feature has been defined as metastability. Note the logarithmic scale used to highlight the transitions that occur primarily in the lower frequency band (0–10 Hz). Local field potentials (LFP) recording was conducted in the anterior cingulate cortex (ACC). **(E)** Stimulation (white bars) of dopaminergic neurons of the ventral tegmental area in the rat shifts cortical states under sedation from slow delta (<4 Hz) towards  $\theta$  power (isoflurane), or towards  $\theta$  and  $\beta$  power (propofol). *Oscillations:* slow (<1 Hz),  $\delta$  (1–4 Hz),  $\theta$  (4–8 Hz),  $\alpha$  (8–15 Hz),  $\beta$  (15–30 Hz), awake- $\gamma$  (30–80 Hz), ketamine- $\gamma$  (25–35 Hz), spindle (9–16 Hz). Adapted with permission from **(A,B)** Purdon et al. (2015), **(B, lower panel)** Akeju and Brown (2017), **(C)** Purdon et al. (2013), **(D)** Hudson (2017), and **(E)** Solt et al. (2014).

inhibition of the postsynaptic cell. Presynaptic actions that impair neurotransmitter release (Hemmings et al., 2005, 2019; Baumgart et al., 2015) further contribute to reducing neural activity across the central nervous system (CNS), including the cortex, thalamus, striatum, brainstem, and even the spinal cord (Figure 1B; Bowery et al., 1987; Hemmings et al., 2005; Brown

et al., 2010; Phillips et al., 2018). This exerts hypnotic effects and gradually renders the animal unconscious in a dose-dependent manner (Akeju and Brown, 2017; Flores et al., 2017).

An initial excitement is often observed at the induction of anesthesia, accompanied by a disinhibition of motor activity and the emergence of relatively fast cortical oscillatory activity in the



**TABLE 1** | Anesthetic compounds used in mouse functional magnetic resonance imaging (fMRI)

	target receptors						
	GABA <sub>A</sub>	α2	NMDA	Glycine	AMPA	Kainate	
propofol	++		−	++	−	−	Schroeter (2014); Grandjean (2014)
etomidate	++			+			Petrinovic (2016); Klee (2017a)
benzodiazepine	++			++	−	−	Sharp (2015*)
barbiturate	++			+	−−	−−	Huang (1996)
isoflurane	++		−	++	−−	++	Heindl-Erdmann (2010); Baltes (2010); Bosshard (2011); Kahn (2011); Hess (2011); Vetter (2012); Farrari (2012); Schroeter (2014; 2017); Grandjean (2014; 2017*; 2019*); Reimann (2016; 2018); Zerbi (2019*)
sevoflurane	++		−−	++	−−	++	Klee (2017b)
halothane	++		−	++	−−	++	Sforazzini (2014); Liska (2015); Giorgi (2017); Gutierrez-Barragan (2018)
medetomidine		++	−				Adamczak (2010); Jonckers (2011); Shah (2013; 2015); Schroeter (2014); Grandjean (2014; 2017*; 2019*); Nasrallah (2014); Niranjana (2016); Li (2017); Blazquez Freches (2017); Ben-Nejma (2019); Zerbi (2019*)
xylazine		++	−				Shim (2018*); Jung (2019*)
ketamine	+		−−				Shim (2018*); Jung (2019*)

Each drug addresses distinct receptor types by either agonistic (+) or antagonistic (−) modulation. Single or double symbols are used to indicate whether potentiation or inhibition is mild or strong. Anesthetics are grouped into three classes based on their main target receptors: GABA<sub>A</sub> agonists, including injectable drugs and volatile ethers (red and yellow, respectively), α2AR agonists (green), and NMDA antagonists (blue). Alongside these key target receptors, the listed drugs modulate a host of further receptor types and may contribute to specific anesthetic effects (see, e.g., Alkire et al., 2008; Franks, 2008; McKinstry-Wu and Kelz, 2019). Halothane is chemically not an ether, yet it belongs to the group of halogenated vapors, like isoflurane and sevoflurane, which share largely overlapping target sites. Note that only drugs suitable for longitudinal studies are listed; the GABA<sub>A</sub> agonists α-chloralose and urethane have also been applied in murine fMRI (Ahrens and Dubowitz, 2001; Xu et al., 2003, 2005; Grandjean et al., 2014; Schroeter et al., 2014) and will be discussed in "Sensory Processing and the Key Challenges in Murine fMRI" and "Functional Connectivity and Murine Resting-State fMRI" sections. References are shortened to first author (and year) to save space and allow for more comprehensive referencing. Asterisks indicate use in multimodal protocols.

beta band (~13–30 Hz; **Figure 2A**; Ching and Brown, 2014; Le Van Quyen et al., 2016). The precise mechanisms responsible for this *paradoxical excitation* have yet to be fully explained. Network models indicate the involvement of hyperpolarization-triggered inward currents that cause cortical circuits of excitatory and inhibitory neurons to produce periodic firing in the beta range (McCarthy et al., 2008). With increasing inhibitory tone and sedation, beta rhythms slow down to alpha frequencies (**Figures 2A, 3A**; McCarthy et al., 2008; Baker et al., 2014; Flores et al., 2017). This is the onset of stable sedation marked by the loss of the righting reflex (LORR; Baker et al., 2014; Flores et al., 2017).

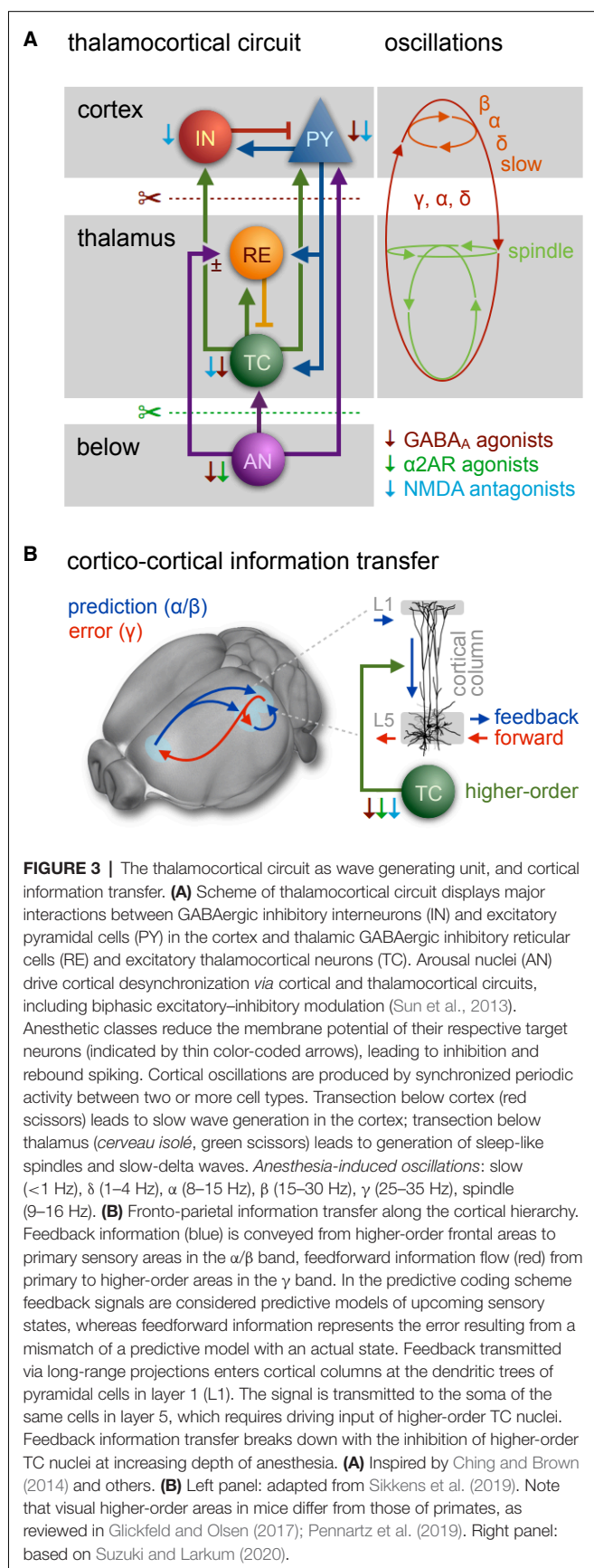
Although alpha oscillations (8–15 Hz) are produced within the cortex (Mukamel et al., 2014), the generation of temporally and spatially coherent alpha rhythms across larger areas requires the participation of the thalamus (Ching et al., 2010; Flores et al., 2017). The GABAergic inhibition of TC cells is thought to trigger enhanced post-inhibitory rebound spiking that is paced by the decay rate of inhibition, and thus depends on the anesthetic dosage (**Figure 3A**; Ching and Brown, 2014). Thalamic and cortical alpha activities reinforce each other by reciprocal coupling, resulting in strong TC synchronicity (**Figure 3A**; Ching et al., 2010; Ching and Brown, 2014; Crunelli et al., 2015; Bhattacharya et al., 2019).

Dense recurrent projections from the thalamus convey broad spatial coherence of alpha oscillations in the frontal cortex (Flores et al., 2017). In contrast, occipital thalamic projections

of high-threshold TC neurons, which produce coherent alpha oscillations in the awake state are suppressed by a reduction of hyperpolarization-activated cation currents, presumably by secondary drug effects (Ching et al., 2010; Cimenser et al., 2011; Vijayan et al., 2013). This causes a characteristic spatial shift in EEG alpha power from the posterior to the anterior part of the brain, a phenomenon known as *anteriorization* (Akeju et al., 2014b).

LORR is likely caused by impaired information transfer and integration across cortical areas due to changing oscillatory modes. When sending and receiving cells are tuned in phase, postsynaptic neurons receive input during their active periods and are more likely to fire. Phase alignment facilitates neural transmission across distant cortical areas, which also involves higher-order TC nuclei (Saalmann, 2014; Sherman, 2017; Mo and Sherman, 2019). Higher-order TC nuclei are engaged with GABAergic drug-induced frontal alpha waves in humans and rodents (Ching et al., 2010; Liu et al., 2013a; Baker et al., 2014; Flores et al., 2017), which might affect phase-frequency relationships between sending and receiving areas and thus contribute to the disruption of cross-cortical communication.

However, the loss of awake-alpha (and higher frequency) waves has been shown to be a more reliable marker for stable sedation (Blain-Moraes et al., 2015; Pavone et al., 2017). A loss of awake-alpha is typical for the first stage of non-rapid eye movement (N1) sleep (Prerau et al., 2017) and was correlated



with a lack of behavioral response to stimuli in N1 sleep (Prerau et al., 2014) and under light sedation *via* sevoflurane (Pavone et al., 2017).

Awake-alpha waves play an essential role in information transfer and integration across the cortex (Jensen et al., 2019; Senzai et al., 2019). For instance, neural activity in the alpha band in contralateral sensorimotor areas has been shown to phase lock robustly to somatosensory stimuli. If these stimuli are perceived consciously, the stimulus locking spreads rapidly to the frontal, parietal, and ipsilateral sensorimotor regions (Palva et al., 2005). For unperceived stimuli, the phase locking is weak and restricted to the initial sensorimotor area.

Synchronous alpha activity in frontal, parietal, and sensorimotor regions facilitates recurrent information transfer between these areas due to a phase alignment of neural activity in transmitting and receiving areas. Such alpha-phase synchrony between primary sensory and higher-order cortical areas is considered a neural basis of attention, and is strongly mediated by frontal and parietal regions. Fronto-parietal information transfer is involved in many higher-order functions, including the formation of higher-order functional networks in the awake, resting state (see “Functional Connectivity and Murine Resting-State fMRI” section; Coull, 1998; Kastner and Ungerleider, 2000; Rees et al., 2002; Zeman, 2004; Klimesch et al., 2007; Palva and Palva, 2007; Sadaghiani et al., 2012; Fries, 2015; Han et al., 2019; Sikkens et al., 2019).

Recurrent processing along the cortical hierarchy accounts for the contextual modulation of stimuli, including perceptual and semantic interpretation, endowing them with spatiotemporal context and behavioral significance (Pennartz et al., 2019). One function of recurrent processing is expressed in the *predictive coding framework*, in which the brain is constantly generating and updating sensory models of the world (as reviewed in Pennartz et al., 2019; Sikkens et al., 2019). Feedback projections from higher-order (e.g., prefrontal) to early sensory areas are thought to convey or facilitate predictive models of upcoming sensory states *via* alpha/beta oscillations. A mismatch of a predictive model with actual sensory input results in error signals that are presumably fed forward from early sensory to higher-order areas *via* gamma oscillations—through different, partially overlapping cortical layers—to update the model in a perpetual feedback/feedforward loop (Figure 3B, left panel; van Kerkoerle et al., 2014; Zhang et al., 2014; Bastos et al., 2015; D’Souza et al., 2016; Chao et al., 2018; Michalareas et al., 2016; Kissinger et al., 2018; Zhang et al., 2019).

Feedback and feedforward information transfer are considered cornerstones of effective signal processing in the conscious brain (as reviewed in Lamme, 2018; Pennartz et al., 2019). LORR can probably be attributed to impaired feedback or feedforward information transfer (Pavone et al., 2017; Sanders et al., 2018; Redinbaugh et al., 2020). Long-range feedback modulation, which in primary sensory areas promotes both alpha waves (Jensen et al., 2019; Sikkens et al., 2019) and desynchronized activity (Harris and Thiele, 2011; Ecker et al., 2016), undergoes a breakdown with increasing anesthetic depth (Imas et al., 2005b; Raz et al., 2014; Hentschke et al., 2017; Sanders et al., 2018; Murphy et al., 2019).

A recent cellular study revealed that cortico-cortical feedback flow through the cortical layers (mediated by L5 pyramidal cells) requires the driving input of higher-order TC nuclei (**Figure 3B**, right panel; Suzuki and Larkum, 2020). Since these higher-order TC nuclei in turn depend on the driving input of cholinergic projections from the brainstem (**Figure 3A**; Trageser et al., 2006; Masri et al., 2006, 2008), they appear to be particularly susceptible to the effects of various anesthetics. At already 1% isoflurane the inhibition of higher-order TC nuclei causes a disruption of cortico-cortical feedback loops.

This causal relationship unifies two central hypotheses of consciousness, which had previously assumed a dependence on either cortical feedback or higher-order TC loops (Alkire et al., 2008; Boly et al., 2011; Mashour, 2014). Also, feedforward signal propagation toward higher-order cortical regions has been reported to be increasingly suppressed with thalamocortical inhibition (Massimini et al., 2005; Sellers et al., 2015; Casarotto et al., 2016; Hentschke et al., 2017; Sanders et al., 2018; Redinbaugh et al., 2020). Therefore, studying sensory pathways within the cortex of anesthetized mice is often limited to brain areas that receive direct sensory thalamic or subcortical input.

Neural responses to sensory stimuli in primary cortices are largely preserved across sensory modalities (Lamme et al., 1998; Detsch et al., 1999; Pack et al., 2001; Imas et al., 2005a; Greenberg et al., 2008; Schumacher et al., 2011; Haider et al., 2013; Milenkovic et al., 2014; Raz et al., 2014; Sellers et al., 2015). Yet, even low anesthetic dosages alter the spatial and temporal structure of neural firing patterns, thereby disrupting information processing, which relies on precise timing of ensemble activity (Luczak et al., 2015; Yuste, 2015). For example, neural responses to visual stimuli in mice anesthetized with 0.25–1% isoflurane extend into larger V1 areas and are temporally prolonged, as they are less shaped by directed inhibition compared to the awake state (Haider et al., 2013; Sellers et al., 2015). This has also been observed for other drugs and sensory modalities (Devonshire et al., 2010).

### Crossing the Borders: GABAergic Slow Waves and Burst Suppression

Increasing doses of anesthetics bring a complete loss of movement (LOM), which coincides with a sudden rise in slow-delta power; i.e., delta oscillations (1–4 Hz) primarily in the frontal areas, and slow waves (<1 Hz) across the entire cortex (**Figure 2A**; Steriade et al., 1993b,c,d; Flores et al., 2017; Chamadia et al., 2019). Slow-delta waves are characterized by alternations of persistent desynchronized network activity (depolarized “UP” states) and generalized neural silence (hyperpolarized “DOWN” states) of varying duration (Steriade et al., 1993b,c,d, 2001; Luczak et al., 2007). Slow-delta rhythms impose strong oscillatory dominance on the cortical firing patterns by nesting oscillations of higher frequencies in distinctive sequences (Steriade, 2001; Fuentealba and Steriade, 2005). Delta oscillations are grouped themselves by slow waves into larger sequences, although both waveforms share a very similar structure.

The exact phase-amplitude coupling of slower and faster waves is a function of the anesthetic depth (Chamadia et al.,

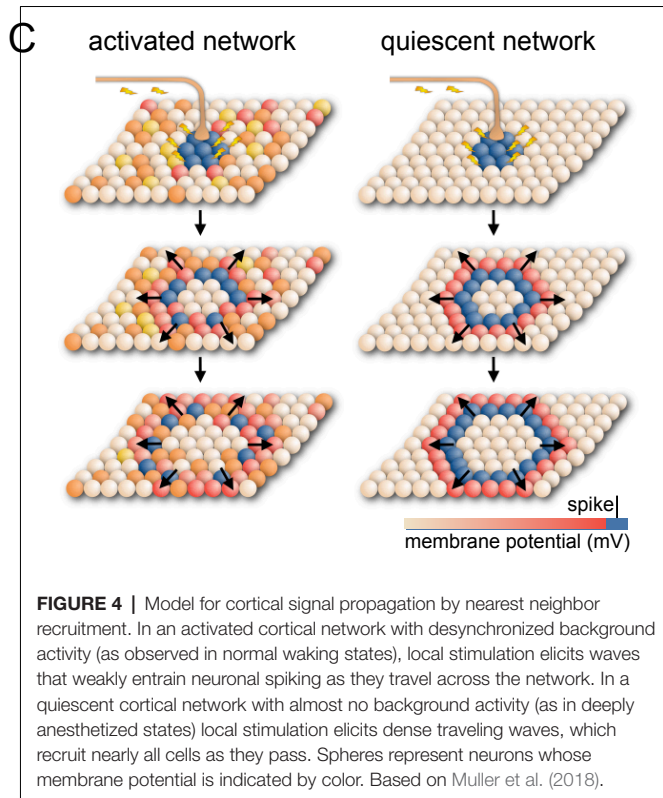
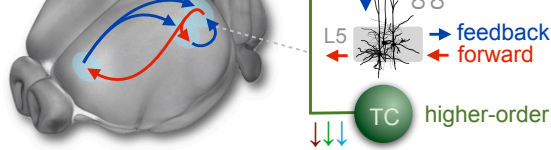
2019). With light sedation, alpha waves lie in the trough of slow-delta oscillations. With increasing anesthetic depth, they experience a shift in phase until they ride on top of the slow waves’ peaks. This phase-amplitude syntax causes a functional decoupling of neural activity, and thus cortical communication to collapse in crucial networks, leading to deep anesthesia (see “Functional Connectivity and Murine Resting-State fMRI” section).

Cortical delta and slow waves persist following removal of the thalamus, although under such conditions, slow waves (at ~0.3 Hz) clearly dominate the EEG (**Figure 3A**, red scissors; Steriade et al., 1993c; Steriade, 2001). Slow waves even occur in small cortical pieces *in vitro* (e.g., in the primary visual cortex) and are robust to various experimental perturbations (as reviewed in Sanchez-Vives et al., 2017). Therefore, they have been proposed to represent the default activity pattern of cortical networks. The UP states bear striking resemblance to the desynchronized firing patterns of the awake state (Destexhe et al., 2007). The transition between the two micro-states originates from activity-dependent adaptation, which accumulates during UP states, attenuates neural transmission, and eventually switches to the DOWN state. In this state of neural silence (or strongly reduced activity), the network recovers its excitability until it elicits the next sudden transition to the UP state, forced up and amplified by the firing rates of neighboring neurons (Compte et al., 2003; Destexhe and Contreras, 2006; Braun and Mattia, 2010; Mattia and Sanchez-Vives, 2012; Neske, 2015; Sanchez-Vives et al., 2017).

Delta oscillations consist of cortical and thalamic components that are generated in overlapping higher-order TC nuclei along with alpha oscillations at increasing inhibitory tone (Steriade et al., 1993c, 1994; Amzica and Steriade, 1997; Steriade, 2001; Flores et al., 2017). Thalamic delta rhythmogenesis results from the interplay between two distinct hyperpolarization-activated currents within TC neurons, whereas GABAergic RE cells provide the hyperpolarizing input under natural conditions (**Figure 3A**; McCormick and Pape, 1990; Soltesz et al., 1991; Steriade et al., 1994). Since TC cells are not synaptically coupled, they fire in frequency, but not necessarily in phase (Amzica and Steriade, 1997; Neske, 2015). Coupling of the clock-like thalamic delta patterns is mediated through interconnected cortical pyramidal cells, which generate bursts at similar intrinsic delta frequencies, thereby synchronizing cortical areas *via* recurrent TC projections (Amzica and Steriade, 1997; Neske, 2015). While delta power rises coherently in anterior areas, slow waves occur all over the cortex but not necessarily in phase, leading to cortical fragmentation at moderate doses and thus impaired information transmission (Lewis et al., 2012).

An essential component in generating slow and delta oscillations is the inhibition of subcortical arousal nuclei, which causes cortical and TC neurons to lose their depolarizing driving input (**Figure 3A**; Steriade et al., 1991, 1993a). An inhibition of key arousal nuclei is indeed sufficient to explain delta and slow wave generation in sleep (reviewed in Weber and Dan, 2016) and quiet wakefulness (Neske, 2015). However, the pattern characteristics of slow waves under natural conditions are distinct from those elicited by GABAergic anesthesia, which are





produced under additional direct thalamic and cortical inhibition (Chauvette et al., 2011; Kenny et al., 2014; Busche et al., 2015; Akeju and Brown, 2017; Arena et al., 2017; Aggarwal et al., 2019). The full inhibitory package results in enhanced suppression of spontaneous desynchronized cortical background activity, which causes successively quiescent networks in the DOWN state with increasing anesthetic depth.

Such quiescent cortical networks of synchronous activity have been proposed to be ideal substrates for the propagation of dense traveling waves (Figure 4; Muller et al., 2018). Given that all neurons act in phase, neurotransmission reaches all post-synaptic cells during episodes of increased excitability. Triggering UP states in this low complexity environment instantly recruits neighboring neurons *via* massive recurrent excitation (Sanchez-Vives et al., 2017), so that even local events (like sensory inputs) can elicit dense traveling waves that entrain nearly all cells as they pass (Muller et al., 2018).

Such dense, traveling slow waves have been found to spread across the cortical surface, irrespective of anatomical boundaries, like a set of waves created by a drop in an oil bath (Figure 4; Massimini et al., 2004, 2007; Luczak et al., 2007; Stroh et al., 2013; Sanchez-Vives et al., 2017; Muller et al., 2018). Whether these waves remain local or propagate depends on the (local) brain state or desynchronization (Massimini et al., 2004, 2007; Nir et al., 2011; Vyazovskiy et al., 2011; Schwalm et al., 2017; Muller et al., 2018) and hence on anesthetic depth.

Aside from functional deafferentation and cortical inhibition, which provides a powerful substrate for long-range spread by nearest-neighbor recruitment, the synchronizing drive from the thalamus may add to the manifestation of slow waves (Crunelli

and Hughes, 2010; Sheroziya and Timofeev, 2014; Crunelli et al., 2015; Neske, 2015). For the sake of translatability, functional brain studies should be generally performed at lower dosages than those leading to the dominance of slow wave and delta power, which affects neuronal excitability and stimulus–response properties (see “Anesthetic Depth and Consciousness—the Virtue of Translational Neuroimaging” section; Steriade et al., 1993b,c; Petersen et al., 2003; McGinley et al., 2015b).

Higher anesthetic dosages lead to *burst suppression*—characterized by increasingly prolonged “DOWN” micro-states alternating with periods of brief bursts of spikes and waves (Figure 2A; Steriade et al., 1994; Lewis et al., 2013; Amzica, 2015; Purdon et al., 2015). This is an increasingly hyperexcitable state in which large networks of cortical neurons suddenly discharge in tight synchrony (Steriade et al., 1994; Kroeger and Amzica, 2007; Ferron et al., 2009). Subsequently they fall into a post-burst refractory period, due to metabolic exhaustion (Hirsch and Taylor, 2010; Ching et al., 2012; Ching and Brown, 2014). This period of complete cortical silence is induced by adenosine triphosphate (ATP)-sensitive cation channels, which open when ATP decreases below critical levels (Cunningham et al., 2006). The resulting cation influx renders neurons unable to produce action potentials. Upon metabolic regeneration, the channels close again, and cortical networks gradually regain hyperexcitability by depolarizing, until they exhibit the next discharge (Kroeger and Amzica, 2007). Hyperexcitability is generated within the cortex as a result of the functional impairment of cortical afferents (Echlin et al., 1952; Henry and Scoville, 1952; Hughes, 1986; Niedermeyer et al., 1999; Kenny et al., 2014; Amzica, 2015).

Thalamic sensory transmission is preserved during isoelectric and burst episodes, although cortical discharge rates are substantially reduced during isoelectric periods (Detsch et al., 2002). During periods of cortical hyperexcitability, even subliminal stimuli are sufficient to trigger bursting activity (Yli-Hankala et al., 1993; Hartikainen et al., 1995; Hudetz and Imas, 2007; Kroeger and Amzica, 2007; Amzica, 2009). Spontaneous and evoked bursts can spread across large cortical areas (Steriade et al., 1993b; Timofeev et al., 2000; Land et al., 2012) and may further engage remote hippocampal (Sirota et al., 2003; Hahn et al., 2006; Ji and Wilson, 2007; Busche et al., 2015) and thalamic structures (Steriade et al., 1993c; Stroh et al., 2013; Sheroziya and Timofeev, 2015), presumably by excitatory projections (Leong et al., 2016). In this regard, burst suppression events can be considered as non- or quasi-periodic slow waves with a long refractory time and high discharge.

Further increasing the anesthetic dosage leads to longer suppression, shorter burst periods, and less reactivity to stimuli (Hartikainen et al., 1995), which culminates in complete neural inexcitability and finally isoelectricity due to increasing (thalamo)cortical and subcortical suppression (Figure 2A; Kroeger and Amzica, 2007).

### **$\alpha$ 2AR Agonists Induce a Sleep-Like State by Suppressing General Arousal**

Sympatholytics or  $\alpha$ 2AR agonists such as xylazine, medetomidine, or its potent dextro enantiomer



dexmedetomidine exert their sedating and antinociceptive effects primarily by acting on presynaptic  $\alpha_2A$  adrenergic receptors of noradrenergic cells that project from the locus coeruleus. This results in a hyperpolarization of the affected neurons and a reduction in the release of noradrenaline to their target sites (Correa-Sales et al., 1992; Jorm and Stamford, 1993; Chiu et al., 1995; Van Bockstaele et al., 1999). The locus coeruleus is a major arousal nucleus that projects to the basal forebrain (a subcortical arousal structure, which regulates cortical states by cholinergic efferents; Nelson et al., 2005; Hoover and Vertes, 2007; Pal et al., 2018), intralaminar nucleus of the thalamus, thalamic reticular nucleus, preoptic area of the hypothalamus, and diffusely into the cortex (**Figure 1C**; Asanuma, 1992; Nelson et al., 2003; Samuels and Szabadi, 2008a; Saper et al., 2010; Zhang et al., 2015; Fu et al., 2017; Brown et al., 2018a). Nociceptive pathways are further affected by the direct activation of  $\alpha_2$  receptors in the spinal cord (Andrieu et al., 2009).

Decreased noradrenaline release in the preoptic area of the hypothalamus causes disinhibition (and thus excitation) of local endogenous sleep-promoting cells that send inhibitory projections to other key arousal nuclei in the midbrain and pons (**Figure 1C**; Sherin et al., 1998; Saper et al., 2005; reviewed in Saper et al., 2010; Scammell et al., 2017). This causes an inhibition of widely-projecting neurons in these arousal nuclei, which in turn decreases the depolarizing input to thalamic and cortical areas, leading to sleep-like, spindle (9–16 Hz), and slow-delta (0.1–4 Hz) oscillations. This pattern is distinct from that induced by GABA<sub>A</sub> agonists due to the lack of direct cortical and thalamic inhibition (**Figures 2A–C, 3A**; Noreika et al., 2011; Baker et al., 2014; Nasrallah et al., 2014a; Akeju and Brown, 2017; Banks et al., 2017).

Similar sleep-like states were observed following a transection below the thalamus in cats, the so-called *cerveau isolé* (the isolated cerebrum; **Figure 3A**, green scissors; Bremer, 1935; Steriade et al., 1993c). This dramatically demonstrates that the disconnection of thalamic and cortical circuits from the input of subcortical arousal nuclei suffices to produce oscillatory and spiking patterns, similar to non-REM sleep at the N2 stage and sedation elicited by  $\alpha_2AR$  agonists (Akeju and Brown, 2017).

Under light sedation, spindle rhythms occur in brief bursts of ~0.5–3 s (**Figures 2A,C**; Baker et al., 2014; Nasrallah et al., 2014a; Purdon et al., 2015). Spindles are generated within the thalamic RE nucleus, which is considered the *spindle pacemaker* (Steriade et al., 1985; Halassa et al., 2011; Kim et al., 2012). Evidence for this includes the deafferentation of this structure from the cortex and the remaining thalamus, upon which the nucleus continues to generate spindles (Steriade et al., 1987). Rhythmogenesis thereby strongly depends on the level of hyperpolarization of thalamic RE neurons and the intactness of their long and thin dendrites, which are richly endowed with low-threshold hyperpolarization-triggered  $Ca^{2+}$  channels (as reviewed in Fuentealba and Steriade, 2005; Crandall et al., 2010; Astori et al., 2011; Zaman et al., 2011).

Suppression of arousal nuclei leads to decreased depolarizing input to RE cells (McCormick, 1992; Saper et al., 2010; Sun et al., 2013), which can cause hyperpolarization below the threshold, leading to strong dendritic  $Ca^{2+}$  spikes. This

results in depolarization and rhythmic bursts of action potentials that are sustained by additional voltage-gated cation channels (reviewed in Fuentealba and Steriade, 2005; Lüthi, 2014). GABAergic transmission and electrical coupling in recurrent networks of RE neurons are sufficient to synchronize oscillations in the range of spindles (Bazhenov et al., 1999; Fuentealba and Steriade, 2005).

In the intact brain, the initiation and synchronization of spindles is supported by TC and cortical circuits. Glutamatergic stimuli can easily trigger low-threshold  $Ca^{2+}$  spikes in RE cells (Crandall et al., 2010). Therefore, spindles can be initiated by diverse inputs, including spontaneous oscillating TC cells or cortical volleys that impinge on RE networks (**Figure 3A**; Destexhe et al., 1996). For example, the cortical transitions to the UP state of slow waves may be quickly followed by a resulting spindle wave, which is a common sequence in slow wave sleep, known as the *K-complex* (Steriade et al., 1993b; Amzica and Steriade, 1997). Notably, K-complexes have not been described for sedation through  $\alpha_2AR$  agonists (Huupponen et al., 2008; Nasrallah et al., 2014a; Akeju et al., 2016a).

A burst of action potentials in an RE neuron causes hyperpolarization, rebound spiking and thus neural firing in multiple TC neurons; this reverberates in large thalamic networks by recurrent synapsing (**Figure 3A**; Bazhenov et al., 1999; Beenhakker and Huguenard, 2009). The strong synchronized rhythms of TC neurons can further entrain cortical pyramidal cells and interneurons in both prefrontal and sensory cortices (Peyrache et al., 2011). Cortico-cortical recruitment may cause further synchronization (Kandel and Buzsáki, 1997). The short spindle episodes are terminated *via* intrinsic ionic mechanisms in both thalamic RE and TC cells that are triggered by high concentrations of accumulated intracellular  $Ca^{2+}$  (alongside other strategies reviewed in Lüthi, 2014).

Delta waves resemble spindles in that they do not appear continuously in sleep or for light  $\alpha_2AR$ -induced sedation (**Figure 2B**; Steriade et al., 1993a,c; Baker et al., 2014). This changes at higher anesthetic dosages through a further suppression of depolarizing arousal inputs to thalamic RE neurons; they further hyperpolarize, and their firing patterns subsequently change from spindle to delta waveforms (**Figures 2A,B**; Nuñez et al., 1992; Destexhe et al., 1994).

The transition to delta rhythms is mirrored in TC circuits and entrains cortical networks. The result is continuous delta oscillation in thalamic and cortical areas, whose onset has been reported to coincide with a discrete drop in frequency at the instant of dexmedetomidine-induced LORR (Baker et al., 2014). This is accompanied by a significant phase shift of delta waves in the central medial thalamus as compared to cortical areas. Such phase shifts in higher-order TC nuclei can disrupt cortico-thalamo-cortical communication, causing a breakdown of information transfer between cortical areas (Slézia et al., 2011; Saalman, 2014; Mo and Sherman, 2019; Suzuki and Larkum, 2020) and thus LORR.

First-order thalamic relay nuclei may not participate in producing continuous delta oscillations (Baker et al., 2014). Instead, they intensify spindle generation following LORR. The

transient inhibition of TC neurons during spindle periods has been reported to prevent them from transferring sensory information to the cortex (Steriade and Contreras, 1995; Fuentealba and Steriade, 2005). Accordingly, spindle density correlates with the gating of sensory inputs in sleep (Dang-Vu et al., 2010; Wimmer et al., 2012; Chen et al., 2016).

However, in  $\alpha 2$ AR-induced sedation, primary sensory routes remain largely intact. Medetomidine has been shown to preserve cortical responsiveness to acoustic stimuli in primary sensory areas in rats (Banks et al., 2017). Subcutaneous electrostimulation of the paw has been reported to evoke slightly reduced potentials in the primary somatosensory cortex (S1) compared to GABAergic drugs (Hayton et al., 1999; Li et al., 2003). Yet, the amplitudes of somatosensory evoked potentials in S1 are not affected with increasing concentrations of medetomidine (Li et al., 2003; Nasrallah et al., 2014a). Spindle activity in the paw region of S1 occurs at medetomidine concentrations that are commonly used for fMRI (Nasrallah et al., 2014a).

The induction of LORR at continuous delta activity has been demonstrated in a study that kept rats in a rotating tube in which they had to constantly adapt their position until they rolled onto their sides and remained supine (Baker et al., 2014). Such constant active behavior increases the activity of arousal nuclei (Marlinski et al., 2012; Furth et al., 2017) and interferes with the “sleep-like” sedative state elicited by  $\alpha 2$ AR agonists (Kamibayashi and Maze, 2000; Venn and Grounds, 2001). In fact, sedation induced by  $\alpha 2$ AR agonists is far more arousable than for GABAergic drugs (see “Sensory Processing and the Key Challenges in Murine fMRI” and “Functional Connectivity and Murine Resting-State fMRI” sections; Noreika et al., 2011; Sanders et al., 2012; Akeju and Brown, 2017; Banks et al., 2017). Thus, higher anesthetic dosages were required to induce LORR in actively moving rats (Baker et al., 2014) than to induce the sedation sufficient for an fMRI experiment (Nasrallah et al., 2014a). For the GABAergic drug propofol, LORR occurred with the onset of alpha waves both in rotating tubes (Baker et al., 2014) and in the freely resting rodent (Flores et al., 2017).

### NMDA Receptor Antagonists Primarily Affect Cortical Neural Activity

Ketamine is an NMDA receptor antagonist that provides dissociating, quasi-hypnotic effects by selectively binding and blocking NMDA receptors that are primarily expressed in cortical inhibitory and excitatory neurons, but also in thalamic TC and RE cells (Deleuze and Huguenard, 2016), and to a lesser extent in subcortical arousal nuclei and peripheral nerves (Gunduz-Bruce, 2009). Ketamine generates opposing effects in a dose-dependent manner. At lower concentrations, it binds preferentially to NMDA receptors on cortical GABAergic inhibitory interneurons, which show about 10-fold higher sensitivity to NMDA blockade than pyramidal neurons (Grunze et al., 1996).

NMDA receptor blockage prevents cation influx and depolarization, resulting in decreased GABAergic transmission to downstream excitatory neurons (Homayoun and Moghaddam, 2007; Seamans, 2008). Thus, excitatory neurons are

disinhibited and become depolarized (**Figure 1D**, Brown et al., 2011; Phillips et al., 2018; Picard et al., 2019). An appropriate depolarization of cortical pyramidal neurons can elicit fast oscillations in the lower gamma band (Steriade et al., 1991; Nuñez et al., 1992; Gray and McCormick, 1996), synchronized by rhythmic inhibition (Buzsáki and Chrobak, 1995; Olufsen et al., 2003; Börgers et al., 2005).

The fast rhythm generation is promoted by ketamine actions in thalamic circuits, which switches the firing patterns of TC and RE neurons from the burst mode to the tonic generation of single action potentials (Anderson et al., 2017; Mahdavi et al., 2019). Such synchronous fast oscillations in the gamma band (25–80 Hz) typically emerge with cortical processing during higher-level mental activity, and in REM sleep, which is associated with dreaming mentation (Llinás and Ribary, 1993).

Low-dose ketamine elicits gamma waves in a narrow frequency band of ~25–35 Hz and diffuse excitatory cortical activity (Akeju et al., 2016b). At this stage, hallucinations, dissociated states, euphoria, and dysphoria have been reported in clinical use: they have been attributed to preserved communication across brain areas at low inhibitory modulation and control, as well as a disruption of dopaminergic neurotransmission in the prefrontal cortex (Moghaddam et al., 1997; Purdon et al., 2015).

At higher doses, ketamine increasingly begins to block NMDA receptors at excitatory pyramidal neurons, causing cortical inhibition to predominate. It further suppresses arousal pathways by blocking excitatory projections from the parabrachial nucleus and from the medial pontine reticular formation in the brainstem to the thalamus and to the basal forebrain (Boon and Milsom, 2008; Fuller et al., 2011; Brown et al., 2018a). In turn, gamma waves become interspersed with slow-delta oscillations (**Figure 2B**; Ruiz-Mejias et al., 2011; Akeju et al., 2016b), which are augmented by direct drug action in the thalamus (Kiss et al., 2011; Zhang et al., 2012). The suppression of higher-order TC loops (Suzuki and Larkum, 2020) and a breakdown of cortical coherence is considered likely to be the mechanism that induces LORR (Pal et al., 2015; de la Salle et al., 2016; Schroeder et al., 2016; Brown et al., 2018a).

Ketamine can cause regional hypo- and hyperactivation across the cortex (Porro et al., 2004) and introduces further complexity to the processing of external stimuli (Oye et al., 1992; Zandieh et al., 2003; Schwertner et al., 2018). Besides antinociception mediated by direct inhibition of peripheral nociceptive afferents expressing NMDA receptors (Sinner and Graf, 2008), ketamine has also been shown to persistently reduce aversive responses to noxious stimuli in a top-down manner by prolonged suppression of hyperactive neurons in the anterior cingulate cortex (ACC) in rodent chronic pain models (Zhou et al., 2018). Low dosages increase SEP in rats and mice, even compared to the awake state (Franceschini et al., 2010; Michelson and Kozai, 2018) although the degree depends on the mouse strain (Maxwell et al., 2006). SEP from the cortex to higher-order TC neurons are decreased, illustrating a disturbed functional state of cortico-thalamo-cortical and thus cortico-cortical circuits (Anderson et al., 2017). Low-dose ketamine increases power in the gamma band, but also delta power

can increase more significantly than for low-dose isoflurane (Michelson and Kozai, 2018).

## Metastability of Brain States, Hysteresis, and Behavioral Monitoring

The comparison of brain states induced by various anesthetic compounds and dosages illustrates the diversity of mechanisms that lead to sedation. Whether an adjusted brain state can be stably maintained throughout the entire duration of an fMRI session remains to be clarified. Volatile ethers and intravenously injectable anesthetics appear to permit relatively stable maintenance of cortical oscillation patterns (Figures 2B,C; Purdon et al., 2013, 2015; Flores et al., 2017). However, transitions between two or more brain states have been observed over longer periods (1 h) at fixed concentrations of isoflurane in rats (Hudson et al., 2014). The transitions occurred predominantly in the lower-frequency band and could be well observed when the data were expressed on a logarithmic scale (Figure 2D).

The authors referred to this condition as “metastable” or potentially “multistable,” given that intermediate brain states shifted between two or more *attractors* distributed in phase space (for further discussion, see Breakspear, 2017; Hudson, 2017). There is evidence that such meta- or multistability is a general feature of brain states under anesthesia for various compounds and across species, although studies dedicated to detail such transitions over long periods are sparse (reviewed in Hudson, 2017). To what extent the transition of brain states may jeopardize the concordance of results across experiments (e.g., by exerting significantly different effects on signal processing) remains an open question, and has to be established for specific anesthetic protocols applied.

The reported state transitions were observed during recovery of consciousness; after an initial concentration of 1.75% isoflurane for 1 h, which reliably produced burst suppression, the concentration was reduced by 0.5%, maintained at that level for 1 h, and the process was repeated over a total of 6 h. Given this design, the dynamics of metastability were likely swayed by neural inertia—an intrinsic feature of neural circuits to resist swift transitions between consciousness and unconsciousness (Friedman et al., 2010; Proekt and Hudson, 2018; Proekt and Kelz, 2018). During recovery of consciousness neural inertia tends to trap the brain in an unconscious state. Due to this “stickiness,” a lower anesthetic dosage is required to maintain a similar anesthetic depth for the emergence from unconsciousness, compared to the induction of anesthesia (Hudson et al., 2014; Hudson, 2017). This dependence of the brain state on its history (hysteresis effect) cannot be explained by pharmacokinetic actions (see “Functional Connectivity and Murine Resting-State fMRI” section; Kelz et al., 2008; Friedman et al., 2010), and should be considered in preclinical studies for which an initial bolus induction of anesthesia is common practice.

Another issue in maintaining an intermediate brain state throughout an fMRI experiment might arise from nociceptive or stressful stimuli. Clinical experience and preclinical studies have shown that noxious stimuli, as well as direct stimulation of key

arousal nuclei, can shift cortical states from slow synchronized towards highly desynchronized oscillations (Figure 2E; Hudetz et al., 2003; Solt et al., 2014; Vazey and Aston-Jones, 2014; Akeju and Brown, 2017; Sanders, 2017; Pal et al., 2018; Hayat et al., 2019). When the activity of the arousal-related nuclei exceeds a certain level, the animal wakes up from light anesthesia, and falls back into sedation when the activity subsides (Figure 2E; Solt et al., 2014). Consequently, an anesthetic protocol should be tailored, in terms of class and dosage, to the experimental task at hand. If the aim of a study is to assess functional connectivity (FC) at rest, there is no need to adjust the anesthetic depth so that the mouse remains unresponsive to nociceptive stimuli.

This calls into question the idea of the minimum alveolar concentration (MAC) in rodent fMRI—a well-established behavioral measure of anesthetic depth for volatile ethers (Steffey, 2017). Briefly, a MAC of 1.0 is defined as the average of the lowest anesthetic concentration that prevents a behavioral response upon a standardized pain stimulus, and the highest concentration that still permits a nocifensive response in 50% of tested subjects (Eger et al., 1965; Quasha et al., 1980). A MAC of 0.7 in rats has been reported to suppress the righting reflex and active attempts to withdraw; at a MAC of 0.3 (“MAC-awake”), frequent movements of the snout including sniffing, chewing, licking, and gross limb movements still occurred (Hudetz, 2002). However, these behavioral markers are not dependent on the application of any nociceptive stimulus that is not applied in the actual experiment.

Anesthetic depth can be monitored by behavioral markers and pupil dilation diameter to assess the level of arousal (Eriskien et al., 2014; McGinley et al., 2015a; Joshi et al., 2016; Reimer et al., 2016; Binda and Gamlin, 2017; Shimaoka et al., 2018). Behavioral observation is indispensable to determine the level of sedation and immobilization of an animal, although the capacity to infer the actual brain state from these responses is limited (Pal et al., 2018). Hence, it is highly encouraged to adjust and compare anesthetic protocols based on oscillatory signatures and neural response properties outside the MR environment. Behavioral monitoring becomes increasingly crucial as different anesthetic classes and combinations of drugs are employed to sedate the animal (see “Multimodal Anesthesia in Translational fMRI” section).

## Anesthetic Depth and Consciousness—the Virtue of Translational Neuroimaging

In general, anesthetic mechanisms are highly conserved across species (Achermann and Borbély, 1997; Steriade et al., 2001; Mölle et al., 2002; Eschenko et al., 2006; Destexhe et al., 2007; Buzsáki and Moser, 2013; Shein-Idelson et al., 2016) and the oscillatory signatures of humans and mice are virtually similar upon applications of specific anesthesia and stages of anesthetic depth (Seth et al., 2005; Flores et al., 2017; Guidera et al., 2017; Hudson, 2017; Storm et al., 2017; Olcese et al., 2018). This relation provides a rough means with which to interpret brain states in terms of perceived anesthetic depth, stress, or pain. In humans, specific brain states can be linked to the introspective experience, as revealed by subjective reports. Using oscillatory signatures, a third-person observer can precisely specify the



moment at which the “lights switch off,” from the first-person perspective (Purdon et al., 2015; Brown et al., 2018b). For GABAergic drugs, this is the emergence of slow-delta oscillations (**Figure 2C**; Purdon et al., 2013, 2015).

Traditionally, this transition point is considered to be the “loss of connected consciousness” (LOC; Sanders et al., 2012) and is defined as a loss of sensory perception and interoception—the perception of internal processes, including pain, anxiety, stress, discomfort, or sense of time (Purdon et al., 2013, 2015; Warnaby et al., 2016; Sleigh et al., 2018; Chamadia et al., 2019). This is distinguished from “disconnected” phenomenal consciousness, like the awareness of pure darkness (Sanders et al., 2012; Sleigh et al., 2018) or dreaming, although dreams have often been reported for medetomidine-induced LOC (Akeju and Brown, 2017; Mashour and Hudetz, 2017, 2018).

In the state of propofol-induced LOC, innocuous sensory stimuli such as words or tones could no longer be perceived, and no hemodynamic responses to these stimuli were detected using fMRI (Ní Mhuircheartaigh et al., 2013; Warnaby et al., 2016; Lichtner et al., 2018). However, BOLD patterns in response to nociceptive stimuli were partially preserved (Lichtner et al., 2018). These findings imply that the oscillatory signatures associated with LOC may be useful as a marker delineating the lower limit of anesthetic depth in sensory fMRI (**Figures 2A,C**).

Another transition point can be determined based on behavioral observation—the “loss of behavioral responsiveness” (LOR), which occurs earlier, approximately with the appearance of frontal alpha oscillations and before slow waves dominate the EEG (**Figure 2C**; Purdon et al., 2013; Warnaby et al., 2016). At this point, subjects fail to exhibit volitional responses to sensory or even noxious stimuli (Sanders et al., 2012; Ní Mhuircheartaigh et al., 2013; Purdon et al., 2013; Warnaby et al., 2016; Sleigh et al., 2018). Stimuli are still perceived and processed when they are applied, as indicated by fMRI data and subjective reports (Warnaby et al., 2016). Volunteers described this as a state of detachment from the stimuli and from “themselves”—i.e., they were not fully unconscious, but also not aware that the stimuli were related to them. Tones that were presented rather subtly and very briefly (1 kHz, 60 ms) were not sufficient to elicit significant BOLD responses beyond the thalamic relay nuclei. The BOLD patterns evoked by words and nociception, on the other hand, were reported to reflect closely those of the conscious state, with substantially lower activity in only the right dorsal anterior insular cortex (Warnaby et al., 2016). This area is reportedly associated with body ownership and self-agency (Warnaby et al., 2016; Lichtner et al., 2018; Sleigh et al., 2018). The anesthetic inhibition of the anterior insula and related interoceptive networks led the authors to reformulate their understanding of the hypnotic effects of anesthesia as a gradual disruption of “selfhood,” which occurs with an increase of anesthetic depth, finally leading to oblivion (Sleigh et al., 2018).

This offers not only a direct link between introspective reports, behavioral and neuroimaging markers, but also a vivid example of how the examination of large-scale networks using fMRI can complement models of LOC that are based

on the suppression of feedback or feedforward information transfer along the cortical hierarchy (see “GABA<sub>A</sub> Agonists Suppress Neural Activity Across the Central Nervous System” section; reviewed in Pennartz et al., 2019; Sikkens et al., 2019). Because similar transitions in oscillatory signatures can be observed in humans and rodents, maintaining and stabilizing a brain state between LOR (corresponding to LORR in animals; Baker et al., 2014; Banks et al., 2017; Flores et al., 2017) and LOC (corresponding to LOM in animals, and the occurrence of slow-delta waves; Flores et al., 2017) could be an attractive target state in which to perform fMRI in mice, at least for sensory perception tasks (**Figures 2A,C**). Hemodynamic coupling does not appear to be directly affected by propofol (Veselis et al., 2005). However, potential vasomodulatory effects of anesthesia have to be taken into account when inferring dose-dependent suppression of neural activity based on fMRI (**Table 2**; Aksenov et al., 2015).

## MURINE fMRI AND HEMODYNAMIC INTEGRITY

Research on murine fMRI has grown exponentially from only a handful of publications in the first decade (Huang et al., 1996; Ahrens and Dubowitz, 2001; Mueggler et al., 2003; Xu et al., 2003, 2005) to more than 50 in the following. Nevertheless, anesthesia in murine fMRI presents a multitude of challenges beyond the determination of functional brain states and neural response properties, and anesthetic protocols are still far from being effectively tailored to meet murine physiological and hemodynamic demands. Preserving hemodynamic integrity that approximates a physiological state is essential to exploit the two key advantages of murine fMRI: non-invasively investigating whole-brain functional dynamics, and the opportunity to directly compare these dynamics across species, including humans. The experiments that have been performed can be classified into at least four categories: (1) sensory perception; (2) nociception and pain; (3) FC based on the resting state; and (4) within-brain stimulation involving the use of opto- and chemogenetics. Each category reveals particular aspects of anesthetic effects in fMRI and deserves consideration in terms of the requirements for anesthetic protocols that are appropriate to obtaining neuroimaging data of translational value.

### Sensory Processing and the Key Challenges in Murine fMRI

Sensory studies were among the earliest and most elementary applications of murine fMRI. Nevertheless, only a handful of publications report on innocuous sensory stimulation tasks that address the “natural” sensory perception routes of primary organs—like eyes, ears, nose, skin, or whiskers—using various anesthetic protocols. The very first mouse fMRI study (Huang et al., 1996) concerned visual perception and applied the GABA<sub>A</sub>-positive allosteric modulator and agonist pentobarbital; subsequent olfactory studies used urethane (Xu et al., 2003, 2005). Later visual (Niranjan et al., 2016) and auditory



(Blazquez Freches et al., 2018) studies applied the  $\alpha 2$ AR agonist medetomidine, and deflection of the vibrissae was conducted under low-dose (0.5–1%) isoflurane (Kahn et al., 2011). All these studies identified modality-relevant sensory pathways, including primary sensory and partially preserved secondary cortical and thalamic structures. Higher-order or association areas have not been reported.

### Technical Detection Limit

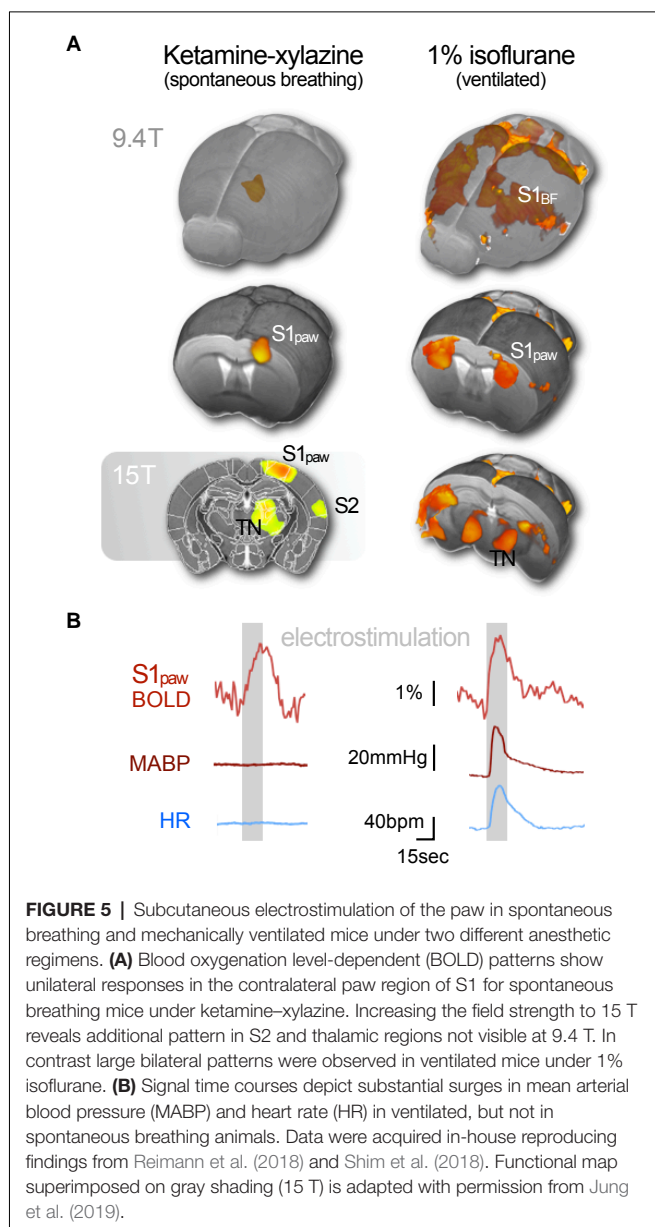
While this is in principle an encouraging situation, the reality of implementing mouse fMRI in the laboratory is fraught with peril. Sensory fMRI requires meticulous fine-tuning of murine physiology to compensate for anesthetic side effects, in conjunction with advanced technical equipment to boost the BOLD effect, which increases linearly with the magnetic

field strength and venous blood volume (Kim, 2018). A recent study vividly illustrated how the field strength determines whether or not activation of entire brain areas may be detected. Subcutaneous electrostimulation of the murine paw is expected to elicit activity in the contralateral ventral thalamic nuclei, which relay the signal along the spinothalamic tract to the primary (S1) and secondary (S2) somatosensory cortices. Nevertheless, BOLD responses to electrostimulation were found to be limited to the contralateral S1 in mice under ketamine–xylazine at 9.4 T (Figure 5A; Shim et al., 2018), and under medetomidine at 9.4 T (Nasrallah et al., 2014a) and 11.7 T (Adamczak et al., 2010). However, the same ketamine–xylazine protocol at 15 T revealed the expected BOLD patterns in all three somatosensory key areas (Figure 5A; Jung et al., 2019).

To boost the BOLD signal, small-bore MR scanners can be equipped with a cryogenically cooled radiofrequency coil that amplifies the temporal SNR by a factor of up to 3 (reviewed in Niendorf et al., 2015). Increasing the statistical power by multiple repetitions of a task similarly facilitates the detection of weak BOLD effects. Both strategies reduce noise, but do not increase the sensitivity to  $T_2^*$ —the measure of relative changes in blood oxygenation, which determines the intrinsic technical detection limit of an MR system. To amplify a signal, it must be above the detection threshold. However, weak BOLD responses have been observed in the S2 and the thalamus at 9.4 T (Jung et al., 2019), strongly suggesting that these areas can be detected *via* signal amplification. The studies reported above employed MR surface coils that lose sensitivity for deeper brain areas. In addition to this technical issue, the low BOLD signal in deeper brain areas is likely physiological in nature, and may be caused by the sparser overall thalamic cell density, compared to the cortex (Meyer et al., 2013), by different densities of noradrenergic afferents in these areas (King et al., 1995; Wang et al., 1996) or otherwise asymmetric impact of reduced subthalamic drive, by unequal anesthetic effects on the vasculature in different brain regions (Lei et al., 2001), or by the experimental setup that can create a thermal gradient between the dorsal–ventral axis of the mouse brain (see below).

### Vasomodulation, Body Temperature, and Respiratory Depression

The modulation of vascular properties across the brain is a key challenge for fMRI under anesthesia—even more so due to the temperature sensitivity of vessels. All anesthetics affect thermoregulation and render mice poikilothermic. Due to their high surface area-to-mass ratio, mice quickly adopt to their environmental temperature. Although blood circulation distributes heat energy across the body, heating pads below the trunk and cool surface MR units attached to the scalp may induce thermal gradients across the mouse, which means the brain temperature is neither homogeneous across all areas, nor accurately reflected by the temperature reading of a rectal probe (Reimann et al., 2016). Aside from a thermal dependency of neural activity that has been shown to alter brain states, and hence anesthetic depth (Volgushev et al., 2000; Reig et al., 2010; Sheroziya and Timofeev, 2015; Schwalm and Easton, 2016),



**TABLE 2** | Effects on hemodynamic integrity of anesthetics used in mouse fMRI.

	hemodynamic effects				
	CBF	CMRO <sub>2</sub>	ICP	AutoReg	CO <sub>2</sub> Reg
propofol	↓↓	↓↓	↓↓	—	↓
etomidate	↓↓	↓↓	↓↓	—	↓
benzodiazepine	↓	↓	↓	—	
barbiturate	↓↓	↓↓	↓↓	—/↓	↓
isoflurane	—/↑	↓↓	—/↑	—/↓	↑
sevoflurane	—	↓↓	—	↓	↑
halothane	↑↑	↓	↑↑	—	↑
medetomidine	↓	—	↓	—	—
xylazine	↓	—	↓	—	—
ketamine	↑	—	↑	↓	

Anesthetics are listed according to their classes: GABA<sub>A</sub> agonists, including injectable drugs (red) and volatile ethers (yellow),  $\alpha$ 2AR agonists (green), and NMDA antagonists (blue). Vascular effects including vasodilation and vasoconstriction may impact on cerebral blood flow (CBF), intracranial pressure (ICP), or cerebral autoregulation (AutoReg). Drugs may further affect the cerebral metabolic rate of oxygen consumption (CMRO<sub>2</sub>) and the regulation of the CO<sub>2</sub> metabolism (CO<sub>2</sub>Reg). Effect: ↓ reduction; ↓↓ strong reduction; ↑ increase; ↑↑ strong increase;—no change. Based on Bittner et al. (2019).

vessels dilate with increasing temperature. Vasoconstriction induced by cooling the scalp can boost cortical BOLD responses to electrostimulation by about 30%, although this is not physiological (Baltes et al., 2011).

Many anesthetics directly affect vascular dilation and reactivity, which can substantially alter baseline CBF and hemodynamic responses (see **Table 2**, Franceschini et al., 2010; Masamoto and Kanno, 2012). In contrast to cooling, vasoconstrictive effects of  $\alpha$ 2AR such as medetomidine and xylazine (Sinclair et al., 2003; Fukuda et al., 2013) might hinder the thorough vasodilation of arterioles in response to neural activity and impair BOLD responses in mice (Nasrallah et al., 2014a; Schroeter et al., 2014). Anesthetic-induced vasodilation, on the other hand, increases the baseline blood flow and hence decreases the magnitude of relative changes in CBV, CBF, and BOLD in response to neural activity. Isoflurane is commonly reported to have strong vasodilatory effects (Sharp et al., 2015; Petrinovic et al., 2016; Cao et al., 2017), causing decreases of up to 75% in CBF responses to whisker stimulation at a concentration of only 1% in spontaneously breathing mice (Takuwa et al., 2012). Although isoflurane has been shown to directly dilate cerebral vessels (Iida et al., 1998; Leoni et al., 2011), its major vasodilatory potency is a result of its suppressive effects on respiration in a dose-dependent manner (van Alst et al., 2019).

Respiratory depression leads to hypercapnia—elevated blood CO<sub>2</sub> levels—which causes strong cerebral vasodilation and increases the baseline CBF (Kety and Schmidt, 1948; Raper and Levasseur, 1971; Shimosegawa et al., 1995; Jones et al., 2005; Leoni et al., 2011; van Alst et al., 2019). Both rats and mice exhibit massively reduced or even fully ablated BOLD

responses, when CO<sub>2</sub> (5–10%) is added to the inspiratory gas (Schlegel, 2017; Munting et al., 2019; van Alst et al., 2019). Mechanically ventilated rats showed substantially higher BOLD responses to subcutaneous electrostimulation, and lower baseline CBF compared to spontaneously breathing rats at the same concentrations of isoflurane (van Alst et al., 2019). A comparison of cortical and thalamic baseline perfusion across anesthetic protocols in mechanically ventilated mice found isoflurane to be in a similar range as urethane, which is known to preserve hemodynamics and vascular diameters (Schroeter et al., 2014).

Hypercapnia has further effects on brain states. Since CO<sub>2</sub> forms carbonic acid in water, elevated CO<sub>2</sub> makes the blood more acidic. Both low pH and CO<sub>2</sub> have been shown to impair neural transmission and excitability (Coulter et al., 1995; Tombaugh and Somjen, 1996; Sun et al., 1997; Meuth et al., 2003, 2006; Putnam et al., 2004; Williams et al., 2007; Sinning and Hübner, 2013). Hyperoxia, on the other hand, has been reported to cause vasoconstriction of arterioles (Duling and Berne, 1970; Pries et al., 1995) and to decrease baseline CBF (Matsuura et al., 2001; Matsuura and Kanno, 2002). This demonstrates that blood gas levels are crucial in fMRI and that respiration should be kept at physiological levels to preserve neural and especially hemodynamic responses.

### Mechanical Ventilation and Stress Response

This has led to the widespread use of mechanical ventilation in murine fMRI, initially to detect highly reproducible BOLD responses to *salient* (“attention grabbing”) stimuli, like subcutaneous electrostimulation (Bosshard et al., 2010; Baltes et al., 2011), and later nociceptive heat stimuli (see “Nociception, Pain, and Physiological Confounds” section; Bosshard et al., 2015; Reimann et al., 2016). Mechanical ventilation was further introduced as part of a robust protocol for pharmacological fMRI in mice (Ferrari et al., 2012) and is utilized in resting-state fMRI (see “Functional Connectivity and Murine Resting-State fMRI” section; Grandjean et al., 2014). The commonly applied protocol features 1.0–1.3% isoflurane in combination with neuromuscular blockage to minimize motion-related noise and prevent stimulus-correlated movements. An application of classical electrostimulation block paradigms to the murine paw—stimulus trains of 15–30 s composed of rectangular pulses of 0.5–2 mA—revealed that this protocol leads to BOLD signal changes of substantial magnitudes, and a very specific pattern of distribution: BOLD effects occurred in bilateral clusters in S1, S2, the thalamus, and the insular cortex (**Figure 5A**; Bosshard et al., 2010; Baltes et al., 2011; Schroeter et al., 2017; Reimann et al., 2018; Schlegel et al., 2018; Shim et al., 2018), and were accompanied by strong, transient elevations in mean arterial blood pressure (MABP) and heart rate (HR; **Figure 5B**; Schroeter et al., 2014; Reimann et al., 2018). This was confirmed for various anesthetic protocols including low-dose isoflurane, medetomidine, propofol, and urethane in ventilated and paralyzed mice (Schroeter et al., 2014).

The same stimulation paradigm applied to spontaneously breathing mice evoked less prominent BOLD responses that were strictly limited to the paw region of the contralateral S1 as described above (**Figure 5A**; Adamczak et al., 2010;

Nasrallah et al., 2014a; Shim et al., 2018). This was observed for medetomidine and ketamine–xylazine anesthesia. The additional patterns observed in contralateral thalamic relay nuclei and S2 at higher field strengths emphasize the somatosensory nature of the response along the spinothalamic tract (Jung et al., 2019). Weak responses in ipsilateral S1 have also been reported for some animals, likely due to inter-hemispheric projections (Schroeter et al., 2017; Shim et al., 2018; Jung et al., 2019), and no cardiovascular changes were observed during the stimulus periods in these studies (Figure 5B).

In contrast, the strong, transient cardiovascular surges in ventilated animals clearly indicate sympathetic activity, which is governed by medullary control areas and triggered by an autonomic stress response (Pfaff, 2005; Samuels and Szabadi, 2008b; Ulrich-Lai and Herman, 2009). Such a response involves activity of the locus coeruleus and other key nuclei of the brainstem and the forebrain arousal system, which innervate multiple cortical and subcortical areas (Figure 1A; Toussay et al., 2013; Lecrux and Hamel, 2016; Lecrux et al., 2019).

The specific reason for this strong sympathetic response in ventilated mice is not fully understood. It is known that endotracheal intubation and forced ventilation can induce excessive stress in humans and animals when the anesthetic depth is too low. We hypothesize that long stimulus trains in ventilated mice accumulate to engage the activity of subcortical arousal nuclei that drive the brain state gradually toward desynchronization (see “Metastability of Brain States, Hysteresis and Behavioral Monitoring” section; Hudetz et al., 2003; Solt et al., 2014; Vazey and Aston-Jones, 2014; Pal et al., 2018; Hayat et al., 2019). Elevated sympathetic activity causes a tightly coupled increase in cardiovascular and respiratory output (Pfaff, 2005; Ulrich-Lai and Herman, 2009). Yet, in ventilated mice, the respiration rate is fixed (to 80–90 bpm), which may lead to an allosteric mismatch (“prediction error”) between autonomic command and sensory feedback (Alheid and McCrimmon, 2008; Kleckner et al., 2017). It is unclear whether this drives the animal’s stress response. However, it appears that ongoing electrostimulation triggers an excitatory feed-forward loop that forces increased activity in reticulo-thalamo-cortical circuits, and causes bilateral patterns in the S1 barrel fields, and strong sympathetic outflow (Toussay et al., 2013; Lecrux and Hamel, 2016; Lecrux et al., 2019).

With isoflurane, this is a transient process, and mice usually fall back into stable sedation after a stimulation period, which is evident from HR and MABP traces (Schroeter et al., 2014; Reimann et al., 2018) as well as from behavioral observations in unparalyzed mice. Mice administered medetomidine while under ventilation have been shown to exhibit increases in their HRs progressively over the length of the fMRI scan following the first stimulation period (Schroeter et al., 2014). Medetomidine exerts its effects predominantly by inhibiting the locus coeruleus (Akeju and Brown, 2017). Once a competing mechanism engages this key arousal nucleus, the sedation is temporarily suppressed (Hayat et al., 2019), and the stimuli might be processed in the awakening, largely unaffected cortex (see “ $\alpha$ 2AR Agonists Induce a Sleep-Like State by Suppressing General Arousal” section). Again, medetomidine produces a state from which arousal is

more likely to occur than through the use of GABAergic drugs, which suppress the activity of excitatory neurons across the entire brain (Figures 1B,C).

Mechanically ventilated mice anesthetized with medetomidine exhibited bilateral BOLD responses even to single electrical pulses applied to the paw (1 mA, 0.5 ms; Schlegel et al., 2015). The same single electrical pulses applied to ventilated mice under isoflurane, propofol, or urethane evoked BOLD responses contralaterally along the spinothalamic tract in S1, S2, and first-order TC nuclei at 9.4 T—very similar to the patterns observed for stimulus trains in spontaneously breathing mice under ketamine–xylazine anesthesia at 15.2 T (Figure 5; Jung et al., 2019), with the latter being more spatially defined. The greatest similarities have been reported for patterns elicited under propofol along with profound hemodynamic responses to electrostimulation. However, the sedation under propofol was found to be unstable for the adjusted dosage (Schroeter et al., 2014; Schlegel et al., 2015). In summary, for the tested regimens, single pulses of electrical stimulation in mechanically ventilated mice elicit patterns predominantly along the somatosensory axis, whereas long pulse trains evoke bilateral patterns that likely reflect a transient recruitment of arousal structures, which can drive the brain state towards consciousness.

Multiple publications have reported on subcutaneous electrostimulation of the paw in mechanically ventilated mice under isoflurane. Whether this protocol allows for the detection of non-salient, low-intensity sensory stimuli has not yet been investigated. There is evidence that stimulus trains in the somatosensory range (0.5 mA) lead either to bilateral patterns or to no response at all (Shim et al., 2018). For a free-breathing sedation adjusted at 0.5–1% isoflurane, significant BOLD responses to whisker deflection have been reported in a murine fMRI study (Kahn et al., 2011). In such lightly sedated mice, tracheal intubation and forced ventilation would act as heavy continuous stressors during an fMRI experiment. Accordingly, mechanical ventilation does not permit further decreases in isoflurane concentrations, and approximately 1–1.3% are advised to prevent panic and withdrawal from the endotracheal tube. Concentrations of 0.75–1.1% isoflurane were found necessary to avoid burst suppression and permit stable sedation in spontaneously breathing mice (Kozai et al., 2015; Michelson and Kozai, 2018). To create conditions as close as possible to those experienced by a calm human volunteer, alternative strategies to ventilation should be considered for sensory tasks.

### Balancing Anesthesia in Spontaneously Breathing Mice

Increasing oxygen concentration in the carrier gas is a popular strategy to prevent hypoxemia in spontaneously breathing mice, although caution is advised: no differences in respiratory rates, pH, blood CO<sub>2</sub> levels, or blood pressure have been found for mice breathing isoflurane in pure oxygen (100%) compared to medical air (21%). This has been explained by a decrease in lung volume for pure oxygen based on alveolar collapse, known as *absorption atelectasis* (Wilding et al., 2017). Hemodynamic responses to whisker stimulation were significantly reduced for mice breathing pure oxygen compared to medical air (Sharp



et al., 2015). The exact concentration of oxygen in the breathing gas may be adjusted based on a bell-shaped stimulus–response curve with respect to the anesthetic protocol that is applied (Blazquez Freches et al., 2018).

A good strategy is to apply anesthesia that causes neither respiratory depression nor vasodilation. This makes  $\alpha 2$ AR agonists attractive to assess sensory processing in murine fMRI, preserving physiological breathing rates at about 120–190 bpm (Adamczak et al., 2010; Nasrallah et al., 2014a). However,  $\alpha 2$ AR agonists induce a sleep-like state and impair BOLD responses due to vasoconstriction (see above; Fukuda et al., 2013; Nasrallah et al., 2014a), which further initiates a transient increase in MABP followed by a reduction based on suppression of noradrenergic sympathetic ganglia (McCallum et al., 1998; Samuels and Szabadi, 2008a). A better choice might be the combination of two or more complementary anesthetics, to balance their respective actions and to tailor the desired effect (Fukuda et al., 2013). Combining the  $\alpha 2$ AR agonist xylazine together with ketamine preserves physiological breathing rates at about 180 bpm and leads to pronounced BOLD responses, at least in the contralateral S1 (Shim et al., 2018). Another promising alternative was introduced by balancing isoflurane anesthesia with fentanyl-fluanisone and midazolam, which allows the isoflurane level to be reduced to 0.5–0.8% (Sharp et al., 2015). This was reported to induce stable sedation and permit the detection of hemodynamic effects for whisker stimulation, with magnitudes and transition times similar to those observed in awake mice.

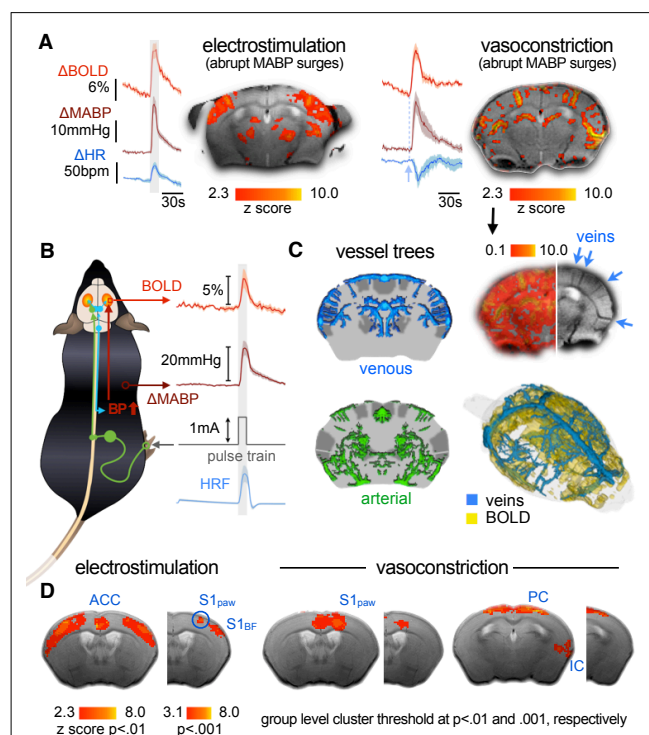
Balanced multimodal anesthesia holds great promise to circumvent the multiple challenges in murine fMRI, with the aim of preserving physiological conditions while sufficiently sedating the animal, and still ensuring a certain level of “connectedness” to sensory perception and interoception (see “Anesthetic Depth and Consciousness—the Virtue of Translational Neuroimaging” section; Sanders et al., 2012). Protocols for balanced multimodal anesthesia will be further discussed in “Multimodal Anesthesia in Translational fMRI” section.

## Nociception, Pain, and Physiological Confounds

Nociception is “the neural process of encoding noxious stimuli.” Pain is defined as an unpleasant *experience* that may or may not arise from nociception (Davis et al., 2017). In contrast to fMRI tasks that probe the central processing of applied stimuli, fMRI of pain explicitly aims to identify neural correlates of a phenomenal quality (Mouraux and Iannetti, 2018). Its assessment is therefore hampered in the (sedated) rodent, which cannot report on the experience or level of pain during an fMRI task (Seth et al., 2005). This is an important domain for translational fMRI: in principle, it should be possible to first correlate nociceptive-evoked BOLD patterns with subjective ratings of pain to identify pain-specific functional neurosignatures in the human brain (Wager et al., 2013; Woo et al., 2017). In a second step, a corresponding signature could be sought in the sedated animal model (Tracey, 2017). This leads to an even more fundamental question in terms of anesthesia: to what extent is consciousness required to encode a pain-specific neurosignature?

Currently, we cannot answer this question, since a pain-specific neurosignature has not yet been found (Mouraux and Iannetti, 2018). Painful stimuli are processed within a widely distributed network that is often referred to as the “pain matrix” (Legrain et al., 2011)—a term derived from Melzack’s original *neuromatrix theory of pain* (see below; Melzack, 1989, 2005). It has been shown that salient and painful stimuli are processed by largely overlapping cortical areas in humans (Mouraux et al., 2011), and subcortical and arousal structures that are involved in nociception are also engaged by acute responses to stress (Reimann et al., 2016; Martins and Tavares, 2017). In other words, the neuromatrix is not pain-specific. This is a major obstacle for identifying pain-specific brain signatures, because painful stimuli are: (1) intrinsically salient (Legrain et al., 2011; Mouraux et al., 2011); and (2) often provoke autonomic stress responses (Reimann et al., 2016, 2018).

Such stress responses introduce severe complications in murine fMRI. They involve the activity of the



**FIGURE 6 |** BOLD patterns evoked by abrupt surges in MABP.

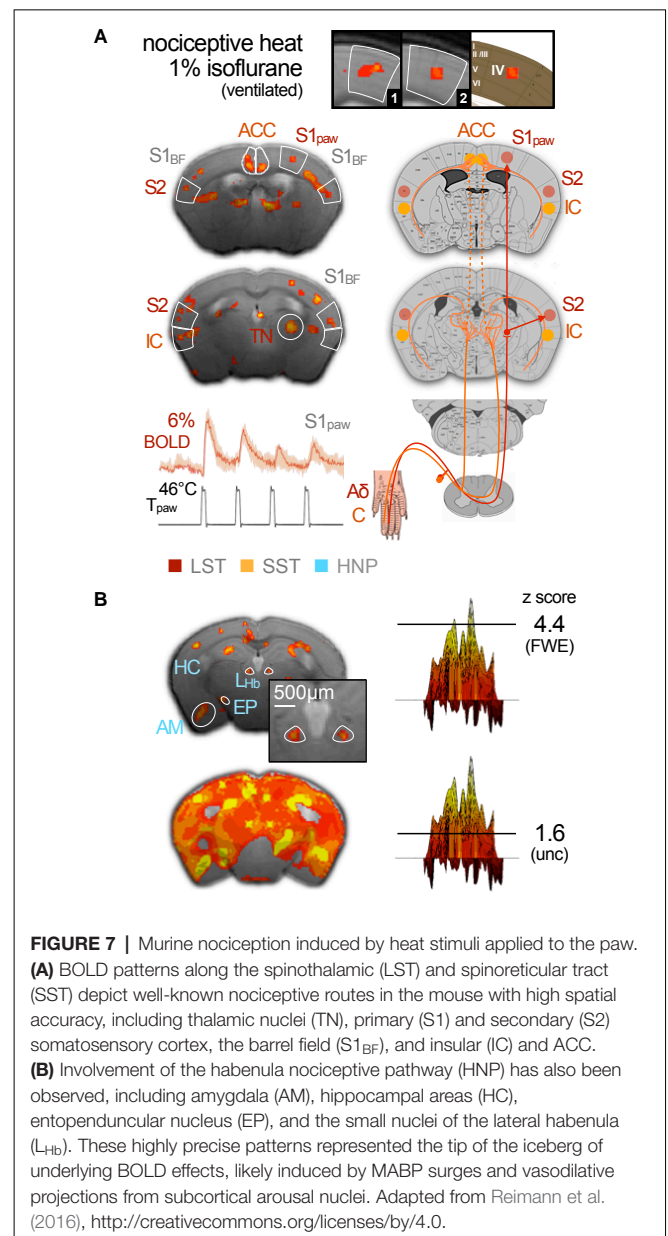
(A) Electro-stimulation evokes abrupt surges in MABP and HR. Mimicking these surges by intravenous injection of vasoconstricting drugs leads to BOLD patterns co-localizing with large veins. Injection time point is indicated by dashed line. (B) Nociceptive stimuli or stress can elicit abrupt surges in MABP that increase the influx of oxygenated blood into the brain vasculature. MABP and BOLD patterns are highly correlated with each other and with the applied stimulus. This causes the hemodynamic response function (HRF)—that is modeled based on the stimulus paradigm—to reveal MABP-induced effects as significant patterns. (C) Significant patterns co-localize with large veins and can therefore be identified and corrected for, whereas widespread patterns remain below the statistical threshold. (D) At the group level, significant effects persist at more liberal thresholds, and few minor clusters even at more conservative standards. Adapted with permission from Reimann et al. (2018).



nociceptive–medullary–autonomic circuit—a fundamental component of the “flight or fight” response, which triggers abrupt surges in HR and MABP (**Figure 6A**; Price, 2000; Pfaff, 2005; Brown et al., 2010). Since cerebral autoregulation has high-pass filter characteristics, slow changes in MABP are buffered to a certain extent, whereas more pronounced and rapid changes are reflected in the CBF (Hamner et al., 2019). The efficiency of dynamic autoregulation might be further impaired by vasomodulatory effects of anesthesia (Lee et al., 1994; Dagal and Lam, 2009; Sanders et al., 2011). Abrupt surges in MABP, well within the physiological range (80–120 mmHg in mice), can lead to stimulus-correlated increases in CBF causing widespread BOLD patterns across the brain (**Figures 6B,C**; Kalisch et al., 2001; Wang et al., 2006; Gozzi et al., 2007; Reimann et al., 2018). Veins are particularly prone to translate changes in CBF, into pronounced BOLD effects due to their low oxygen concentration (Reimann et al., 2018). In fact, significant clusters were observed along large draining veins (cluster threshold  $z = 3.1$ ) that were highly correlated with transient MABP surges induced by methods of controlled pharmacological vasoconstriction (**Figures 6A,D**). Since large veins co-localize with key cortical regions of the neuromatrix (S1, S2, insula, and parietal cortex), this is a confounding factor in current murine fMRI studies of nociception. Aside from strategies to identify and correct the affected regions, fMRI protocols that track changes in CBV, such as vascular space occupancy (VASO) or iron-oxide-based CBV-fMRI, are less prone to such large-vein effects (for a review, see Huber et al., 2017) and can be used either as alternative techniques, or to normalize and correct BOLD data.

Further complexity is introduced by outreaching vasodilatory projections of specific arousal nuclei such as the locus coeruleus, or the nucleus basalis of Meynert in the basal forebrain that are thought to have a substantial impact on the cortical hemodynamic responses to noxious stimuli (Lecrux and Hamel, 2016; Uchida and Kagitani, 2018; Paquette et al., 2019). The extent to which these BOLD responses actually reflect neural activity remains unclear (Lecrux et al., 2019). Both abrupt elevations in MABP and remote vasodilatory projections of arousal nuclei can produce hemodynamic readouts that are indistinguishable from neurovascular coupling. This problem has hardly been explored and is currently considered a major hurdle for interpreting experimental data related to nociceptive fMRI in (ventilated) rodents (Paquette et al., 2018).

Despite these considerations, nociceptive pathways are largely preserved in murine BOLD fMRI using low-dose mono-anesthetic isoflurane anesthesia in mechanically ventilated and paralyzed mice (Reimann et al., 2016). Brief heat stimuli just above the A $\delta$  nociceptor threshold (46°C) applied to the murine paw evoked BOLD effects in brain areas including the spinothalamic and spinoreticular tract, which have been well documented as having a functional involvement in murine and human nociception (**Figure 7A**). Employing a cryogenic transmit-receive unit (Niendorf et al., 2015) at 9.4 T has permitted resolving even the small nuclei of the habenular nociceptive pathway (**Figure 7B**, Reimann et al., 2016)—a pain modulating circuit known from human fMRI



**FIGURE 7 |** Murine nociception induced by heat stimuli applied to the paw. **(A)** BOLD patterns along the spinothalamic (LST) and spinoreticular tract (SST) depict well-known nociceptive routes in the mouse with high spatial accuracy, including thalamic nuclei (TN), primary (S1) and secondary (S2) somatosensory cortex, the barrel field (S1<sub>BF</sub>), and insular (IC) and ACC. **(B)** Involvement of the habenula nociceptive pathway (HNP) has also been observed, including amygdala (AM), hippocampal areas (HC), entopenduncular nucleus (EP), and the small nuclei of the lateral habenula (L<sub>hb</sub>). These highly precise patterns represented the tip of the iceberg of underlying BOLD effects, likely induced by MABP surges and vasodilative projections from subcortical arousal nuclei. Adapted from Reimann et al. (2016), <http://creativecommons.org/licenses/by/4.0>.

and from histological c-fos immunostaining in the mouse (Shelton et al., 2012a,b).

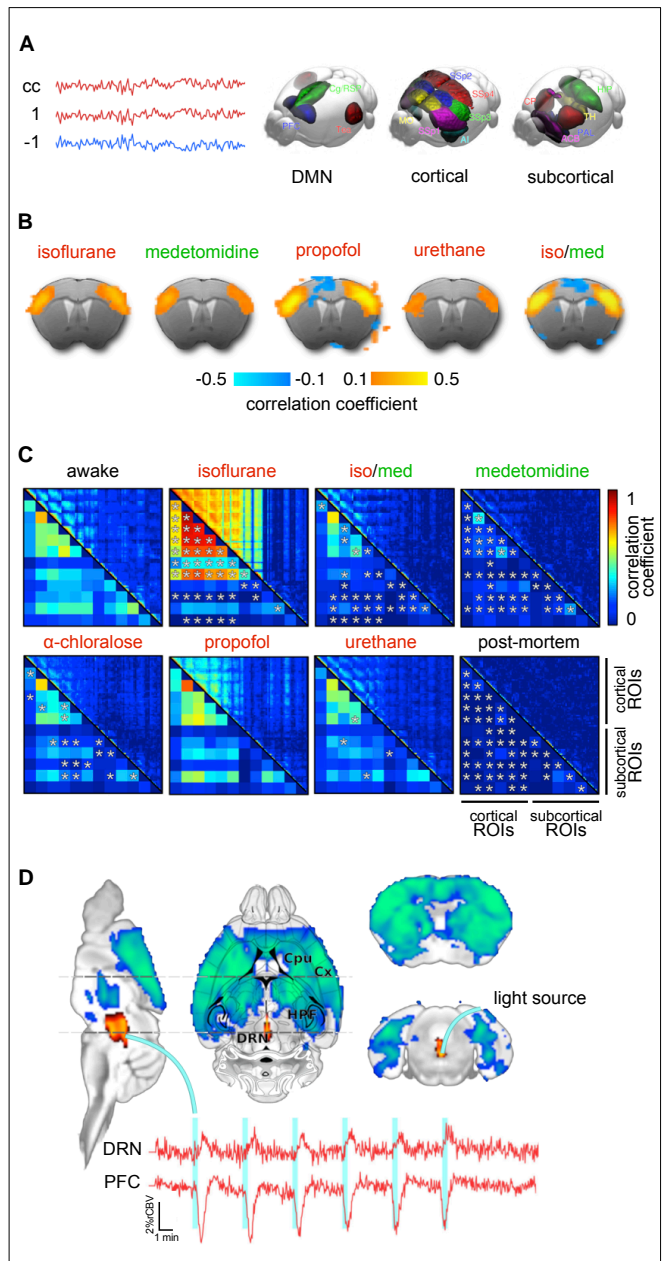
The detail of the BOLD patterns observed in this study was achieved by highly conservative thresholding, and has exposed peak values of extensive patterns spreading across large parts of the brain (**Figure 7B**). More liberal thresholding of heat-evoked BOLD effects in mice under comparable conditions revealed pattern distributions very similar to those observed for electrostimulation (Bosshard et al., 2015). In fact, nociceptive and arousal pathways are tightly interconnected (Craig, 2013). It is thus difficult to disentangle or exclude a potential contribution of ventilation-driven stress responses from those involved in nociception (as discussed in “Sensory Processing and the Key Challenges in Murine fMRI” section). Nociception drives arousal pathway activity, and hence cortical desynchronization, toward a

transient emergence from anesthesia (**Figure 2E**; Hudetz et al., 2003; Solt et al., 2014; Vazey and Aston-Jones, 2014; Akeju and Brown, 2017; Sanders, 2017; Hayat et al., 2019). Therefore, it is important to develop protocols that permit stable sedation in receptive, spontaneously breathing animals. The continuous monitoring of MABP permits the tracking of this important confounder, but it is highly invasive and introduces postoperative nociception. Sympathetic activity is similarly reflected in HR and respiration. Both may rise with the stimulus, but should subsequently return to the baseline (Reimann et al., 2016, 2018). A progressively rising HR requires adjustment of the anesthetic dosage to avoid the animal waking (Nasrallah et al., 2014b; Schroeter et al., 2014; Shim et al., 2018).

The interconnection of nociceptive and arousal routes causes a transient lightening of anesthesia in response to nociceptive stimuli. This could explain why BOLD responses to noxious electrostimulation are partially preserved in humans even after propofol-induced LOC, whereas sensory evoked patterns vanish (see “Anesthetic Depth and Consciousness—the Virtue of Translational Neuroimaging” section; Lichtner et al., 2018). In relation to our initial question, of whether consciousness is required to evoke a pain-specific neurosignature, it could be said that nociception itself drives the brain state toward consciousness. The proposed architecture of the *body-self neuromatrix* is composed of sensory, affective, and cognitive subnetworks of brain areas. Their output patterns produce the multiple dimensions of the pain experience (sensory-discriminative, cognitive-evaluative, motivational-affective), which coincide with homeostatic and behavioral responses (Melzack, 2005; Craig, 2013; Seth and Tsakiris, 2018). Dynamic changes in the FC of these respective networks, triggered either by external nociceptive stimuli or by internal processes in animal models of chronic pain, may offer a more concrete representation of the neuromatrix, as was recently addressed as the dynamic pain connectome (Kucyi and Davis, 2017).

## Functional Connectivity and Murine Resting-State fMRI

FC between two or more brain areas is defined by the correlation coefficient of their signal intensity courses over time: the higher the correlation, the higher the FC (**Figure 8A**; Petersen and Sporns, 2015). This simple relation is the key principle in defining functional networks across the entire brain—an opportunity that is unique to fMRI. FC is considered to be in large part the hemodynamic translation of simultaneous activity of neural populations, which is inherently limited by the physical connections between neurons (Mateo et al., 2017). This structural scaffold permits a repertoire of possible functional configurations that changes dynamically on a time scale depending on the level of organization: on the order of 1 ms for neurons, 10 ms for local circuits, 100 ms for EEG, and 1 s for fMRI (Mashour and Hudetz, 2018). This has two important implications: (1) the temporal changes in FC hold important information that is simply discarded when large fMRI time series are averaged assuming steady-state conditions (Preti et al., 2017); and (2) brain states can, in principle, be translated into FC and assessed *via* fMRI (Van de Ville et al., 2010).



**FIGURE 8 |** Functional connectivity (FC) in the mouse brain. **(A)** FC is defined by the correlation coefficient (cc) of remote brain areas. The left panel shows an assembly of important areas in murine FC, including the default mode network (DMN), cortical sensory-motor networks, and subcortical networks. **(B)** Murine FC correlation maps for five anesthetic protocols. Seed voxels in the anterior primary somatosensory cortex show anticorrelated time courses in the cingulate cortex. Anticorrelation is an essential criterion for anesthetic states in fMRI in which the FC is not restricted to its structural scaffold. **(C)** Correlation matrices in the rat for six different anesthetic protocols referenced against awake rats. Asterisks indicate statistically significant differences compared with the awake group ( $t$ -test,  $p < 0.05$ , false discovery rate corrected). **(D)** Optogenetic stimulation of the dorsal raphe nuclei that send serotonergic projections into large parts of the brain leading to wide patterns of negative BOLD effects. Adapted from **(A)** Grandjean et al. (2019a); **(B)** based on Grandjean et al. (2014), reproducible using original data from the open repository, <https://central.xnat.org>; **(C)** adapted with permission from Paasonen et al. (2018); **(D)** Grandjean et al. (2019b).

The integrity or disruption of FC is commonly used as an indicator of various types and stages of neurodegenerative diseases and neurophysiological malfunctions, or to study the effects of pharmacological interventions, genetically modified proteins or neural populations under optogenetic control (see “Deep Brain Stimulation, Opto- and Chemogenetics” section). Notably, the majority of murine resting-state studies have been conducted under anesthesia (Jonckers et al., 2011, 2014, 2015; Guilfoyle et al., 2013; Grandjean et al., 2014; Mechling et al., 2014; Nasrallah et al., 2014a; Sforazzini et al., 2014; Liska et al., 2015; Bukhari et al., 2017; Wu et al., 2017; Pan et al., 2018). The extent to which various anesthetic protocols intrinsically disrupt FC is therefore a critical issue that must be comprehensively addressed. A number of excellent reviews are already available on this issue for humans (Mashour and Hudetz, 2017, 2018) and mice (Hoyer et al., 2014; Jonckers et al., 2014, 2015; Pan et al., 2015; Gozzi and Schwarz, 2016; Chuang and Nasrallah, 2017; Sumiyoshi et al., 2019; Grandjean et al., 2020; Mandino et al., 2020). Here, we add to the existing literature by focusing on basic concepts of anesthesia-specific changes in FC with respect to the underlying mechanisms.

FC is remarkably similar across mammalian species. In 2014, 3 years after the first study on murine FC (Jonckers et al., 2011), it was shown that mice exhibit a medial fronto-parietal or “default mode” network (DMN) during the resting state (Stafford et al., 2014) similar to that of rats (Lu et al., 2012), monkeys (Vincent et al., 2007), and humans (Raichle et al., 2001). The DMN plays a fundamental role in baseline functions of the mammalian brain (Raichle, 2015). In humans, it is associated with mind-wandering and rumination. Whether the DMN entails comparable phenomenal qualities in mice remains speculative and may be interesting from an evolutionary or philosophical point of view (Seth et al., 2005; Havlik, 2017). In the presence of a salient stimulus, DMN connectivity becomes disrupted and shifts towards connectivity in the dorsal attention network (Raichle, 2015), in the same way that cortical states change with attention toward a stimulus (see “Cortical States of Wakefulness” and “Brain States Under Anesthesia” sections).

The DMN involves similar higher-order cortical regions along the medial fronto-parietal axis across species (**Figure 8A**; Raichle, 2015; Grandjean et al., 2020), although prefrontal and cingulate regions appear to perform the tasks of the precuneus, which is absent in rodents (as reviewed in Gozzi and Schwarz, 2016). Further functional networks have been described in the mouse, including a salience network and latero-cortical network, as well as various sensory, sensorimotor, cerebellar, limbic system, and basal ganglia networks (see **Figure 8B** for murine FC maps based on a seed voxel in S1 for different anesthetics; Jonckers et al., 2011, 2014; Grandjean et al., 2014, 2017b, 2020; Nasrallah et al., 2014b; Sforazzini et al., 2014; Zerbi et al., 2015; Shim et al., 2018; Gutierrez-Barragan et al., 2018).

## Anesthesia-Induced Disruptions in FC

The network architecture is very similar between humans and mice (Bullmore and Sporns, 2012; Gămănuț et al., 2018; Balsters et al., 2019; Fulcher et al., 2019), which has been utilized

to pursue questions about human network configurations in the mouse (Fulcher and Fornito, 2016). Brain networks are organized around highly connected hubs. Substantial disruption of FC induced by anesthesia at these central nodes can cause widespread communication failure associated with LOC (Schröter et al., 2012; Lee H. et al., 2013; Moon et al., 2015; Bonhomme et al., 2016; Tononi et al., 2016; Mashour and Hudetz, 2018). Abnormal reorganization of hubs during recovery from anesthesia likely explains differences in FC strength and configuration observed prior to and following LOC (Långsjö et al., 2012; Liu et al., 2013b; Monti et al., 2013). This sheds light on the anesthetic hysteresis effect: namely, that different anesthetic dosages are required to induce a similar depth of anesthesia during the induction and recovery phases (see “Brain States Under Anesthesia” section). The translational utility of murine FC studies may depend on the capacity of the anesthetic protocol to preserve or restore the integrity of specific functional networks across the brain.

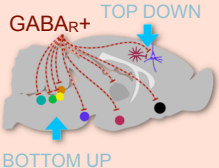
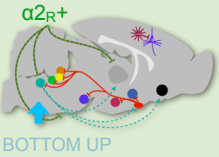
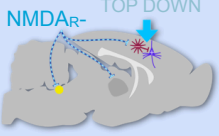
Of course, the technical equipment, animal preparation protocols, arousal states due to stress, murine susceptibility to vascular effects of anesthesia, and the applied statistical methods all have a crucial impact on the reliability of the data (Grandjean et al., 2020). It is therefore helpful to reference murine FC data against: (1) FC data derived from more stable systems like humans, monkeys, or rats under the same anesthesia; and (2) the molecular pathways that are known to be targeted by specific anesthetic protocols. Disruption of neural communication across brain areas occurs according to the principal target routes of an anesthetic regimen (see “Brain States Under Anesthesia” section) and may be translated into a discrete FC signature (**Table 3**). This relation makes the assessment of FC a highly relevant point of convergence for detailing and understanding brain states induced by anesthesia.

## Anticorrelation and Rich Dynamic Repertoires—Criteria for Appropriate FC

Discrete signatures of GABAergic drugs and volatile ethers, which suppress brain-wide neural activity include dose-dependent disruption of FC across the entire brain (for details, see **Table 3**; Boveroux et al., 2010; Schröter et al., 2012; Jordan et al., 2013; Ní Mhuircheartaigh et al., 2013; Palanca et al., 2015; Bonhomme et al., 2016; Ranft et al., 2016). This is reflected in a reduction in global connectedness, with stronger effects on higher-order cortical and thalamic FC compared to sensory areas, and fewer transitions between functional networks (Uhrig et al., 2018; Golkowski et al., 2019). This rigidity in the dynamics between networks correlates with a reduced ability to interact with external stimuli (see “GABAA Agonists Suppress Neural Activity Across the Central Nervous System” section). LOC has consistently been found to be correlated with a breakdown of fronto-parietal FC in humans and non-human primates induced by GABAergic drugs (Purdon et al., 2009; Boveroux et al., 2010; Jordan et al., 2013; Monti et al., 2013; Palanca et al., 2015; Ranft et al., 2016; Mashour and Hudetz, 2017; Uhrig et al., 2018), likely due to impaired frontal-to-parietal cortical information transfer (Ku et al., 2011; Jordan et al., 2013; Ranft et al., 2016). Although this relationship is also



**TABLE 3 |** Functional connectivity (FC) disruptions reflect principal mechanisms of anesthesia.

	<p><b>target sites.</b> cortico-cortical, thalamo-cortical, subcortico-cortical, subcortico-subcortical [1–6]</p> <p><b>LOC.</b> breakdown of fronto-parietal FC (dose-dependent loss of frontal-to-parietal information transfer) [1,2,4,6,7–10]</p> <ul style="list-style-type: none"> <li>• disruptions in thalamo-cortical FC: midline and intralaminar thalamic nuclei that drive cortical arousal, higher order thalamic nuclei that relay cortical sensory input [6,11–18]</li> <li>• FC of first order thalamo-cortical sensory connections and primary sensory networks usually remain preserved [2,14,15,18,19,20]</li> <li>• disruptions of cortico-cortical FC likely originate from reduced communication between neighboring neurons, which causes fluctuations in local sum activity [21–23]</li> <li>• this also impairs the propagation of stimulus-related cortical activity at low to moderate anesthetic depth [23–25]</li> </ul>	<p>[1] Boveroux (2010); [2] Jordan (2013); [3] Ni Mhuircheartaigh (2013); [4] Palanca (2015); [5] Bonhomme (2016); [6] Ranft (2016)</p> <p>[7] Purdon (2009); [8] Ku (2011); [9] Monti (2013); [10] Mashour &amp; Hudetz (2017)</p> <p>[11] Sherman (2005); [12] Alkire (2009); [13] Liu (2013c); [14] Saalman (2014); [15] Baker (2014); [16] Panagiotaropoulos (2014); [17] Kanai (2015); [18] Kundishora (2017)</p> <p>[19] Longmuir &amp; Pashko (1976); [20] Liu (2013a)</p> <p>[21] Lewis (2012); [22] Vizuet (2014); [23] Mashour &amp; Hudetz (2018)</p>
	<p><b>target sites.</b> locus ceruleus projections to thalamic, brain stem, forebrain arousal nuclei; disinhibition of sleep promoting hypothalamic nuclei decreases arousal level by direct or indirect cortical projections [28–34]</p> <p><b>LOC.</b> FC breakdown between locus ceruleus, thalamus and posterior or anterior cingulate cortex may eventually affect fronto-parietal networks [33,35–37]</p> <ul style="list-style-type: none"> <li>• FC characteristics closely resemble those of sleep in the same way as cortical states do, thereby reflecting the functional disconnection of the cortex from subthalamic areas like in the <i>cerveau isolé</i> (see Fig. 3A, green scissors) [38–41]</li> <li>• disrupted thalamo-cortical FC, but preserved cortico-cortical FC (e.g. saliency network) and primary sensory routes allow for easy <i>arousability</i> (comparable to sleep) [36,40,42,43]</li> <li>• with increasing doses, suppression of arousal nuclei alter cortical states [44]</li> <li>• local cortical activity gets coupled with slower rhythms of their subcortical afferents building uncoupled islands across the cortex [21,45]</li> </ul>	<p>[28] Jones &amp; Yang (1985); [29] Vogt &amp; Pandya (1987); [30] Buckwalter (2008); [31] Vogt (2008); [32] Brown (2011); [33] Akeju (2014); [34] Leung (2014)</p> <p>[35] Langsjo (2012); [36] Picchioni (2014); [37] Fernandez (2017)</p> <p>[38] Akeju &amp; Brown (2017); [39] Lewis &amp; Akeju (2017); [40] Guldenmund (2017); [41] Bremer (1935)</p> <p>[42] Horowitz (2008); [43] Larson-Prior (2009)</p> <p>[44] Song (2017)</p> <p>[45] Nir (2011)</p>
	<p><b>target sites.</b> primarily cortico-cortical [23,46–49]</p> <p><b>LOC.</b> direct suppression of fronto-parietal and further cortico-cortical FC [5,23,46–51]</p> <ul style="list-style-type: none"> <li>• in contrast to the effects of GABAergic and <math>\alpha_2</math>AR agonists, activation of preoptic area and other wake-promoting nuclei, and increase of thalamic metabolism [51,52]</li> <li>• FC widely preserved in subcortical areas [50–52]</li> <li>• at low dosage, global increases in FC have been observed [53]</li> </ul>	<p>[46] Voss (2012); [47] Lee (2013b); [48] Blain-Moraes (2014); [49] Vlisides (2017)</p> <p>[50] Muthukumaraswamy (2015)</p> <p>[51] Mashour (2014); [52] Alkire &amp; Miller (2005)</p> <p>[53] Driesen (2013)</p>

The table summarizes key effects on FC for the three major anesthetic classes: disruption of whole brain FC by GABA<sub>A</sub> receptor agonists (GABA<sub>R</sub>+, red), bottom-up disruption by  $\alpha_2$  adrenoceptor agonists ( $\alpha_2$ R+, green), and disruption of cortical FC by NMDA receptor antagonists (NMDA<sub>R</sub>–, blue). Loss of consciousness (LOC) is most likely mediated by suppression of frontal-to-parietal information transfer in the cortex and can be achieved by all three anesthetic classes either directly via disruption of cortico-cortical FC or indirectly via bottom-up pathways. Principal mechanisms of anesthesia are described in “Brain States Under Anesthesia” section and their reflection in FC patterns is discussed in “Functional Connectivity and Murine Resting-State fMRI” section. References are shortened to first author (and year) to save space and allow for comprehensive referencing.

thought to exist in mice, further studies are required to assess the modes of murine fronto-parietal FC and their significance for wakefulness (Imas et al., 2005b; Mashour and Alkire, 2013; Sforazzini et al., 2014; Pal et al., 2018; Shofty et al., 2019; Grandjean et al., 2020).

On the other hand, burst suppression and slow waves that propagate unabated across the cortical surface (see “Crossing the Borders: GABAergic Slow Waves and Burst Suppression” section) promote FC that can span extensive cortical networks in humans and rats, occasionally reaching into the basal ganglia and thalamic structures (Liu et al., 2011, 2013b,c; Golkowski et al., 2017; Schwalm et al., 2017; Aedo-Jury et al., 2019). Isoflurane concentrations as low as 1.3% in ventilated rats were found sufficient to induce highly synchronized large-scale FC patterns in fronto-cortical and striatal regions, and to suppress most thalamo-cortical and intra-subcortical connections (Figure 8C; Paasonen et al., 2018).

Loss of information transfer and integration is further reflected in a reduced number of possible FC configurations

(Hutchison et al., 2014; Barttfeld et al., 2015; Hudetz and Mashour, 2016; Ma et al., 2017; Uhrig et al., 2018). The repertoire of FC configurations can be evaluated *via* a sliding-window approach. In this approach, a window length of about 1 min is defined and shifted along the entire length of the resting-state time series, with a step size equal to the fMRI temporal resolution (typically 1–2 s). Each time-shifted segment then undergoes a correlation analysis, which gives access to dynamic aspects of FC, which have been reported to be well conserved across rats, monkeys, and humans (Majeed et al., 2011). Although blurred in temporal resolution, this technique permits the assessment of dynamic whole-brain signatures based on FC. Combining Ca<sup>2+</sup> imaging or electrophysiology with fMRI (Thompson et al., 2013; Keilholz, 2014; Schmid et al., 2016; Schwalm et al., 2017; Lurie et al., 2018; Schlegel et al., 2018; Matsui et al., 2018; Aedo-Jury et al., 2019; van Alst et al., 2019), these signatures can be linked to the cortical states, and hence be employed to determine the nature and level of sedation (see “Brain States Under Anesthesia” and “Limitations in Maintaining Stable Brain States” sections).



The dynamical elaboration of a rich, flexible repertoire in the awake brain becomes increasingly constrained with increasing anesthetic dosage (Barttfeld et al., 2015; Uhrig et al., 2018). This is accompanied by a lack of anticorrelated areas—a further symptom of inflexible FC and confined information traffic pointing to LOC. A lack of anticorrelated areas was reported in mice ventilated at 1% isoflurane (**Figure 8B**; Grandjean et al., 2014). At 1.3% isoflurane broad cortical FC has been observed along with synchronized cortical slow wave activity in the  $\text{Ca}^{2+}$  signal assessed in the murine S1 (Schlegel et al., 2018). Further reducing the isoflurane concentration below 1–1.3% is not appropriate if the mice are ventilated and paralyzed (see “Sensory Processing and the Key Challenges in Murine fMRI” section). As head movements, including rhythmic, respiration-induced motion, account for a large part of the noise in FC (Kalthoff et al., 2011; Pais-Roldán et al., 2018), a trade-off has to be made between data quality and anesthetic depth. In ventilated rats, a reduction of isoflurane from 1.5% to 1% was sufficient to preserve distinct cortical networks and anticorrelated FC at a largely desynchronized frequency band in the EEG (Liu et al., 2013c). Due to the narrow dose range between stable sedation and burst suppression, careful adjustment of isoflurane concentrations is required to avoid unphysiological FC.

Compared to isoflurane and other volatile ethers, halothane exhibits a remarkably wide dose range between stable sedation and burst suppression (Murrell et al., 2008; Brown et al., 2018; McIlhorne et al., 2018). A rich repertoire of dynamic FC patterns was observed for halothane-anesthetized mice, based on a varied set of recurring configurations of FC, with abundant anticorrelated areas (Sforazzini et al., 2014; Gutierrez-Barragan et al., 2018). At low dosages, other GABAergic drugs have been shown to preserve FC quite well with respect to the above criteria. The use of low-dose propofol and urethane was reported to achieve the best resemblance to FC in awake rats out of six different anesthetic regimens tested (**Figure 8C**; Paasonen et al., 2018). A similar observation was reported for mice, with propofol preserving spatially defined symmetrical cortico-cortical FC with anticorrelated cingulate areas (**Figure 8B**; Grandjean et al., 2014). Nevertheless, unstable sedation has been reported for this protocol in mice, which could be due to ventilation conditions, considering the general arousability of mice relative to rats. Comparing the capacity of anesthetic protocols to approximate the FC of the awake state in rats, propofol and urethane showed the best performance, followed by  $\alpha$ -chloralose, while a protocol combining isoflurane and medetomidine (see below) was ranked 4th (**Figure 8C**).

### “One-Dimensional” Disruption of FC

Medetomidine and other  $\alpha 2\text{AR}$  agonists, which suppress subcortical arousal nuclei (see “Brain States Under Anesthesia” section) disrupt FC in a bottom-up manner (**Table 3**). Disinhibition of sleep-promoting cells in the hypothalamus disrupts subcortico-subcortical and hence thalamo-cortical FC. This causes a discrete FC signature that reflects the functional disconnection of the cortex from subcortical areas, as anticipated on the basis of the *cerveau isolé* (“Brain States Under Anesthesia”

section; Bremer, 1935; Guldenmund et al., 2017; Lewis and Akeju, 2017; Akeju and Brown, 2017). With increasing dosages, local cortical activity gets coupled with slower rhythms of subcortical afferents from the suppressed arousal nuclei, building uncoupled islands across the cortex (Nir et al., 2011; Lewis et al., 2012). This is reflected in disrupted interhemispheric cortico-cortical FC, which has been observed with increasing dosages of medetomidine in rats (Nasrallah et al., 2014a) and mice (Nasrallah et al., 2014b). The mechanism of bottom-up-induced LOC likely involves a breakdown of FC between the locus ceruleus, the thalamus, and the posterior or anterior cingulate cortex, which may eventually affect fronto-parietal FC (Långsjö et al., 2012; Fernandez et al., 2017). Medetomidine was found to preserve moderate intercortical connectivity in ventilated rats, although this was significantly reduced compared to that of awake rats, whereas thalamo-cortical activity was almost absent (**Figure 8C**; Pawela et al., 2008; Zhao et al., 2008; Paasonen et al., 2018).

Ketamine, on the other hand, which acts primarily in a top-down manner in the cortex, mediates LOC by direct suppression of fronto-parietal and further cortico-cortical FC (**Table 3**; Voss et al., 2012; Lee U. et al., 2013; Blain-Moraes et al., 2014; Muthukumaraswamy et al., 2015; Bonhomme et al., 2016; Vlisides et al., 2017; Mashour and Hudetz, 2018; Uhrig et al., 2018). In contrast to GABAergic and  $\alpha 2\text{AR}$  agonists, ketamine has been reported to largely preserve subthalamic FC, which is consistent with its activating effects on the preoptic area and other wake-promoting nuclei (Alkire and Miller, 2005; Mashour, 2014). The top-down and bottom-up actions of ketamine and medetomidine may be thought of as “one-dimensional” (Mashour and Hudetz, 2017). The dual effect of GABAergic drugs makes them more suitable for stable sedation by suppressing both subthalamo-cortical and cortico-cortical FC (Boveroux et al., 2010). To achieve a similar anesthetic stability by not overly stressing one pathway, drugs targeting either of these “one-dimensional” mechanisms are commonly combined (e.g., ketamine-xylazine or medetomidine-isoflurane) in order to target the complementary pathways.

### Multimodal Anesthesia in Murine Resting-State fMRI

Combining isoflurane and medetomidine at levels roughly half their mono-anesthetic dosages (Lu et al., 2012; Paasonen et al., 2018; Sumiyoshi et al., 2019) has been shown to produce FC resembling that of awake rats substantially better than either mono-anesthetic protocol alone, yielding a correlation matrix that roughly corresponds to that of  $\alpha$ -chloralose (**Figure 8C**; Paasonen et al., 2018). This medetomidine-isoflurane protocol was adapted for murine resting-state fMRI, where it was also described as providing the “best of both worlds”: reproducible, spatially defined networks, including symmetrical patterns of FC in bilateral sensory areas anticorrelated with cingulate regions, at a stable sedation in ventilated mice (**Figure 8B**; Grandjean et al., 2014).

Later studies revealed a high correspondence between FC and structural connectivity of cortico-cortical and cortico-striatal regions with the same medetomidine-isoflurane protocol in

mice, whereas thalamo-frontal cortical FC was found suppressed (Grandjean et al., 2017b). Although a high correspondence between functional and structural connectivity has been associated with a constrained repertoire of FC configurations (Barttfeld et al., 2015; Uhrig et al., 2018), the sliding-window approach ultimately revealed rich patterns of dynamic FC under this protocol (Grandjean et al., 2017a). This incongruity could be due to methodological differences between the studies (e.g., the use of MR diffusion-based vs. neural tracer-based structural connectivity; Straathof et al., 2019). The exact relationship between the dynamics of FC and its correspondence to its scaffold as a function of anesthetic depth remains to be clarified across anesthetic protocols in mice.

Taken together, the preservation of a rich repertoire of FC configurations, along with spatially defined and anticorrelated FC in relevant networks, has made medetomidine-isoflurane the protocol of choice for studying murine FC. Low-dose halothane performs similarly well in mice, with the advantage of preserving thalamo-frontal cortical FC (Gutierrez-Barragan et al., 2018; Bertero et al., 2018; Mandino et al., 2020). Halothane is rarely used due to a risk of causing liver damage, although the risk is possibly no higher than that of other halogenated vapors (Mizobe, 2019).

A recent meta-analysis comparison across numerous labs, MR scanners, and animal preparation protocols (Grandjean et al., 2020) revealed greater specificity within elements of the DMN under medetomidine-isoflurane anesthesia in ventilated mice than in awake habituated mice, although the latter exhibited higher overall FC. This can be interpreted as an effect of environmental or internal distraction in awake mice, since DMN connectivity is known to be disrupted in the presence of salient stimuli (see above; Raichle, 2015). Contradictory findings have been reported for anesthetic effects on the DMN, although it appears that FC is principally supported at low anesthesia and disrupted with increasing anesthetic depth (Akeju et al., 2014a; Huang et al., 2014; Palanca et al., 2015; Ranft et al., 2016).

### Resting-State fMRI in the Awake Mouse

Medetomidine-isoflurane and low-dose halothane are currently considered the most suitable anesthetic protocols available for the assessment of murine FC. Yet, carefully titrated multimodal protocols specifically designed to balance the complementary effects of their anesthetic compounds offer far more room for improvement (see “Multimodal Anesthesia in Translational fMRI” section). Comparing FC matrices for medetomidine-isoflurane against propofol with reference to the “ground truth” of an awake habituated rat gives an immediate impression of what we are heading for (Figure 8C). With respect to the *non-voluntary task conundrum*, resting-state applications have the best chance to produce reliable data from awake habituated mice. Nevertheless, training awake mice to tolerate the noisy scanner environment is far more challenging than for awake rats (Jonckers et al., 2014, 2015; Low et al., 2016b; Dopfel and Zhang, 2018), although further strategies are constantly being developed to facilitate this procedure (Yoshida et al., 2016; Madularu et al., 2017; Han et al., 2019; Chen et al., 2020).

Since brain states during wakefulness alter with the level of arousal, such a “ground truth” is still prone to reflect stress and restraint rather than the resting state, which becomes critical when comparing results with human volunteers (Bergmann et al., 2016). It might be useful to administer minimal anesthetic dosages (e.g., of anxiolytics) in the actual experiment to relax mice, which are already habituated. This would create a state of drowsy wakefulness that approximates the “ground truth” against which novel anesthetic protocols could be referenced. Ultimately, multimodal sedation may be instrumental to lock the mouse into a defined spectrum of brain states to minimize variations due to stress, arousal, and distraction, and to prevent effects due to habituation and reward.

### Deep Brain Stimulation, Opto- and Chemogenetics

Functional reorganization and network properties can be further evaluated by using techniques that modulate neural activity from within the brain. The most popular methods are deep brain stimulation, opto- and chemogenetics. These techniques allow one, for example, to evaluate the reorganization of FC in response to locally targeted perturbations. This can be used as a measure for functional complexity of information processing in the anesthetized brain (perturbational complexity index; Casali et al., 2013; Mashour and Hudetz, 2018). Ketamine, for instance, has been shown to preserve cortical functional complexity, which is decreased by propofol with increasing dosage (Sarasso et al., 2015). The signal propagation of locally induced neural activity also provides useful information on FC, which has been shown to be impaired during anesthesia, sleep, and pathological unconsciousness (Massimini et al., 2005; Ferrarelli et al., 2010; Casali et al., 2013). In humans, such investigations are often limited to studying cortical FC using non-invasive transcranial magnetic stimulation. In preclinical fMRI deep brain stimulation can be more easily utilized to induce neural activity in subcortical areas (McIntyre and Anderson, 2016), although the electrodes are prone to cause imaging artifacts, distortions, and local signal loss (Lehto et al., 2018).

Optogenetics employs optical fiber probes that do not directly interfere with functional MR image acquisition. Excitation or inhibition of genetically modified neurons governed by light control can be combined with fMRI to detail spatial and temporal effects on evoked neural activity (Desai et al., 2011; Lai et al., 2015) or FC in remote brain regions (Ryali et al., 2016; Chan et al., 2017). Chemogenetics is a powerful minimally invasive complement to optogenetics, in which neurons are controlled by artificially designed ligands that are injected into the bloodstream (Roth, 2016). Both are important techniques to further investigate the mechanisms of FC (as recently summarized in Mandino et al., 2020), including the impact of distinct transmitter systems in mice. Exclusively selecting, for example, serotonergic neurons of the raphe nuclei (Figure 8D; Grandjean et al., 2019b) or noradrenergic neurons of the locus coeruleus (Zerbi et al., 2019) based on specific gene expression patterns can be used to detail the underpinnings of brain states

in anesthesia and the circuits that drive awakening (Carter et al., 2010; Taylor et al., 2016).

## TRANSLATIONAL fMRI ACROSS BRAIN STATES AND SPECIES

The key challenges of preclinical fMRI with a view to translation are: (1) to induce and maintain brain states that correspond with those of calm and relaxed human volunteers; and (2) to assure hemodynamic integrity—in other words, to preserve neurovascular coupling and physiological baseline conditions. Up to this point, we have discussed the general principles of anesthesia and have highlighted the need to recognize and overcome the shortcomings of current anesthetic protocols. This will require strategies to shape, maintain, and unambiguously identify distinct brain states in fMRI that meet the above criteria, to conduct reliable, reproducible studies and translate findings from mice to humans.

### Multimodal Anesthesia in Translational fMRI

Preclinical fMRI aims for an animal that is sufficiently sedated, yet with a fully functional brain. This challenge puts high demands on anesthetic protocols to induce and maintain an intermediate brain state that closely approximates an awake, calm, relaxed, and undistracted state during the fMRI task. Current standard practices involving simple mono-anesthesia are outdated. Balanced multimodal anesthesia, in which two or more drugs are combined to balance their complementary effects, has come into use in the clinic (Brown et al., 2018a) and should be further developed for preclinical studies (see “Sensory Processing and the Key Challenges in Murine fMRI” and “Functional Connectivity and Murine Resting-State fMRI” section; Grandjean et al., 2014; Sharp et al., 2015; Shim et al., 2018).

By reducing the dose of individual compounds, a multimodal protocol reduces excessive stress on a single pathway, taking advantage of complementary modes of action to shape and maintain desired brain states. In principle, respiratory depression or hemodynamic side effects from specific drugs can be attenuated by balancing them with the effects of others. At the same time, combining anesthetic compounds through multimodal anesthesia introduces a higher level of complexity and potential cross-effects that require caution. A rigorous characterization of anesthetic actions will be necessary to understand and adjust the effects of individual drugs in multimodal approaches. This will require that brain states, hemodynamics, and physiology be monitored, evaluated, and reported to the community.

### Defining Brain States in fMRI: Qualifiers and MIND Signature

To properly identify and communicate an anesthetic state in translational fMRI, we first need to define a set of functional characteristics that can identify an induced brain state with high specificity. Their purpose is to evaluate the anesthetic maintenance of an intermediate brain state, and qualify its

approximation to the awake state. We will refer to these functional characteristics as *qualifiers*.

Among these qualifiers, we choose one *identifier* that allows the unambiguous assessment of whether a desired intermediate brain state has been maintained during the full length of the fMRI scan, based on a reliable neurophysiological signature pattern. It should permit: (1) identifying a brain state and the integrity of its qualifiers; and (2) comparing this state across subjects, trials, scanners, labs, and species. We will refer to this identifier of a specific brain state as a maintained, intermediate neurophysiologically-determined (MIND) signature.

### Qualifiers

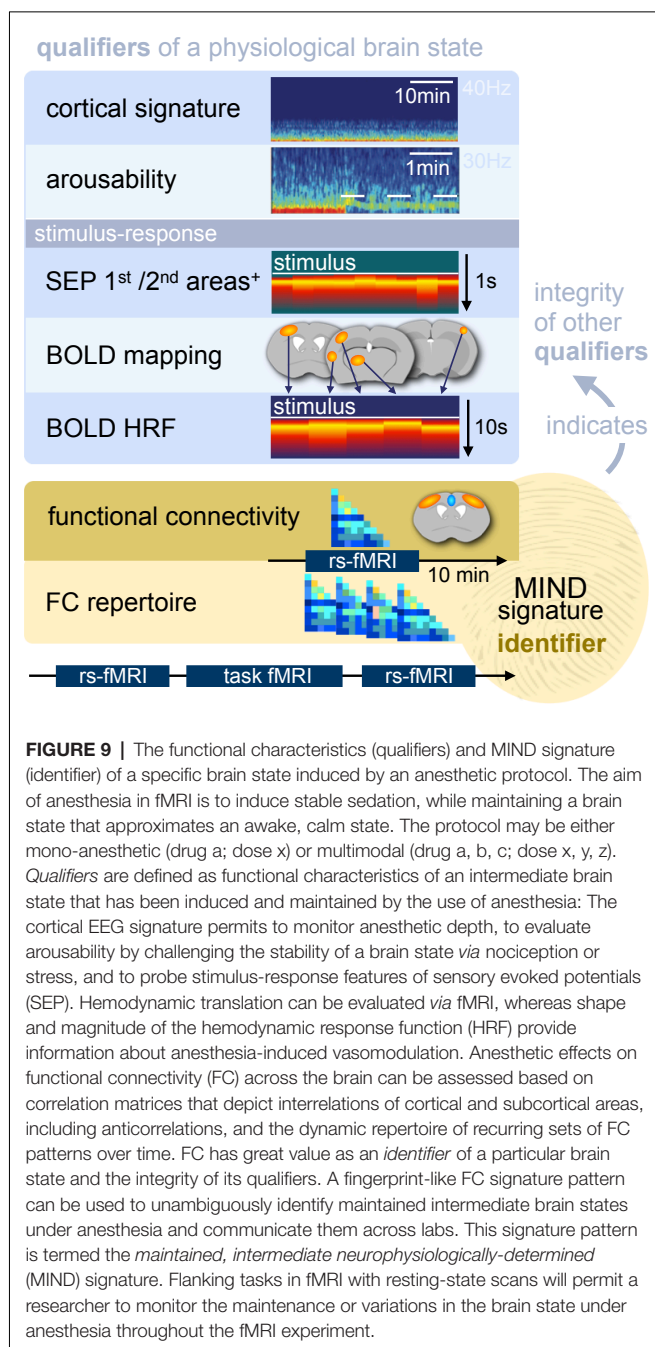
A number of qualifiers have the potential to illuminate characteristics of brain states and the functional implications that accompany them. The set of such qualifiers is growing rapidly through new findings and the development of novel analytical tools (as summarized in Bonhomme et al., 2019). Here, for the sake of brevity, we list only those that we consider most relevant and sufficient for defining brain states for fMRI in anesthetized animals.

Under these conditions, the set of functional characteristics needed to identify an anesthetically induced brain state should include measures of maintenance, arousability, neuronal stimulus–response features, and their hemodynamic translation, as well as FC and its dynamic repertoire (**Figure 9**). These qualifiers cover a range of informative features to identify and evaluate the degree to which an induced brain state approximates the calm, awake state.

Cortical EEG signatures provide valuable information about cortical synchronization at particular frequency bands (including correlates of phase-amplitude coupling, which can influence cortical information transfer and sensory processing; see “Brain States Under Anesthesia” section; Chamadia et al., 2019), the stability of a maintained intermediate brain state during rest and its response to nociception and stress (see “Metastability of Brain States, Hysteresis, and Behavioral Monitoring” section). Due to its real-time accessibility and high temporal resolution, EEG can be used to assess anesthetic depth and arousability, particularly when aligned with behavioral observations (outside the scanner).

The dynamics of neural stimulus–responses can be directly assessed *via* EEG (whose spatial resolution is limited due to the smoothing properties of the scalp and skull) or invasive electrophysiological recordings of SEP and  $\text{Ca}^{2+}$  signals in primary, secondary sensory, and higher-order cortical areas. Multichannel recordings (Muller et al., 2018) and large-scale assessment techniques of neural activity, including high-density EEG (Massimini et al., 2005, 2007),  $\text{Ca}^{2+}$ - and voltage-sensitive dye imaging (McVea et al., 2016), can be used to assess the degree of complexity of signal processing and propagation across the cortex at different spatial and temporal scales (see also “Brain States Under Anesthesia” and “Deep Brain Stimulation, Opto- and Chemogenetics” sections). This provides insights into the effects of anesthesia on thalamo-cortical and cortico-cortical sensory transmission, as well as cortical information transfer (see “Brain States Under Anesthesia” section).





Hemodynamic integrity can be evaluated *via* the translation of neural responses into BOLD effects. Comparing SEP or evoked  $\text{Ca}^{2+}$  signals with BOLD effects reveals whether neural responses are triggered but become lost in hemodynamic translation (Aksenov et al., 2015). Neurovascular coupling varies across cortical areas and has been found to be more reliable in sensory vs. higher-order cortices (Ojemann et al., 2002, 2013; Conner et al., 2011). Therefore, an overall impairment of neurovascular coupling by anesthesia might lead to a selective loss of BOLD effects. Further information about vasoconstrictive or vasodilatory drug effects can be derived from the shape of the

hemodynamic response function (HRF; see “Sensory Processing and the Key Challenges in Murine fMRI” section).

Since anesthesia primarily disrupts neural communication across the brain, its effects are most comprehensively reflected by measures of FC. An ideal sedation for fMRI would meet criteria of a close resemblance of the assessed correlation matrices with those of the awake references, including anticorrelated regions, and a rich repertoire of dynamic FC (see “Functional Connectivity and Murine Resting-State fMRI” section).

### MIND Signature

Another advantage of FC in fMRI is that it can be acquired simply in the form of a resting-state scan without any additional equipment. This makes it a strong candidate measure for determining a MIND signature in translational fMRI. fMRI tasks can be flanked with brief resting-state scans that permit the brain state to be identified at regular points throughout an experiment. Such flanking approaches of FC have previously been shown to reveal valuable information on anesthetic effects on murine fMRI tasks (Nasrallah et al., 2014b; Schroeter et al., 2017; Shim et al., 2018). More advanced FC applications have been used to track spatial and temporal aspects of intermediate brain state transitions as a function of anesthesia in mice (Grandjean et al., 2017a; Gutierrez-Barragan et al., 2018) and rats (Schwalm et al., 2017; Aedo-Jury et al., 2019). Finally, the relations of FC and its structural scaffold have been already used to formulate resting-state dynamics as cortical signatures of anesthesia in monkeys (Barttfeld et al., 2015; Uhrig et al., 2018) and humans (Roberts et al., 2019).

Approaches like these have laid the groundwork for the development of a simple, yet comprehensive signature scheme for the unambiguous identification of anesthetically induced and maintained, intermediate brain states that (1) are adjusted to functionally approximate the awake state and (2) indicate the integrity of a defined set of functional qualifiers. This concept assumes a mechanistic connection between FC and the other qualifiers, so that specific functional properties can be inferred from the detection of a specific MIND signature (Figure 9). Such a mechanistic connection remains to be established—as well as the circumstances in which this assumption can be made with confidence.

### Linking the MIND Signature to Other Qualifiers

Defining a MIND signature that could indicate the cortical state, stimulus-response properties, their hemodynamic translation, and dynamic FC in the form of one unique identifier would be extremely useful. The unambiguous identification of brain states during fMRI would also substantially improve the process of developing (and communicating) novel anesthetic protocols. A verifiable linkage between these qualifiers requires the simultaneous assessment of neural and hemodynamic readouts (see below). However, a large fraction of qualifiers can be assessed in separate measurements. It is reasonable to perform initial anesthetic adjustments outside the MR scanner, based on electrophysiological and behavioral responses to stimuli and cortical signatures during rest (see “Metastability of Brain



States, Hysteresis, and Behavioral Monitoring” section), while under a tight control of physiological parameters. It is therefore important to select conditions under which the qualifiers measured outside the scanner can be linked to others obtained during fMRI.

Brain states, and thus stimulus response properties, are likely to change during fMRI scans due to the loud gradient noise, which can trigger activity in subcortical arousal nuclei (see “Metastability of Brain States, Hysteresis, and Behavioral Monitoring” section). In fact, fMRI at high levels of sound pressure substantially boosts BOLD responses to subcutaneous electrostimulation of the rat under  $\alpha$ -chloralose, compared to silent fMRI protocols or after deafening by cochleotomy (Burke et al., 2000). Thus, hearing protection is recommended (Reimann et al., 2016). Sound pressure levels depend on the gradient coils, and can be reduced *via* sophisticated design of hardware and pulse sequences (Winkler et al., 2018).

Current efforts are underway to advance silent MR techniques using sweep imaging with Fourier transform (SWIFT; Idiyatullin et al., 2006, 2015). This approach renders obsolete the rapid and pulsed on/off switching of strong magnetic field gradients used for spatial encoding, and substantially lowers the acoustic noise level. SWIFT also permits zero echo time TE, which yields distortion- and artifact-free fMRI, even in the face of severe body motion and with electrodes implanted close to the region of interest (Lehto et al., 2018; Paasonen et al., 2020). This is a major advance for the development of sedation protocols to study FC in freely breathing animals and is further attractive for electrophysiological recordings or deep brain stimulation in fMRI. Until SWIFT is readily available for small-bore animal MR scanners, typical sound levels and vibrations produced by the gradients may be mimicked outside the scanner in the electrophysiological setup to test their effects on cortical states. Simultaneous fMRI and EEG (Sadaghiani et al., 2012; Thompson et al., 2013) or  $\text{Ca}^{2+}$  imaging (Schwalm et al., 2017; Schlegel et al., 2018) can permit a direct translation between the cortical state and whole-brain hemodynamic responses.

It remains to be clarified to what extent a signature based on FC is sufficient to indicate the integrity of the underlying functional properties of a brain state. Recent studies on the simultaneous imaging of murine cortex-wide FC of hemodynamic and neural calcium fluctuations in the resting state have found a strong relationship between both measurements (Murphy et al., 2017). Both measurements show sufficient resemblance in their intercortical FC for tracking similar brain states under anesthesia (Brier et al., 2019) and similar sets of dynamic FC configurations using a sliding-window approach (Matsui et al., 2018). Transitions in the dynamic FC configurations can be tracked *via* fMRI within a time scale of 1 or 2 s, and a sliding-window length below 1 min (Barttfeld et al., 2015, movie S1; Grandjean et al., 2017a; Lurie et al., 2018), which would be sufficient to resolve effects such as the metastable cortical state transitions that were observed *via* LFP recordings under constant isoflurane concentrations (see “Metastability of Brain States, Hysteresis, and Behavioral Monitoring” sections, **Figure 2D**; Hudson et al., 2014; Hudson, 2017). However, studies that investigate such brain state transitions under anesthesia

linking electrophysiological or calcium recordings with dynamic FC are still pending.

## Limitations in Maintaining Stable Brain States

Even when a brain state is stably maintained over time, it will exhibit a restricted set of recurrent FC configurations that vary within a defined dynamic range (Uhrig et al., 2018). Transitions between two intermediate brain states, as observed for metastable LFP cortical signatures (Hudson et al., 2014), should thus be classifiable as two distinct sets of recurrent FC configurations. Each of these sets varies within the internal dynamic range of its respective brain state. The MIND signature will comprise an indication of the variance of this range (narrow or broad) and the FC patterns that recur (Gutierrez-Barragan et al., 2018). A loss of recurrence and the appearance of new FC configurations would indicate a drift in brain states, and hence a change in anesthetic depth. Transitions in brain states become a confounding factor when they significantly alter the specific functional properties that are relevant for a given fMRI investigation.

The continuous anesthetic infusion *via* intraperitoneal or subcutaneous administration routes may present a practical experimental challenge because the uptake and distribution of agents depends on various biopharmaceutical and biological factors (comprehensively reviewed in Claassen, 1994). Whether and to what extent it is possible to maintain stable brain states under these conditions has to be established according to specific application procedures (Simpilatz et al., 2019). Future technical developments and novel methods should produce increasingly detailed functional information that will need to be integrated into single composite fMRI measurements (Breakspear, 2017; Belloy et al., 2018; Roberts et al., 2019; Song et al., 2019). This will produce a MIND signature that identifies brain states with increasing precision.

## The MIND Signature in Translational fMRI

The MIND signature would be valuable as an indicator of specific brain states independent of minor changes in brain anatomy and thus hold even further value for translational fMRI. It would provide a framework for comparing the MIND signatures of humans and mice and facilitate the development of a scheme for systematic evaluation of functional similarity and divergence across species. This would allow for better translation of anesthetic effects from humans to mice and back, and will have three major benefits: (1) anesthetic effects could be referenced against an awake, but calm and relaxed volunteer—even within the same subject; (2) introspective reports could link specific MIND signatures to LOC and to the overall experience; and (3) determining precise translational limitations in general brain functions and specific anesthetic effects, to assess the scope for murine transgenic and optogenetic applications to explore human cognition and consciousness.

## Future Perspectives

The need for effective anesthetic protocols and a better comprehension of their underpinnings is not exclusive to

translational fMRI, but is equally important for preclinical researchers studying electrophysiology,  $\text{Ca}^{2+}$  imaging, and other modalities to probe brain functions. In clinical practice, balanced multimodal anesthesia is common and increasingly being used as a substitute for opioids, especially in light of the current opioid crisis in the US (Brown et al., 2018a). In basic research, cognitive scientists employ anesthesia to probe the neural correlates of consciousness. All these disparate fields produce valuable data that should be taken into account across research domains in pursuit of a deeper understanding of anesthetic effects on animal and human brain functions.

Yet accessing knowledge on anesthetic effects is not always straightforward. Much valuable information is buried in publications that focus on biological research questions and use anesthesia only as an experimental tool. Many anecdotal observations of crucial effects remain unpublished “lab lore” and are not available to the community. Having access to this information could substantially accelerate our progress in detailing anesthetic effects to develop multimodal protocols for a stable sedation with the best possible preservation of brain functions. A community-driven open database of key observations—collected from publications and lab notes—would make this information more accessible, promoting reproducibility and preventing research groups from running into the same “dead ends” already discovered by others. We are currently developing a web-based infrastructure for such an open initiative across neuroimaging communities, to facilitate the exchange of this information. We invite anyone interested in this initiative to contact the corresponding author.

Within the fMRI community, the additional assessments of resting-state fMRI at selected points throughout the experiment would help to specify the actual brain state and make comparisons of data across labs more reliable. The concept of the MIND signature could support this process by defining a set of qualifiers and one identifier readout that permits a reliable indication of a specific brain state and its related functional properties. Our future work is dedicated to further refining and establishing the MIND signature in a collaborative effort.

## CONCLUSIONS

To make proper use of translational fMRI for a direct, non-invasive comparison between large-scale brain functions in humans and mice, it is essential to understand the

impact of anesthesia on neural physiology. The animal is usually anesthetized to avoid variations due to stress and distraction, with the ultimate aim of achieving stable sedation while maintaining a functional brain state close to that of an awake, resting subject. Maintaining functional brain states and neurovascular coupling are the two major challenges of preclinical fMRI. Most anesthetics affect neural and hemodynamic integrity in different ways, based on their target receptors, which typically assigns them to one of the three main classes of anesthesia. Despite addressing different pathways, all classes exert their effects by spoiling the phase-frequency relationships of remote neuronal populations, which eventually leads to a breakdown of information transfer along the cortical hierarchy and thus LOC. The resulting disruption of higher-order circuits is reflected in FC, although the specificity of markers for discrete functional stages remain to be clarified. Better knowledge of the mechanisms of action allows to balance different compounds in order to develop multimodal anesthetic protocols for translational fMRI, ideally revealing mainly species-specific variations and not the influence of anesthesia. The concept of the MIND signature has been proposed as an approach to determine and re-identify specific brain states and related functional properties induced by different anesthetic drugs and levels across studies, laboratories, and species.

## AUTHOR CONTRIBUTIONS

HR: conceiving and drafting the article, agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. TN: revising the article critically, final approval of the submitted version.

## ACKNOWLEDGMENTS

We thank Russ Hodge (Max Delbrück Center for Molecular Medicine, Berlin, Germany), Andreas Pohlmann, Jurjen Heij, Jason Millward (Berlin Ultrahigh Field Facility, Max Delbrück Center for Molecular Medicine, Berlin, Germany), and Nikoloz Sirmipilatz (Functional Imaging Laboratory, Deutsches Primatenzentrum, Göttingen, Germany) for valuable discussions, reading and commenting on the manuscript.

## REFERENCES

- Achermann, P., and Borbély, A. A. (1997). Low-frequency (<1 Hz) oscillations in the human sleep electroencephalogram. *Neuroscience* 81, 213–222. doi: 10.1016/s0306-4522(97)00186-3
- Adamczak, J. M., Farr, T. D., Seehafer, J. U., Kalthoff, D., and Hoehn, M. (2010). High field BOLD response to forepaw stimulation in the mouse. *NeuroImage* 51, 704–712. doi: 10.1016/j.neuroimage.2010.02.083
- Aedo-Jury, F., Schwalm, M., Hamzehpour, L., and Stroh, A. (2019). Brain states govern the spatio-temporal dynamics of resting state functional connectivity. *bioRxiv* [Preprint]. doi: 10.1101/832998
- Aggarwal, A., Brennan, C., Shortal, B., Contreras, D., Kelz, M. B., and Proekt, A. (2019). Coherence of visual-evoked  $\gamma$  oscillations is disrupted by propofol but preserved under equipotent doses of isoflurane. *Front. Syst. Neurosci.* 13:19. doi: 10.3389/fnsys.2019.00019
- Ahrens, E. T., and Dubowitz, D. J. (2001). Peripheral somatosensory fMRI in mouse at 11.7 T. *NMR Biomed.* 14, 318–324. doi: 10.1002/nbm.709
- Akeju, O., and Brown, E. N. (2017). Neural oscillations demonstrate that general anesthesia and sedative states are neurophysiologically distinct from sleep. *Curr. Opin. Neurobiol.* 44, 178–185. doi: 10.1016/j.conb.2017.04.011

- Akeju, O., Kim, S. E., Vazquez, R., Rhee, J., Pavone, K. J., Hobbs, L. E., et al. (2016a). Spatiotemporal dynamics of dexmedetomidine-induced electroencephalogram oscillations. *PLoS One* 11:e0163431. doi: 10.1371/journal.pone.0163431
- Akeju, O., Song, A. H., Hamilos, A. E., Pavone, K. J., Flores, F. J., Brown, E. N., et al. (2016b). Electroencephalogram signatures of ketamine anesthesia-induced unconsciousness. *Clin. Neurophysiol.* 127, 2414–2422. doi: 10.1016/j.clinph.2016.03.005
- Akeju, O., Loggia, M. L., Catana, C., Pavone, K. J., Vazquez, R., Rhee, J., et al. (2014a). Disruption of thalamic functional connectivity is a neural correlate of dexmedetomidine-induced unconsciousness. *Elife* 3:e04499. doi: 10.7554/eLife.04499
- Akeju, O., Westover, B., Pavone, K. J., Sampson, A. L., Hartnack, K. E., Brown, E. N., et al. (2014b). Effects of sevoflurane and propofol on frontal electroencephalogram power and coherence. *Anesthesiology* 121, 990–998. doi: 10.1097/aln.0000000000000436
- Aksenov, D. P., Li, L., Miller, M. J., Iordanescu, G., and Wyrwicz, A. M. (2015). Effects of anesthesia on BOLD signal and neuronal activity in the somatosensory cortex. *J. Cereb. Blood Flow Metab.* 35, 1819–1826. doi: 10.1038/jcbfm.2015.130
- Alheid, G. F., and McCrimmon, D. R. (2008). The chemical neuroanatomy of breathing. *Respir. Physiol. Neurobiol.* 164, 3–11. doi: 10.1016/j.resp.2008.07.014
- Alkire, M. T., Asher, C. D., Franciscus, A. M., and Hahn, E. L. (2009). Thalamic microinfusion of antibody to a voltage-gated potassium channel restores consciousness during anesthesia. *Anesthesiology* 110, 766–773. doi: 10.1097/aln.0b013e31819c461c
- Alkire, M. T., Hudetz, A. G., and Tononi, G. (2008). Consciousness and anesthesia. *Science* 322, 876–880. doi: 10.1126/science.1149213
- Alkire, M. T., and Miller, J. (2005). General anesthesia and the neural correlates of consciousness. *Prog. Brain Res.* 150, 229–244. doi: 10.1016/s0079-6123(05)50017-7
- Amzica, F. (2009). Basic physiology of burst-suppression. *Epilepsia* 50, 38–39. doi: 10.1111/j.1528-1167.2009.02345.x
- Amzica, F. (2015). What does burst suppression really mean? *Epilepsy Behav.* 49, 234–237. doi: 10.1016/j.yebeh.2015.06.012
- Amzica, F., and Steriade, M. (1997). The K-complex: its slow (<1 Hz) rhythmicity and relation to delta waves. *Neurology* 49, 952–959. doi: 10.1212/wnl.49.4.952
- Anderson, P. M., Jones, N., O'Brien, T. J., and Pinault, D. (2017). The N-Methyl D-Aspartate glutamate receptor antagonist ketamine disrupts the functional state of the corticothalamic pathway. *Cereb. Cortex* 27, 3172–3185. doi: 10.1093/cercor/bhw168
- Andrieu, G., Roth, B., Ousmane, L., Castaner, M., Petillot, P., Vallet, B., et al. (2009). The efficacy of intrathecal morphine with or without clonidine for postoperative analgesia after radical prostatectomy. *Anesth. Analg.* 108, 1954–1957. doi: 10.1213/ane.0b013e3181a30182
- Arena, A., Lamanna, J., Gemma, M., Ripamonti, M., Ravasio, G., Zimarino, V., et al. (2017). Linear transformation of the encoding mechanism for light intensity underlies the paradoxical enhancement of cortical visual responses by sevoflurane. *J. Physiol.* 595, 321–339. doi: 10.1113/jp272215
- Asanuma, C. (1992). Noradrenergic innervation of the thalamic reticular nucleus: a light and electron microscopic immunohistochemical study in rats. *J. Comp. Neurol.* 319, 299–311. doi: 10.1002/cne.903190209
- Aston-Jones, G., Rajkowski, J., and Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biol. Psychiatry* 46, 1309–1320. doi: 10.1016/s0006-3223(99)00140-7
- Astori, S., Wimmer, R. D., Prosser, H. M., Corti, C., Corsi, M., Liaudet, N., et al. (2011). The CaV3.3 calcium channel is the major sleep spindle pacemaker in thalamus. *Proc. Natl. Acad. Sci. U S A* 108, 13823–13828. doi: 10.1073/pnas.1105115108
- Bai, D., Pennefather, P. S., MacDonald, J. F., and Orser, B. A. (1999). The general anesthetic propofol slows deactivation and desensitization of GABA<sub>A</sub> receptors. *J. Neurosci.* 19, 10635–10646. doi: 10.1523/jneurosci.19-24-10635.1999
- Balsters, J. H., Zerbi, V., Sallet, J., Wenderoth, N., and Mars, R. B. (2019). Primate homologs of mouse cortico-striatal circuits. *bioRxiv* [Preprint]. doi: 10.1101/834481
- Baker, R., Gent, T. C., Yang, Q., Parker, S., Vyssotski, A. L., Wisden, W., et al. (2014). Altered activity in the central medial thalamus precedes changes in the neocortex during transitions into both sleep and propofol anesthesia. *J. Neurosci.* 34, 13326–13335. doi: 10.1523/JNEUROSCI.1519-14.2014
- Bakker, R., Tiesinga, P., and Köster, R. (2015). The scalable brain atlas: instant web-based access to public brain atlases and related content. *Neuroinformatics* 13, 353–366. doi: 10.1007/s12021-014-9258-x
- Baltes, C., Bosshard, S., Mueggler, T., Ratering, D., and Rudin, M. (2011). Increased blood oxygen level-dependent (BOLD) sensitivity in the mouse somatosensory cortex during electrical forepaw stimulation using a cryogenic radiofrequency probe. *NMR Biomed.* 24, 439–446. doi: 10.1002/nbm.1613
- Banks, M. I., Moran, N. S., Krause, B. M., Grady, S. M., Uhrlich, D. J., and Manning, K. A. (2017). Evidence that loss of consciousness under anesthesia is not associated with impaired stimulus representation in auditory cortex. *bioRxiv* [Preprint]. doi: 10.1101/213355
- Barttfeld, P., Uhrig, L., Sitt, J. D., Sigman, M., Jarraya, B., and Dehaene, S. (2015). Signature of consciousness in the dynamics of resting-state brain activity. *Proc. Natl. Acad. Sci. U S A* 112, 887–892. doi: 10.1073/pnas.1418031112
- Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J. M., Oostenveld, R., Dowdall, J. R., et al. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85, 390–401. doi: 10.1016/j.neuron.2014.12.018
- Baumgart, J. P., Zhou, Z.-Y., Hara, M., Cook, D. C., Hoppa, M. B., Ryan, T. A., et al. (2015). Isoflurane inhibits synaptic vesicle exocytosis through reduced Ca<sup>2+</sup> influx, not Ca<sup>2+</sup>-exocytosis coupling. *Proc. Natl. Acad. Sci. U S A* 112, 11959–11964. doi: 10.1073/pnas.1500525112
- Bazhenov, M., Timofeev, I., Steriade, M., and Sejnowski, T. J. (1999). Self-sustained rhythmic activity in the thalamic reticular nucleus mediated by depolarizing GABA<sub>A</sub> receptor potentials. *Nat. Neurosci.* 2, 168–174. doi: 10.1038/5729
- Beenhakker, M. P., and Huguenard, J. R. (2009). Neurons that fire together also conspire together: is normal sleep circuitry hijacked to generate epilepsy? *Neuron* 62, 612–632. doi: 10.1016/j.neuron.2009.05.015
- Belloy, M. E., Naeyaert, M., Abbas, A., Shah, D., Vanreusel, V., Van Audekerke, J., et al. (2018). Dynamic resting state fMRI analysis in mice reveals a set of Quasi-Periodic Patterns and illustrates their relationship with the global signal. *NeuroImage* 180, 463–484. doi: 10.1016/j.neuroimage.2018.01.075
- Bergmann, E., Zur, G., Bershadsky, G., and Kahn, I. (2016). The organization of mouse and human cortico-hippocampal networks estimated by intrinsic functional connectivity. *Cereb. Cortex* 26, 4497–4512. doi: 10.1093/cercor/bhw327
- Bertero, A., Liska, A., Pagani, M., Parolisi, R., Masferrer, M. E., Gritti, M., et al. (2018). Autism-associated 16p11.2 microdeletion impairs prefrontal functional connectivity in mouse and human. *Brain* 141, 2055–2065. doi: 10.1093/brain/awy111
- Bhattacharya, S., Le Cauchois, M. B., Iglesias, P. A., and Chen, Z. S. (2019). Thalamocortical network connectivity controls spatiotemporal dynamics of cortical and thalamic traveling waves. *bioRxiv* 6:7169. doi: 10.1101/780239
- Binda, P., and Gamlin, P. D. (2017). Renewed attention on the pupil light reflex. *Trends Neurosci.* 40, 455–457. doi: 10.1016/j.tins.2017.06.007
- Bittner, S., Gobel, K., and Meuth, S. G. (2019). *Cerebral Physiology. Oxford Textbook of Neuroscience and Anaesthesiology*. Ann Arbor, MI: Oxford University Press.
- Blain-Moraes, S., Lee, U., Ku, S., Noh, G., and Mashour, G. A. (2014). Electroencephalographic effects of ketamine on power, cross-frequency coupling and connectivity in the  $\alpha$  bandwidth. *Front. Syst. Neurosci.* 8:114. doi: 10.3389/fnsys.2014.00114
- Blain-Moraes, S., Tarnal, V., Vanini, G., Alexander, A., Rosen, D., Shortal, B., et al. (2015). Neurophysiological correlates of sevoflurane-induced unconsciousness. *Anesthesiology* 122, 307–316. doi: 10.1097/ALN.0000000000000482
- Blazquez Freches, G., Chavarrias, C., and Shemesh, N. (2018). BOLD-fMRI in the mouse auditory pathway. *NeuroImage* 165, 265–277. doi: 10.1016/j.neuroimage.2017.10.027
- Boly, M., Garrido, M. I., Gosseries, O., Bruno, M. A., Boveroux, P., Schnakers, C., et al. (2011). Preserved feedforward but impaired top-down processes in the vegetative state. *Science* 332, 858–862. doi: 10.1126/science.1202043
- Bonhomme, V., Staquet, C., Montupil, J., Defresne, A., Kirsch, M., Martial, C., et al. (2019). General anesthesia: a probe to explore consciousness. *Front. Syst. Neurosci.* 13:36. doi: 10.3389/fnsys.2019.00036



- Bonhomme, V., Vanhaudenhuyse, A., Demertzi, A., Bruno, M. A., Jaquet, O., Bahri, M. A., et al. (2016). Resting-state network-specific breakdown of functional connectivity during ketamine alteration of consciousness in volunteers. *Anesthesiology* 125, 873–888. doi: 10.1097/aln.0000000000001275
- Boon, J. A., and Milsom, W. K. (2008). NMDA receptor-mediated processes in the Parabrachial/Kolliker fuse complex influence respiratory responses directly and indirectly via changes in cortical activation state. *Respir. Physiol. Neurobiol.* 162, 63–72. doi: 10.1016/j.resp.2008.04.002
- Börger, C., Epstein, S., and Kopell, N. J. (2005). Background gamma rhythmicity and attention in cortical local circuits: a computational study. *Proc. Natl. Acad. Sci. U S A* 102, 7002–7007. doi: 10.1073/pnas.0502366102
- Bosshard, S. C., Baltes, C., Wyss, M. T., Mueggler, T., Weber, B., and Rudin, M. (2010). Assessment of brain responses to innocuous and noxious electrical forepaw stimulation in mice using BOLD fMRI. *Pain* 151, 655–663. doi: 10.1016/j.pain.2010.08.025
- Bosshard, S. C., Stuker, F., Von Deuster, C., Schroeter, A., and Rudin, M. (2015). BOLD fMRI of C-fiber mediated nociceptive processing in mouse brain in response to thermal stimulation of the forepaws. *PLoS One* 10:e0126513. doi: 10.1371/journal.pone.0126513
- Boveroux, P., Vanhaudenhuyse, A., Bruno, M. A., Noirhomme, Q., Lauwrick, S., Luxen, A., et al. (2010). Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. *Anesthesiology* 113, 1038–1053. doi: 10.1097/ALN.0b013e3181f697f5
- Bowery, N. G., Hudson, A. L., and Price, G. W. (1987). GABA<sub>A</sub> and GABA<sub>B</sub> receptor site distribution in the rat central nervous system. *Neuroscience* 20, 365–383. doi: 10.1016/0306-4522(87)90098-4
- Braun, J., and Mattia, M. (2010). Attractors and noise: twin drivers of decisions and multistability. *NeuroImage* 52, 740–751. doi: 10.1016/j.neuroimage.2009.12.126
- Breakspear, M. (2017). Dynamic models of large-scale brain activity. *Nat. Neurosci.* 20, 340–352. doi: 10.1038/nn.4497
- Bremer, F. (1935). Cerveau isole et Physiologie du sommeil. *C. R. Soc. Biol.* 118, 1235–1241.
- Brier, L. M., Landsness, E. C., Snyder, A. Z., Wright, P. W., Baxter, G. A., Bauer, A. Q., et al. (2019). Separability of calcium slow waves and functional connectivity during wake, sleep and anesthesia. *Neurophotonics* 6:035002. doi: 10.1117/1.nph.6.3.035002
- Brown, E. N., Lydic, R., and Schiff, N. D. (2010). General anesthesia, sleep and coma. *N. Engl. J. Med.* 363, 2638–2650. doi: 10.1056/NEJMra0808281
- Brown, P. L., Zanos, P., Wang, L., Elmer, G. I., Gould, T. D., and Shepard, P. D. (2018). Isoflurane but not halothane prevents and reverses helpless behavior: a role for eeg burst suppression?. *Int. J. Neuropsychopharmacol.* 21, 777–785. doi: 10.1093/ijnp/pyy029
- Brown, E. N., Pavone, K. J., and Naranjo, M. (2018a). Multimodal general anesthesia: theory and practice. *Anesth. Analg.* 127, 1246–1258. doi: 10.1213/ane.0000000000003668
- Brown, E. N., Purdon, P. L., Akeju, O., and An, J. (2018b). Using EEG markers to make inferences about anaesthetic-induced altered states of arousal. *Br. J. Anaesth.* 121, 325–327. doi: 10.1016/j.bja.2017.12.034
- Brown, E. N., Purdon, P. L., and Van Dort, C. J. (2011). General anesthesia and altered states of arousal: a systems neuroscience analysis. *Annu. Rev. Neurosci.* 34, 601–628. doi: 10.1146/annurev-neuro-060909-153200
- Buckwalter, J. A., Parvizi, J., Morecraft, R. J., and Van Hoesen, G. W. (2008). Thalamic projections to the posteromedial cortex in the macaque. *J. Comp. Neurol.* 507, 1709–1733. doi: 10.1002/cne.21647
- Bukhari, Q., Schroeter, A., Cole, D. M., and Rudin, M. (2017). Resting state fMRI in mice reveals anesthesia specific signatures of brain functional networks and their interactions. *Front. Neural Circuits* 11:5. doi: 10.3389/fncir.2017.00005
- Bullmore, E., and Sporns, O. (2012). The economy of brain network organization. *Nat. Rev. Neurosci.* 13, 336–349. doi: 10.1038/nrn3214
- Burke, M., Schwindt, W., Ludwig, U., Hennig, J., and Hoehn, M. (2000). Facilitation of electric forepaw stimulation-induced somatosensory activation in rats by additional acoustic stimulation: an fMRI investigation. *Magn. Reson. Med.* 44, 317–321. doi: 10.1002/1522-2594(200008)44:2<317::aid-mrm20>3.0.co;2-r
- Busche, M. A., Kekuš, M., Adelsberger, H., Noda, T., Förstl, H., Nelken, I., et al. (2015). Rescue of long-range circuit dysfunction in Alzheimer's disease models. *Nat. Neurosci.* 18, 1623–1630. doi: 10.1038/nn.4137
- Buzsáki, G., and Chrobak, J. J. (1995). Temporal structure in spatially organized neuronal ensembles: a role for interneuronal networks. *Curr. Opin. Neurobiol.* 5, 504–510. doi: 10.1016/0959-4388(95)80012-3
- Buzsáki, G., and Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nat. Neurosci.* 16, 130–138. doi: 10.1038/nn.3304
- Campagna, J. A., Miller, K. W., and Forman, S. A. (2003). Mechanisms of actions of inhaled anesthetics. *N. Engl. J. Med.* 348, 2110–2124. doi: 10.1056/nejmra021261
- Cao, R., Li, J., Ning, B., Sun, N., Wang, T., Zuo, Z., et al. (2017). Functional and oxygen-metabolic photoacoustic microscopy of the awake mouse brain. *NeuroImage* 150, 77–87. doi: 10.1016/j.neuroimage.2017.01.049
- Carter, M. E., Yizhar, O., Chikahisa, S., Nguyen, H., Adamantidis, A., Nishino, S., et al. (2010). Tuning arousal with optogenetic modulation of locus coeruleus neurons. *Nat. Neurosci.* 13, 1526–1533. doi: 10.1038/nn.2682
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* 5:198ra105. doi: 10.1126/scitranslmed.3006294
- Casarotto, S., Rosanova, M., Gosseries, O., Boly, M., Massimini, M., and Sarasso, S. (2016). “Exploring the neurophysiological correlates of loss and recovery of consciousness: perturbational complexity,” in *Brain Function and Responsiveness in Disorders of Consciousness*, eds M. M. Monti and W. G. Sannita (Cham: Springer International Publishing), 93–104.
- Chamadia, S., Pedemonte, J. C., Hahm, E. Y., Mekonnen, J., Ibalá, R., Gitlin, J., et al. (2019). Delta oscillations phase limit neural activity during sevoflurane anesthesia. *Commun. Biol.* 2:415. doi: 10.1038/s42003-019-0664-3
- Chan, R. W., Leong, A. T. L., Ho, L. C., Gao, P. P., Wong, E. C., Dong, C. M., et al. (2017). Low-frequency hippocampal-cortical activity drives brain-wide resting-state functional MRI connectivity. *Proc. Natl. Acad. Sci. U S A* 114, E6972–E6981. doi: 10.1073/pnas.1703309114
- Chen, X., Tong, C., Han, Z., Zhang, K., Bo, B., Feng, Y., et al. (2020). Sensory evoked fMRI paradigms in awake mice. *NeuroImage* 204:116242. doi: 10.1016/j.neuroimage.2019.116242
- Chao, Z. C., Takaura, K., Wang, L., Fujii, N., and Dehaene, S. (2018). Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. *Neuron* 100, 1252.e3–1266.e3. doi: 10.1016/j.neuron.2018.10.004
- Chauvette, S., Crochet, S., Volgushev, M., and Timofeev, I. (2011). Properties of slow oscillation during slow-wave sleep and anesthesia in cats. *J. Neurosci.* 31, 14998–15008. doi: 10.1523/JNEUROSCI.2339-11.2011
- Chen, Z., Wimmer, R. D., Wilson, M. A., and Halassa, M. M. (2016). Thalamic circuit mechanisms link sensory processing in sleep and attention. *Front. Neural Circuits* 9:83. doi: 10.3389/fncir.2015.00083
- Ching, S., and Brown, E. N. (2014). Modeling the dynamical effects of anesthesia on brain circuits. *Curr. Opin. Neurobiol.* 25, 116–122. doi: 10.1016/j.conb.2013.12.011
- Ching, S., Cimenser, A., Purdon, P. L., Brown, E. N., and Kopell, N. J. (2010). Thalamocortical model for a propofol-induced  $\alpha$ -rhythm associated with loss of consciousness. *Proc. Natl. Acad. Sci. U S A* 107, 22665–22670. doi: 10.1073/pnas.1017069108
- Ching, S., Purdon, P. L., Vijayan, S., Kopell, N. J., and Brown, E. N. (2012). A neurophysiological-metabolic model for burst suppression. *Proc. Natl. Acad. Sci. U S A* 109, 3095–3100. doi: 10.1073/pnas.1121461109
- Chiu, T. H., Chen, M. J., Yang, Y. R., Yang, J. J., and Tang, F. I. (1995). Action of dexmedetomidine on rat locus coeruleus neurones: intracellular recording *in vitro*. *Eur. J. Pharmacol.* 285, 261–268. doi: 10.1016/0014-2999(95)00417-j
- Chuang, K. H., and Nasrallah, F. A. (2017). Functional networks and network perturbations in rodents. *NeuroImage* 163, 419–436. doi: 10.1016/j.neuroimage.2017.09.038
- Cimenser, A., Purdon, P. L., Pierce, E. T., Walsh, J. L., Salazar-Gomez, A. F., Harrell, P. G., et al. (2011). Tracking brain states under general anesthesia by using global coherence analysis. *Proc. Natl. Acad. Sci. U S A* 108, 8832–8837. doi: 10.1073/pnas.1017041108



- Claassen, V. (1994). "Neglected factors in pharmacology and neuroscience research: biopharmaceutics, animal characteristics, maintenance, testing conditions," in *Techniques in the Behavioral and Neural Sciences*, Vol. 12, ed. J. P. Huston (New York, NY: Elsevier).
- Compte, A., Sanchez-Vives, M. V., McCormick, D. A., and Wang, X. J. (2003). Cellular and network mechanisms of slow oscillatory activity (<1 Hz) and wave propagations in a cortical network model. *J. Neurophysiol.* 89, 2707–2725. doi: 10.1152/jn.00845.2002
- Conner, C. R., Ellmore, T. M., Peters, T. A., DiSano, M. A., and Tandon, N. (2011). Variability of the relationship between electrophysiology and BOLD-fMRI across cortical regions in humans. *J. Neurosci.* 31, 12855–12865. doi: 10.1523/JNEUROSCI.1457-11.2011
- Correa-Sales, C., Rabin, B. C., and Maze, M. (1992). A hypnotic response to dexmedetomidine, an  $\alpha$  2 agonist, is mediated in the locus coeruleus in rats. *Anesthesiology* 76, 948–952. doi: 10.1097/0000542-199206000-00013
- Coull, J. T. (1998). Neural correlates of attention and arousal: insights from electrophysiology, functional neuroimaging and psychopharmacology. *Prog. Neurobiol.* 55, 343–361. doi: 10.1016/s0301-0082(98)00011-2
- Coulter, K. L., Perier, F., Radeke, C. M., and Vandenberg, C. A. (1995). Identification and molecular localization of a pH-sensing domain for the inward rectifier potassium channel HIR. *Neuron* 15, 1157–1168. doi: 10.1016/0896-6273(95)90103-5
- Craig, A. D. (2013). Cooling, pain and other feelings from the body in relation to the autonomic nervous system. *Handb. Clin. Neurol.* 117, 103–109. doi: 10.1016/b978-0-444-53491-0.00009-2
- Crandall, S. R., Govindaiah, G., and Cox, C. L. (2010). Low-threshold  $\text{Ca}^{2+}$  current amplifies distal dendritic signaling in thalamic reticular neurons. *J. Neurosci.* 30, 15419–15429. doi: 10.1523/JNEUROSCI.3636-10.2010
- Crunelli, V., David, F., Lörincz, M. L., and Hughes, S. W. (2015). The thalamocortical network as a single slow wave-generating unit. *Curr. Opin. Neurobiol.* 31, 72–80. doi: 10.1016/j.conb.2014.09.001
- Crunelli, V., and Hughes, S. W. (2010). The slow (<1 Hz) rhythm of non-REM sleep: a dialogue between three cardinal oscillators. *Nat. Neurosci.* 13, 9–17. doi: 10.1038/nn.2445
- Cunningham, M. O., Pervouchine, D. D., Racca, C., Kopell, N. J., Davies, C. H., Jones, R. S., et al. (2006). Neuronal metabolism governs cortical network response state. *Proc. Natl. Acad. Sci. U S A* 103, 5597–5601. doi: 10.1073/pnas.060604103
- Dagal, A., and Lam, A. M. (2009). Cerebral autoregulation and anesthesia. *Curr. Opin. Anaesthesiol.* 22, 547–552. doi: 10.1097/ACO.0b013e32833020be
- Dang-Vu, T. T., McKinney, S. M., Buxton, O. M., Solet, J. M., and Ellenbogen, J. M. (2010). Spontaneous brain rhythms predict sleep stability in the face of noise. *Curr. Biol.* 20, R626–R627. doi: 10.1016/j.cub.2010.06.032
- Davis, K. D., Flor, H., Greely, H. T., Iannetti, G. D., Mackey, S., Ploner, M., et al. (2017). Brain imaging tests for chronic pain: medical, legal and ethical issues and recommendations. *Nat Rev Neurol* 13, 624–638. doi: 10.1038/nrneurol.2017.122
- de la Salle, S., Choueiry, J., Shah, D., Bowers, H., McIntosh, J., Ilivitsky, V., et al. (2016). Effects of ketamine on resting-state EEG activity and their relationship to perceptual/dissociative symptoms in healthy humans. *Front. Pharmacol.* 7, 348–348. doi: 10.3389/fphar.2016.00348
- Deleuze, C., and Huguenard, J. R. (2016). Two classes of excitatory synaptic responses in rat thalamic reticular neurons. *J. Neurophysiol.* 116, 995–1011. doi: 10.1152/jn.01121.2015
- Desai, M., Kahn, I., Knoblich, U., Bernstein, J., Atallah, H., Yang, A., et al. (2011). Mapping brain networks in awake mice using combined optical neural control and fMRI. *J. Neurophysiol.* 105, 1393–1405. doi: 10.1152/jn.00828.2010
- Destexhe, A., and Contreras, D. (2006). Neuronal computations with stochastic network states. *Science* 314, 85–90. doi: 10.1126/science.1127241
- Destexhe, A., Contreras, D., Sejnowski, T. J., and Steriade, M. (1994). A model of spindle rhythmicity in the isolated thalamic reticular nucleus. *J. Neurophysiol.* 72, 803–818. doi: 10.1152/jn.1994.72.2.803
- Destexhe, A., Contreras, D., Steriade, M., Sejnowski, T. J., and Huguenard, J. R. (1996). *In vivo*, *in vitro*, and computational analysis of dendritic calcium currents in thalamic reticular neurons. *J. Neurosci.* 16, 169–185. doi: 10.1523/JNEUROSCI.16-01-00169.1996
- Destexhe, A., Hughes, S. W., Rudolph, M., and Crunelli, V. (2007). Are corticothalamic 'up' states fragments of wakefulness? *Trends Neurosci.* 30, 334–342. doi: 10.1016/j.tins.2007.04.006
- Detsch, O., Kochs, E., Siemers, M., Bromm, B., and Vahle-Hinz, C. (2002). Increased responsiveness of cortical neurons in contrast to thalamic neurons during isoflurane-induced EEG bursts in rats. *Neurosci. Lett.* 317, 9–12. doi: 10.1016/s0304-3940(01)02419-3
- Detsch, O., Vahle-Hinz, C., Kochs, E., Siemers, M., and Bromm, B. (1999). Isoflurane induces dose-dependent changes of thalamic somatosensory information transfer. *Brain Res.* 829, 77–89. doi: 10.1016/s0006-8993(99)01341-4
- Devonshire, I. M., Grandy, T. H., Domett, E. J., and Greenfield, S. A. (2010). Effects of urethane anaesthesia on sensory processing in the rat barrel cortex revealed by combined optical imaging and electrophysiology. *Eur. J. Neurosci.* 32, 786–797. doi: 10.1111/j.1460-9568.2010.07322.x
- Dopfel, D., and Zhang, N. (2018). Mapping stress networks using functional magnetic resonance imaging in awake animals. *Neurobiol. Stress* 9, 251–263. doi: 10.1016/j.ynstr.2018.06.002
- D'Souza, R. D., Meier, A. M., Bista, P., Wang, Q., and Burkhalter, A. (2016). Recruitment of inhibition and excitation across mouse visual cortex depends on the hierarchy of interconnecting areas. *Elife* 5:e19332. doi: 10.7554/eLife.19332
- Duling, B. R., and Berne, R. M. (1970). Longitudinal gradients in periarteriolar oxygen tension. A possible mechanism for the participation of oxygen in local regulation of blood flow. *Circ. Res.* 27, 669–678. doi: 10.1161/01.res.27.5.669
- Echlin, F. A., Arnett, V., and Zoll, J. (1952). Paroxysmal high voltage discharges from isolated and partially isolated human and animal cerebral cortex. *Electroencephalogr. Clin. Neurophysiol.* 4, 147–164. doi: 10.1016/0013-4694(52)90004-7
- Ecker, A. S., Denfield, G. H., Bethge, M., and Tolias, A. S. (2016). On the structure of neuronal population activity under fluctuations in attentional state. *J. Neurosci.* 36, 1775–1789. doi: 10.1523/JNEUROSCI.2044-15.2016
- Eger, E. I. II., Saidman, L. J., and Brandstater, B. (1965). Minimum alveolar anesthetic concentration: a standard of anesthetic potency. *Anesthesiology* 26, 756–763. doi: 10.1097/0000542-196511000-00010
- Eriskens, S., Vaiceliunaite, A., Jurjut, O., Fiorini, M., Katzner, S., and Busse, L. (2014). Effects of locomotion extend throughout the mouse early visual system. *Curr. Biol.* 24, 2899–2907. doi: 10.1016/j.cub.2014.10.045
- Erö, C., Gewaltig, M.-O., Keller, D., and Markram, H. (2018). A cell atlas for the mouse brain. *Front. Neuroinform.* 12:84. doi: 10.3389/fninf.2018.00084
- Eschenko, O., Mölle, M., Born, J., and Sara, S. J. (2006). Elevated sleep spindle density after learning or after retrieval in rats. *J. Neurosci.* 26, 12914–12920. doi: 10.1523/JNEUROSCI.3175-06.2006
- Fernandez, L. M. J., Pellegrini, C., Vantomme, G., Béard, E., Lüthi, A., and Astori, S. (2017). Cortical afferents onto the nucleus Reticularis thalami promote plasticity of low-threshold excitability through GluN2C-NMDARs. *Sci. Rep.* 7:12271. doi: 10.1038/s41598-017-12552-8
- Ferrarelli, F., Massimini, M., Sarasso, S., Casali, A., Riedner, B. A., Angelini, G., et al. (2010). Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proc. Natl. Acad. Sci. U S A* 107, 2681–2686. doi: 10.1073/pnas.0913008107
- Ferrari, L., Turrini, G., Crestan, V., Bertani, S., Cristofori, P., Bifone, A., et al. (2012). A robust experimental protocol for pharmacological fMRI in rats and mice. *J. Neurosci. Methods* 204, 9–18. doi: 10.1016/j.jneumeth.2011.10.020
- Ferron, J.-F., Kroeger, D., Chever, O., and Amzica, F. (2009). Cortical inhibition during burst suppression induced with isoflurane anesthesia. *J. Neurosci.* 29, 9850–9860. doi: 10.1523/JNEUROSCI.5176-08.2009
- Flores, F. J., Hartnack, K. E., Fath, A. B., Kim, S.-E., Wilson, M. A., Brown, E. N., et al. (2017). Thalamocortical synchronization during induction and emergence from propofol-induced unconsciousness. *Proc. Natl. Acad. Sci. U S A* 114, E6660–E6668. doi: 10.1073/pnas.1700148114
- Franceschini, M. A., Radhakrishnan, H., Thakur, K., Wu, W., Ruvinskaya, S., Carp, S., et al. (2010). The effect of different anesthetics on neurovascular coupling. *NeuroImage* 51, 1367–1377. doi: 10.1016/j.neuroimage.2010.03.060

- Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* 47, D766–D773. doi: 10.1093/nar/gky955
- Franks, N. P. (2008). General anaesthesia: from molecular targets to neuronal pathways of sleep and arousal. *Nat. Rev. Neurosci.* 9, 370–386. doi: 10.1038/nrn2372
- Friedman, E. B., Sun, Y., Moore, J. T., Hung, H. T., Meng, Q. C., Perera, P., et al. (2010). A conserved behavioral state barrier impedes transitions between anesthetic-induced unconsciousness and wakefulness: evidence for neural inertia. *PLoS One* 5:e11903. doi: 10.1371/journal.pone.0011903
- Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron* 88, 220–235. doi: 10.1016/j.neuron.2015.09.034
- Fries, P., Nikolić, D., and Singer, W. (2007). The  $\gamma$  cycle. *Trends Neurosci.* 30, 309–316. doi: 10.1016/j.tins.2007.05.005
- Fu, B., Yu, T., Yuan, J., Gong, X., and Zhang, M. (2017). Noradrenergic transmission in the central medial thalamic nucleus modulates the electroencephalographic activity and emergence from propofol anesthesia in rats. *J. Neurochem.* 140, 862–873. doi: 10.1111/jnc.13939
- Fuentealba, P., and Steriade, M. (2005). The reticular nucleus revisited: intrinsic and network properties of a thalamic pacemaker. *Prog. Neurobiol.* 75, 125–141. doi: 10.1016/j.pneurobio.2005.01.002
- Fukuda, M., Vazquez, A. L., Zong, X., and Kim, S. G. (2013). Effects of the  $\alpha_2$ -adrenergic receptor agonist dexmedetomidine on neural, vascular and BOLD fMRI responses in the somatosensory cortex. *Eur. J. Neurosci.* 37, 80–95. doi: 10.1111/ejn.12024
- Fulcher, B. D., and Fornito, A. (2016). A transcriptional signature of hub connectivity in the mouse connectome. *Proc. Natl. Acad. Sci. U S A* 113, 1435–1440. doi: 10.1073/pnas.1513302113
- Fulcher, B. D., Murray, J. D., Zerbi, V., and Wang, X.-J. (2019). Multimodal gradients across mouse cortex. *Proc. Natl. Acad. Sci. U S A* 116, 4689–4695. doi: 10.1073/pnas.1814144116
- Fuller, P. M., Sherman, D., Pedersen, N. P., Saper, C. B., and Lu, J. (2011). Reassessment of the structural basis of the ascending arousal system. *J. Comp. Neurol.* 519, 933–956. doi: 10.1002/cne.22559
- Furth, K. E., McCoy, A. J., Dodge, C., Walters, J. R., Buonanno, A., and Delaville, C. (2017). Neuronal correlates of ketamine and walking induced  $\gamma$  oscillations in the medial prefrontal cortex and mediodorsal thalamus. *PLoS One* 12:e0186732. doi: 10.1371/journal.pone.0186732
- Gămănuț, R., Kennedy, H., Toroczkai, Z., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., et al. (2018). The mouse cortical connectome, characterized by an ultra-dense cortical graph, maintains specificity by distinct connectivity profiles. *Neuron* 97, 698.e10–715.e10. doi: 10.1016/j.neuron.2017.12.037
- Giorgi, A., Migliarini, S., Galbusera, A., Maddaloni, G., Mereu, M., Margiani, G., et al. (2017). Brainwide mapping of endogenous serotonergic transmission via chemogenetic-fMRI. *Cell Rep.* 21, 910–918. doi: 10.1016/j.celrep.2017.09.087
- Glickfeld, L. L., and Olsen, S. R. (2017). Higher-order areas of the mouse visual cortex. *Annu. Rev. Vis. Sci.* 3, 251–273. doi: 10.1146/annurev-vision-102016-061331
- Golkowski, D., Ranft, A., Kiel, T., Riedl, V., Kohl, P., Rohrer, G., et al. (2017). Coherence of BOLD signal and electrical activity in the human brain during deep sevoflurane anesthesia. *Brain Behav.* 7:e00679. doi: 10.1002/brb3.679
- Golkowski, D., Larroque, S. K., Vanhaudenhuyse, A., Plenevaux, A., Boly, M., Di Perri, C., et al. (2019). Changes in whole brain dynamics and connectivity patterns during sevoflurane and propofol-induced unconsciousness identified by functional magnetic resonance imaging. *Anesthesiology* 130, 898–911. doi: 10.1097/aln.0000000000002704
- Gozzi, A., Ceolin, L., Schwarz, A., Reese, T., Bertani, S., Crestan, V., et al. (2007). A multimodality investigation of cerebral hemodynamics and autoregulation in pharmacological MRI. *Magn. Reson. Imaging* 25, 826–833. doi: 10.1016/j.mri.2007.03.003
- Gozzi, A., and Schwarz, A. J. (2016). Large-scale functional connectivity networks in the rodent brain. *NeuroImage* 127, 496–509. doi: 10.1016/j.neuroimage.2015.12.017
- Grandjean, J., Canella, C., Anckaerts, C., Ayranci, G., Bougacha, S., Bienert, T., et al. (2020). Common functional networks in the mouse brain revealed by multi-centre resting-state fMRI analysis. *NeuroImage* 116278. doi: 10.1016/j.neuroimage.2019.116278
- Grandjean, J., Canella, C., Anckaerts, C., Ayranci, G., Bougacha, S., Bienert, T., et al. (2019a). Common functional networks in the mouse brain revealed by multi-centre resting-state fMRI analysis. *bioRxiv* [Preprint]
- Grandjean, J., Corcoba, A., Kahn, M. C., Upton, A. L., Deneris, E. S., Seifritz, E., et al. (2019b). A brain-wide functional map of the serotonergic responses to acute stress and fluoxetine. *Nat. Commun.* 10:350. doi: 10.1038/s41467-018-08256-w
- Grandjean, J., Preti, M. G., Bolton, T. A. W., Buerge, M., Seifritz, E., and Pryce, C. R. et al. (2017a). Dynamic reorganization of intrinsic functional networks in the mouse brain. *NeuroImage* 152, 497–508. doi: 10.1016/j.neuroimage.2017.03.026
- Grandjean, J., Zerbi, V., Balsters, J. H., Wenderoth, N., and Rudin, M. (2017b). Structural basis of large-scale functional connectivity in the mouse. *J. Neurosci.* 37, 8092–8101. doi: 10.1523/JNEUROSCI.0438-17.2017
- Grandjean, J., Schroeter, A., Batata, I., and Rudin, M. (2014). Optimization of anesthesia protocol for resting-state fMRI in mice based on differential effects of anesthetics on functional connectivity patterns. *NeuroImage* 102, 838–847. doi: 10.1016/j.neuroimage.2014.08.043
- Gray, C. M., and McCormick, D. A. (1996). Chattering cells: superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. *Science* 274, 109–113. doi: 10.1126/science.274.5284.109
- Greenberg, D. S., Houweling, A. R., and Kerr, J. N. (2008). Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat. Neurosci.* 11, 749–751. doi: 10.1038/nn.2140
- Grunze, H. C. R., Rainnie, D. G., Hasselmo, M. E., Barkai, E., Hearn, E. F., McCarley, R. W., et al. (1996). NMDA-dependent modulation of CA1 local circuit inhibition. *J. Neurosci.* 16, 2034–2043. doi: 10.1523/JNEUROSCI.16-06-02034.1996
- Guidera, J. A., Taylor, N. E., Lee, J. T., Vlasov, K. Y., Pei, J., Stephen, E. P., et al. (2017). Sevoflurane induces coherent slow-delta oscillations in rats. *Front. Neural Circuits* 11:36. doi: 10.3389/fncir.2017.00036
- Guilfoyle, D. N., Gerum, S. V., Sanchez, J. L., Balla, A., Sershen, H., Javitt, D. C., et al. (2013). Functional connectivity fMRI in mouse brain at 7T using isoflurane. *J. Neurosci. Methods* 214, 144–148. doi: 10.1016/j.jneumeth.2013.01.019
- Guldenmund, P., Vanhaudenhuyse, A., Sanders, R., Sleight, J., Bruno, M. A., Demertzi, A., et al. (2017). Brainfunctional connectivity differentiates dexmedetomidine from propofol and natural sleep. *Br. J. Anaesth.* 119, 674–684. doi: 10.1093/bja/aex257
- Gunduz-Bruce, H. (2009). The acute effects of NMDA antagonism: from the rodent to the human brain. *Brain Res. Rev.* 60, 279–286. doi: 10.1016/j.brainresrev.2008.07.006
- Gutierrez-Barragan, D., Basson, M. A., Panzeri, S., and Gozzi, A. (2018). Oscillatory brain states govern spontaneous fMRI network dynamics. *bioRxiv* [Preprint]. doi: 10.1101/393389
- Hahn, T. T. G., Sakmann, B., and Mehta, M. R. (2006). Phase-locking of hippocampal interneurons' membrane potential to neocortical up-down states. *Nat. Neurosci.* 9, 1359–1361. doi: 10.1038/nn1788
- Haider, B., Häusser, M., and Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. *Nature* 493, 97–100. doi: 10.1038/nature11665
- Halassa, M. M., Siegle, J. H., Ritt, J. T., Ting, J. T., Feng, G., and Moore, C. I. (2011). Selective optical drive of thalamic reticular nucleus generates thalamic bursts and cortical spindles. *Nat. Neurosci.* 14, 1118–1120. doi: 10.1038/nn.2880
- Hall, C. N., Reynell, C., Gesslein, B., Hamilton, N. B., Mishra, A., Sutherland, B. A., et al. (2014). Capillary pericytes regulate cerebral blood flow in health and disease. *Nature* 508, 55–60. doi: 10.1038/nature13165
- Halley, A. C., and Krubitzer, L. (2019). Not all cortical expansions are the same: the coevolution of the neocortex and the dorsal thalamus in mammals. *Curr. Opin. Neurobiol.* 56, 78–86. doi: 10.1016/j.conb.2018.12.003
- Hamilton, N. B., Attwell, D., and Hall, C. N. (2010). Pericyte-mediated regulation of capillary diameter: a component of neurovascular coupling in health and disease. *Front. Neuroenergetics* 2:5. doi: 10.3389/fnene.2010.00005
- Hamner, J. W., Ishibashi, K., and Tan, C. O. (2019). Revisiting human cerebral blood flow responses to augmented blood pressure oscillations. *J. Physiol.* 597, 1553–1564. doi: 10.1113/jp277321

- Han, Z., Chen, W., Chen, X., Zhang, K., Tong, C., Zhang, X., et al. (2019). Awake and behaving mouse fMRI during Go/No-Go task. *NeuroImage* 188, 733–742. doi: 10.1016/j.neuroimage.2019.01.002
- Han, H.-B., Lee, K. E., and Choi, J. H. (2019). Functional dissociation of  $\theta$  oscillations in the frontal and visual cortices and their long-range network during sustained attention. *eNeuro* 6:ENEURO.0248-19.2019. doi: 10.1523/ENEURO.0248-19.2019
- Harris, K. D., and Thiele, A. (2011). Cortical state and attention. *Nat. Rev. Neurosci.* 12, 509–523. doi: 10.1038/nrn3084
- Hartikainen, K. M., Rorarius, M., Peräkylä, J. J., Laippala, P. J., and Jäntti, V. (1995). Cortical reactivity during isoflurane burst-suppression anesthesia. *Anesth. Analg.* 81, 1223–1228. doi: 10.1213/00000539-199512000-00018
- Havlik, M. (2017). Missing piece of the puzzle in the science of consciousness: resting state and endogenous correlates of consciousness. *Conscious. Cogn.* 49, 70–85. doi: 10.1016/j.concog.2017.01.006
- Hayat, H., Regev, N., Matosevich, N., Sales, A., Paredes-Rodriguez, E., Krom, A. J., et al. (2019). Locus-coeruleus norepinephrine activity gates sensory-evoked awakenings from sleep. *bioRxiv* [Preprint]. doi: 10.1101/539502
- Hayton, S. M., Kriss, A., and Muller, D. P. (1999). Comparison of the effects of four anaesthetic agents on somatosensory evoked potentials in the rat. *Lab. Anim.* 33, 243–251. doi: 10.1258/002367799780578219
- Heindl-Erdmann, C., Axmann, R., Kreitz, S., Zwerina, J., Penninger, J., Schett, G., et al. (2010). Combining functional magnetic resonance imaging with mouse genomics: new options in pain research. *Neuroreport* 21, 29–33. doi: 10.1097/wnr.0b013e3283324faf
- Hemmings, H. C. Jr., Akabas, M. H., Goldstein, P. A., Trudell, J. R., Orser, B. A., and Harrison, N. L. (2005). Emerging molecular mechanisms of general anesthetic action. *Trends Pharmacol. Sci.* 26, 503–510. doi: 10.1016/j.tips.2005.08.006
- Hemmings, H. C. Jr., Riegelhaupt, P. M., Kelz, M. B., Solt, K., Eckenhoof, R. G., Orser, B. A., et al. (2019). Towards a comprehensive understanding of anesthetic mechanisms of action: a decade of discovery. *Trends Pharmacol. Sci.* 40, 464–481. doi: 10.1016/j.tips.2019.05.001
- Henry, C. E., and Scoville, W. B. (1952). Suppression-burst activity from isolated cerebral cortex in man. *Electroencephalogr. Clin. Neurophysiol.* 4, 1–22. doi: 10.1016/0013-4694(52)90027-8
- Hentschke, H., Raz, A., Krause, B. M., Murphy, C. A., and Banks, M. I. (2017). Disruption of cortical network activity by the general anaesthetic isoflurane. *Br. J. Anaesth.* 119, 685–696. doi: 10.1093/bja/aex199
- Herrera, C. G., Cadavieco, M. C., Jegu, S., Ponomarenko, A., Korotkova, T., and Adamantidis, A. (2016). Hypothalamic feedforward inhibition of thalamocortical network controls arousal and consciousness. *Nat. Neurosci.* 19, 290–298. doi: 10.1038/nn.4209
- Hess, A., Axmann, R., Rech, J., Finzel, S., Heindl, C., Kreitz, S., et al. (2011). From the cover: blockade of TNF- $\alpha$  rapidly inhibits pain responses in the central nervous system. *Proc. Natl. Acad. Sci. U S A* 108, 3731–3736. doi: 10.1073/pnas.1011774108
- Hirsch, N., and Taylor, C. (2010). Pharmacological and pathological modulation of cerebral physiology. *Anaesth. Intens. Care Med.* 11, 349–354. doi: 10.1016/j.mpaic.2010.05.011
- Hofman, M. A. (2014). Evolution of the human brain: when bigger is better. *Front. Neuroanat.* 8:15. doi: 10.3389/fnana.2014.00015
- Homayoun, H., and Moghaddam, B. (2007). NMDA receptor hypofunction produces opposite effects on prefrontal cortex interneurons and pyramidal neurons. *J. Neurosci.* 27, 11496–11500. doi: 10.1523/JNEUROSCI.2213-07.2007
- Hoover, W. B., and Vertes, R. P. (2007). Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Struct. Funct.* 212, 149–179. doi: 10.1007/s00429-007-0150-4
- Hopfield, J. J. (1995). Pattern recognition computation using action potential timing for stimulus representation. *Nature* 376, 33–36. doi: 10.1038/376033a0
- Horowitz, S. G., Fukunaga, M., de Zwart, J. A., van Gelderen, P., Fulton, S. C., Balkin, T. J., et al. (2008). Low frequency BOLD fluctuations during resting wakefulness and light sleep: a simultaneous EEG-fMRI study. *Hum. Brain Mapp.* 29, 671–682. doi: 10.1002/hbm.20428
- Hoyer, C., Gass, N., Weber-Fahr, W., and Sartorius, A. (2014). Advantages and challenges of small animal magnetic resonance imaging as a translational tool. *Neuropsychobiology* 69, 187–201. doi: 10.1159/000360859
- Huang, W., Plyka, I., Li, H., Eisenstein, E. M., Volkow, N. D., and Springer, C. S. Jr. (1996). Magnetic resonance imaging (MRI) detection of the murine brain response to light: temporal differentiation and negative functional MRI changes. *Proc. Natl. Acad. Sci. U S A* 93, 6037–6042. doi: 10.1073/pnas.93.12.6037
- Huang, Z., Wang, Z., Zhang, J., Dai, R., Wu, J., Li, Y., et al. (2014). Altered temporal variance and neural synchronization of spontaneous brain activity in anesthesia. *Hum. Brain Mapp.* 35, 5368–5378. doi: 10.1002/hbm.22556
- Huber, L., Handwerker, D. A., Jangraw, D. C., Chen, G., Hall, A., Stüber, C., et al. (2017). High-resolution CBV-fMRI allows mapping of laminar activity and connectivity of cortical input and output in human M1. *Neuron* 96, 1253.e7–1263.e7. doi: 10.1016/j.neuron.2017.11.005
- Hudetz, A. G. (2002). Effect of volatile anesthetics on interhemispheric EEG cross-approximate entropy in the rat. *Brain Res.* 954, 123–131. doi: 10.1016/s0006-8993(02)03358-9
- Hudetz, A. G., and Imas, O. A. (2007). Burst activation of the cerebral cortex by flash stimuli during isoflurane anesthesia in rats. *Anesthesiology* 107, 983–991. doi: 10.1097/01.anes.0000291471.80659.55
- Hudetz, A. G., and Mashour, G. A. (2016). Disconnecting consciousness: is there a common anesthetic end point? *Anesth. Analg.* 123, 1228–1240. doi: 10.1213/ane.0000000000001353
- Hudetz, A. G., Wood, J. D., and Kampine, J. P. (2003). Cholinergic reversal of isoflurane anesthesia in rats as measured by cross-approximate entropy of the electroencephalogram. *Anesthesiology* 99, 1125–1131. doi: 10.1097/0000542-200311000-00019
- Hudson, A. E. (2017). Metastability of neuronal dynamics during general anesthesia: time for a change in our assumptions? *Front. Neural Circuits* 11:58. doi: 10.3389/fncir.2017.00058
- Hudson, A. E., Calderon, D. P., Pfaff, D. W., and Proekt, A. (2014). Recovery of consciousness is mediated by a network of discrete metastable activity states. *Proc. Natl. Acad. Sci. U S A* 111, 9283–9288. doi: 10.1073/pnas.1408296111
- Hughes, R. J. (1986). Extreme stereotypy in the burst suppression pattern. *Clin. Electroencephalogr.* 17, 162–168.
- Hutchison, R. M., Hutchison, M., Manning, K. Y., Menon, R. S., and Everling, S. (2014). Isoflurane induces dose-dependent alterations in the cortical connectivity profiles and dynamic properties of the brain's functional architecture. *Hum. Brain Mapp.* 35, 5754–5775. doi: 10.1002/hbm.22583
- Huupponen, E., Maksimow, A., Lapinlampi, P., Särkelä, M., Saastamoinen, A., Snapir, A., et al. (2008). Electroencephalogram spindle activity during dexmedetomidine sedation and physiological sleep. *Acta Anaesthesiol. Scand.* 52, 289–294. doi: 10.1111/j.1399-6576.2007.01537.x
- Idiyatullin, D., Corum, C. A., and Garwood, M. (2015). Multi-band-SWIFT. *J. Magn. Reson.* 251, 19–25. doi: 10.1016/j.jmr.2014.11.014
- Idiyatullin, D., Corum, C., Park, J. Y., and Garwood, M. (2006). Fast and quiet MRI using a swept radiofrequency. *J. Magn. Reson.* 181, 342–349. doi: 10.1016/j.jmr.2006.05.014
- Iida, H., Ohata, H., Iida, M., Watanabe, Y., and Dohi, S. (1998). Isoflurane and sevoflurane induce vasodilation of cerebral vessels via ATP-sensitive K<sup>+</sup> channel activation. *Anesthesiology* 89, 954–960. doi: 10.1097/0000542-199810000-00020
- Imas, O. A., Ropella, K. M., Ward, B. D., Wood, J. D., and Hudetz, A. G. (2005a). Volatile anesthetics enhance flash-induced  $\gamma$  oscillations in rat visual cortex. *Anesthesiology* 102, 937–947. doi: 10.1097/0000542-200505000-00012
- Imas, O. A., Ropella, K. M., Ward, B. D., Wood, J. D., and Hudetz, A. G. (2005b). Volatile anesthetics disrupt frontal-posterior recurrent information transfer at gamma frequencies in rat. *Neurosci. Lett.* 387, 145–150. doi: 10.1016/j.neulet.2005.06.018
- Isaacson, J. S., and Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron* 72, 231–243. doi: 10.1016/j.neuron.2011.09.027
- Jensen, O., Spaak, E., and Zumer, J. M. (2019). “Human brain oscillations: from physiological mechanisms to analysis and cognition,” in *Magnetoencephalography*, eds S. Supek and C. Aine (Cham: Springer), 1–46.
- Ji, D., and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107. doi: 10.1038/nn1825



- Jonckers, E., Delgado Y Palacios, R., Shah, D., Guglielmetti, C., Verhoye, M., and Van der Linden, A. (2014). Different anesthesia regimes modulate the functional connectivity outcome in mice. *Magn. Reson. Med.* 72, 1103–1112. doi: 10.1002/mrm.24990
- Jonckers, E., Shah, D., Hamaide, J., Verhoye, M., and Van der Linden, A. (2015). The power of using functional fMRI on small rodents to study brain pharmacology and disease. *Front. Pharmacol.* 6:231. doi: 10.3389/fphar.2015.00231
- Jonckers, E., Van Audekerke, J., De Visscher, G., Van der Linden, A., and Verhoye, M. (2011). Functional connectivity fMRI of the rodent brain: comparison of functional connectivity networks in rat and mouse. *PLoS One* 6:e18876. doi: 10.1371/journal.pone.0018876
- Jones, M., Berwick, J., Hewson-Stoate, N., Gias, C., and Mayhew, J. (2005). The effect of hypercapnia on the neural and hemodynamic responses to somatosensory stimulation. *NeuroImage* 27, 609–623. doi: 10.1016/j.neuroimage.2005.04.036
- Jones, B. E., and Yang, T. Z. (1985). The efferent projections from the reticular formation and the locus coeruleus studied by anterograde and retrograde axonal transport in the rat. *J. Comp. Neurol.* 242, 56–92. doi: 10.1002/cne.902420105
- Jordan, D., Ilg, R., Riedl, V., Schorer, A., Grimberg, S., Neufang, S., et al. (2013). Simultaneous electroencephalographic and functional magnetic resonance imaging indicate impaired cortical top-down processing in association with anesthetic-induced unconsciousness. *Anesthesiology* 119, 1031–1042. doi: 10.3410/f.718082755.793486169
- Jorm, C. M., and Stamford, J. A. (1993). Actions of the hypnotic anaesthetic, dexmedetomidine, on noradrenaline release and cell firing in rat locus coeruleus slices. *Br. J. Anaesth.* 71, 447–449. doi: 10.1093/bja/71.3.447
- Joshi, S., Li, Y., Kalwani, R. M., and Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron* 89, 221–234. doi: 10.1016/j.neuron.2015.11.028
- Jung, W. B., Shim, H. J., and Kim, S. G. (2019). Mouse BOLD fMRI at ultrahigh field detects somatosensory networks including thalamic nuclei. *NeuroImage* 195, 203–214. doi: 10.1016/j.neuroimage.2019.03.063
- Kahn, I., Desai, M., Knoblich, U., Bernstein, J., Henninger, M., Graybiel, A. M., et al. (2011). Characterization of the functional MRI response temporal linearity via optical control of neocortical pyramidal neurons. *J. Neurosci.* 31, 15086–15091. doi: 10.1523/JNEUROSCI.0007-11.2011
- Kalisch, R., Elbel, G.-K., Gössel, C., Czisch, M., and Auer, D. P. (2001). Blood pressure changes induced by arterial blood withdrawal influence bold signal in anesthetized rats at 7 tesla: implications for pharmacologic MRI. *NeuroImage* 14, 891–898. doi: 10.1006/nimg.2001.0890
- Kalshoff, D., Seehafer, J. U., Po, C., Wiedermann, D., and Hoehn, M. (2011). Functional connectivity in the rat at 11.7T: impact of physiological noise in resting state fMRI. *NeuroImage* 54, 2828–2839. doi: 10.1016/j.neuroimage.2010.10.053
- Kamibayashi, T., and Maze, M. (2000). Clinical uses of  $\alpha$ -2 adrenergic agonists. *Anesthesiology* 93, 1345–1349. doi: 10.1097/0000542-200011000-00030
- Kanai, R., Komura, Y., Shipp, S., and Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140169. doi: 10.1098/rstb.2014.0169
- Kandel, A., and Buzsáki, G. (1997). Cellular-synaptic generation of sleep spindles, spike-and-wave discharges, and evoked thalamocortical responses in the neocortex of the rat. *J. Neurosci.* 17, 6783–6797. doi: 10.1523/JNEUROSCI.17-17-06783.1997
- Kastner, S., and Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23, 315–341. doi: 10.1146/annurev.neuro.23.1.315
- Kayser, C., Montemurro, M. A., Logothetis, N. K., and Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron* 61, 597–608. doi: 10.1016/j.neuron.2009.01.008
- Keilholz, S. D. (2014). The neural basis of time-varying resting-state functional connectivity. *Brain Connect.* 4, 769–779. doi: 10.1089/brain.2014.0250
- Keller, D., Erö, C., and Markram, H. (2018). Cell densities in the mouse brain: a systematic review. *Front Neuroanat* 12:83. doi: 10.3389/fnana.2018.00083
- Kelz, M. B., Sun, Y., Chen, J., Cheng Meng, Q., Moore, J. T., Veasey, S. C., et al. (2008). An essential role for orexins in emergence from general anesthesia. *Proc. Natl. Acad. Sci. U S A* 105, 1309–1314. doi: 10.1073/pnas.0707146105
- Kenny, J. D., Westover, M. B., Ching, S., Brown, E. N., and Solt, K. (2014). Propofol and sevoflurane induce distinct burst suppression patterns in rats. *Front. Syst. Neurosci.* 8:237. doi: 10.3389/fnsys.2014.00237
- Kety, S. S., and Schmidt, C. F. (1948). The effects of altered arterial tensions of carbon dioxide and oxygen on cerebral blood flow and cerebral oxygen consumption of normal young men. *J. Clin. Invest.* 27, 484–492. doi: 10.1172/jci101995
- Kim, S.-G. (2018). Biophysics of BOLD fMRI investigated with animal models. *J. Magn. Reson.* 292, 82–89. doi: 10.1016/j.jmr.2018.04.006
- Kim, J.-H., Jung, A.-H., Jeong, D., Choi, I., Kim, K., Shin, S., et al. (2016). Selectivity of neuromodulatory projections from the basal forebrain and locus coeruleus to primary sensory cortices. *J. Neurosci.* 36, 5314–5327. doi: 10.1523/JNEUROSCI.4333-15.2016
- Kim, A., Latchoumane, C., Lee, S., Kim, G. B., Cheong, E., Augustine, G. J., et al. (2012). Optogenetically induced sleep spindle rhythms alter sleep architectures in mice. *Proc. Natl. Acad. Sci. U S A* 109, 20673–20678. doi: 10.1073/pnas.1217897109
- King, P. R., Gundlach, A. L., and Louis, W. J. (1995). Quantitative autoradiographic localization in rat brain of  $\alpha$ 2-adrenergic and non-adrenergic I-receptor binding sites labelled by [3H]rilmenidine. *Brain Res.* 675, 264–278. doi: 10.1016/0006-8993(95)00083-3
- Kiss, T., Hoffmann, W. E., Scott, L., Kawabe, T. T., Milici, A. J., Nilsen, E. A., et al. (2011). Role of thalamic projection in NMDA receptor-induced disruption of cortical slow oscillation and short-term plasticity. *Front. Psychiatry* 2:14. doi: 10.3389/fpsy.2011.00014
- Kissinger, S. T., Pak, A., Tang, Y., Masmanidis, S. C., and Chubykin, A. A. (2018). Oscillatory encoding of visual stimulus familiarity. *J. Neurosci.* 38, 6223–6240. doi: 10.1523/jneurosci.3646-17.2018
- Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., et al. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nat. Hum. Behav.* 1:0069. doi: 10.1038/s41562-017-0069
- Klee, R., Mueggler, T., Bruns, A., Kienlin, M., and Kunnecke, B. (2017a). “Is sevoflurane a viable alternative anaesthetic for functional MRI in mice?” in *Conference Paper: Annual Meeting and Exhibition of the International Society for Magnetic Resonance in Medicine at Volume: 25th* (Honolulu, HI).
- Klee, R., Mueggler, T., Bruns, A., Wyttenbach, N., Ricci, A., Gasser, R., et al. (2017b). “Etomidate anaesthesia for fmri in mice revisited: subcutaneous administration facilitates experimental procedures,” in *Conference Paper: Annual Meeting and Exhibition of the International Society for Magnetic Resonance in Medicine at Volume: 25th* (Honolulu, HI).
- Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG  $\alpha$  oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88. doi: 10.1016/j.brainresrev.2006.06.003
- Kozai, T. D., Du, Z., Gugel, Z. V., Smith, M. A., Chase, S. M., Bodily, L. M., et al. (2015). Comprehensive chronic laminar single-unit, multi-unit, and local field potential recording performance with planar single shank electrode arrays. *J. Neurosci. Methods* 242, 15–40. doi: 10.1016/j.jneumeth.2014.12.010
- Kroeger, D., and Amzica, F. (2007). Hypersensitivity of the anesthesia-induced comatose brain. *J. Neurosci.* 27, 10597–10607. doi: 10.1523/JNEUROSCI.3440-07.2007
- Ku, S. W., Lee, U., Noh, G. J., Jun, I. G., and Mashour, G. A. (2011). Preferential inhibition of frontal-to-parietal feedback connectivity is a neurophysiologic correlate of general anesthesia in surgical patients. *PLoS One* 6:e25155. doi: 10.1371/journal.pone.0025155
- Kucyi, A., and Davis, K. D. (2017). The neural code for pain: from single-cell electrophysiology to the dynamic pain connectome. *Neuroscientist* 23, 397–414. doi: 10.1177/1073858416667716
- Kundishora, A. J., Gummadavelli, A., Ma, C., Liu, M., Mccafferty, C., Schiff, N. D., et al. (2017). Restoring conscious arousal during focal limbic seizures with



- deep brain stimulation. *Cereb. Cortex* 27, 1964–1975. doi: 10.1093/cercor/bhw035
- Kuroki, S., Yoshida, T., Tsutsui, H., Iwama, M., Ando, R., Michikawa, T., et al. (2018). Excitatory neuronal hubs configure multisensory integration of slow waves in association cortex. *Cell Rep.* 22, 2873–2885. doi: 10.1016/j.celrep.2018.02.056
- Lai, H.-Y., Albaugh, D. L., Kao, Y.-C. J., Younce, J. R., and Shih, Y.-Y. I. (2015). Robust deep brain stimulation functional MRI procedures in rats and mice using an MR-compatible tungsten microwire electrode. *Magn. Reson. Med.* 73, 1246–1251. doi: 10.1002/mrm.25239
- Lamme, V. A. F. (2018). Challenges for theories of consciousness: seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373:20170344. doi: 10.1098/rstb.2017.0344
- Lamme, V. A. F., Zipser, K., and Spekreijse, H. (1998). Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proc. Natl. Acad. Sci. U S A* 95, 3263–3268. doi: 10.1073/pnas.95.6.3263
- Land, R., Engler, G., Kral, A., and Engel, A. K. (2012). Auditory evoked bursts in mouse visual cortex during isoflurane anesthesia. *PLoS One* 7:e49855. doi: 10.1371/journal.pone.0049855
- Långsjö, J. W., Alkire, M. T., Kaskinoro, K., Hayama, H., Maksimow, A., Kaisti, K. K., et al. (2012). Returning from oblivion: imaging the neural core of consciousness. *J. Neurosci.* 32, 4935–4943. doi: 10.1523/JNEUROSCI.4962-11.2012
- Larson-Prior, L. J., Zempel, J. M., Nolan, T. S., Prior, F. W., Snyder, A. Z., and Raichle, M. E. (2009). Cortical network functional connectivity in the descent to sleep. *Proc. Natl. Acad. Sci. U S A* 106, 4489–4494. doi: 10.1073/pnas.0900924106
- Le Van Quyen, M., Muller, L. E. II., Telenczuk, B., Halgren, E., Cash, S., Hatsopoulos, N. G., et al. (2016). High-frequency oscillations in human and monkey neocortex during the wake-sleep cycle. *Proc. Natl. Acad. Sci. U S A* 113, 9363–9368. doi: 10.1073/pnas.1523583113
- Lecrux, C., and Hamel, E. (2016). Neuronal networks and mediators of cortical neurovascular coupling responses in normal and altered brain states. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371:20150350. doi: 10.1098/rstb.2015.0350
- Lecrux, C., Bourourou, M., and Hamel, E. (2019). How reliable is cerebral blood flow to map changes in neuronal activity? *Auton. Neurosci.* 217, 71–79. doi: 10.1016/j.autneu.2019.01.005
- Lee, S.-H., and Dan, Y. (2012). Neuromodulation of brain states. *Neuron* 76, 209–222. doi: 10.1016/j.neuron.2012.09.012
- Lee, J. G., Hudetz, A. G., Smith, J. J., Hillard, C. J., Bosnjak, Z. J., and Kampine, J. P. (1994). The effects of halothane and isoflurane on cerebrocortical microcirculation and autoregulation as assessed by laser-Doppler flowmetry. *Anesth. Analg.* 79, 58–65. doi: 10.1213/00000539-199407000-00012
- Lee, U., Ku, S., Noh, G., Baek, S., Choi, B., and Mashour, G. A. (2013). Disruption of frontal-parietal communication by ketamine, propofol, and sevoflurane. *Anesthesiology* 118, 1264–1275. doi: 10.1097/ALN.0b013e31829103f5
- Lee, V., and Maguire, J. (2014). The impact of tonic GABA<sub>A</sub> receptor-mediated inhibition on neuronal excitability varies across brain region and cell type. *Front. Neural Circuits* 8:3. doi: 10.3389/fncir.2014.00003
- Lee, H., Mashour, G. A., Noh, G. J., Kim, S., and Lee, U. (2013). Reconfiguration of network hub structure after propofol-induced unconsciousness. *Anesthesiology* 119, 1347–1359. doi: 10.1097/aln.0b013e3182a8ec8c
- Legrain, V., Iannetti, G. D., Plaghki, L., and Mouraux, A. (2011). The pain matrix Reloaded: a salience detection system for the body. *Prog. Neurobiol.* 93, 111–124. doi: 10.1016/j.pneurobio.2010.10.005
- Lehto, L. J., Filip, P., Laakso, H., Sierra, A., Slopesma, J. P., Johnson, M. D., et al. (2018). Tuning neuromodulation effects by orientation selective deep brain stimulation in the rat medial frontal cortex. *Front. Neurosci.* 12:899. doi: 10.3389/fnins.2018.00899
- Lei, H., Grinberg, O., Nwaigwe, C. I., Hou, H. G., Williams, H., Swartz, H. M., et al. (2001). The effects of ketamine-xylazine anesthesia on cerebral blood flow and oxygenation observed using nuclear magnetic resonance perfusion imaging and electron paramagnetic resonance oximetry. *Brain Res.* 913, 174–179. doi: 10.1016/s0006-8993(01)02786-x
- Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176. doi: 10.1038/nature05453
- Leong, A. T. L., Chan, R. W., Gao, P. P., Chan, Y.-S., Tsia, K. K., Yung, W.-H., et al. (2016). Long-range projections coordinate distributed brain-wide neural activity with a specific spatiotemporal profile. *Proc. Natl. Acad. Sci. U S A* 113, E8306–E8315. doi: 10.1073/pnas.1616361113
- Leoni, R. F., Paiva, F. F., Henning, E. C., Nascimento, G. C., Tannus, A., de Araujo, D. B., et al. (2011). Magnetic resonance imaging quantification of regional cerebral blood flow and cerebrovascular reactivity to carbon dioxide in normotensive and hypertensive rats. *NeuroImage* 58, 75–81. doi: 10.1016/j.neuroimage.2011.06.030
- Leung, L. S., Luo, T., Ma, J., and Herrick, I. (2014). Brain areas that influence general anesthesia. *Prog. Neurobiol.* 122, 24–44. doi: 10.1016/j.pneurobio.2014.08.001
- Lewis, L. D., and Akeju, O. (2017). Hierarchy in disruption of large-scale networks across altered arousal states. *Br. J. Anaesth.* 119, 566–568. doi: 10.1093/bja/aex297
- Lewis, L. D., Ching, S., Weiner, V. S., Peterfreund, R. A., Eskandar, E. N., Cash, S. S., et al. (2013). Local cortical dynamics of burst suppression in the anesthetized brain. *Brain* 136, 2727–2737. doi: 10.1093/brain/awt174
- Lewis, L. D., Voigts, J., Flores, F. J., Schmitt, L. I., Wilson, M. A., Halassa, M. M., et al. (2015). Thalamic reticular nucleus induces fast and local modulation of arousal state. *Elife* 4:e08760. doi: 10.7554/eLife.08760
- Lewis, L. D., Weiner, V. S., Mukamel, E. A., Donoghue, J. A., Eskandar, E. N., Madsen, J. R., et al. (2012). Rapid fragmentation of neuronal networks at the onset of propofol-induced unconsciousness. *Proc. Natl. Acad. Sci. U S A* 109, E3377–E3386. doi: 10.1073/pnas.1210907109
- Li, Q., Li, G., Wu, D., Lu, H., Hou, Z., Ross, C. A., et al. (2017). Resting-state functional MRI reveals altered brain connectivity and its correlation with motor dysfunction in a mouse model of Huntington's disease. *Sci. Rep.* 7:16742. doi: 10.1038/s41598-017-17026-5
- Li, B. H., Lohmann, J. S., Schuler, H. G., and Cronin, A. J. (2003). Preservation of the cortical somatosensory-evoked potential during dexmedetomidine infusion in rats. *Anesth. Analg.* 96, 1155–1160. doi: 10.1213/01.ane.0000053239.62623.32
- Lichtner, G., Aukstulewicz, R., Kirilina, E., Velten, H., Mavrodiss, D., Scheel, M., et al. (2018). Effects of propofol anesthesia on the processing of noxious stimuli in the spinal cord and the brain. *NeuroImage* 172, 642–653. doi: 10.1016/j.neuroimage.2018.02.003
- Liska, A., Galbusera, A., Schwarz, A. J., and Gozzi, A. (2015). Functional connectivity hubs of the mouse brain. *NeuroImage* 115, 281–291. doi: 10.1016/j.neuroimage.2015.04.033
- Liu, X., Lauer, K. K., Ward, B. D., Li, S. J., and Hudetz, A. G. (2013a). Differential effects of deep sedation with propofol on the specific and nonspecific thalamocortical systems: a functional magnetic resonance imaging study. *Anesthesiology* 118, 59–69. doi: 10.1097/ALN.0b013e318277a801
- Liu, X., Pillay, S., Li, R., Vizuete, J. A., Pechman, K. R., Schmainda, K. M., et al. (2013b). Multiphasic modification of intrinsic functional connectivity of the rat brain during increasing levels of propofol. *NeuroImage* 83, 581–592. doi: 10.1016/j.neuroimage.2013.07.003
- Liu, X., Zhu, X. H., Zhang, Y., and Chen, W. (2013c). The change of functional connectivity specificity in rats under various anesthesia levels and its neural origin. *Brain Topogr.* 26, 363–377. doi: 10.1007/s10548-012-0267-5
- Liu, X., Zhu, X. H., Zhang, Y., and Chen, W. (2011). Neural origin of spontaneous hemodynamic fluctuations in rats under burst-suppression anesthesia condition. *Cereb. Cortex* 21, 374–384. doi: 10.1093/cercor/bhq105
- Llinás, R., and Ribary, U. (1993). Coherent 40-Hz oscillation characterizes dream state in humans. *Proc. Natl. Acad. Sci. U S A* 90, 2078–2081. doi: 10.1073/pnas.90.5.2078
- Longmuir, I. S., and Pashko, L. (1976). The induction of cytochrome P-450 by hypoxia. *Adv. Exp. Med. Biol.* 75, 171–175. doi: 10.1007/978-1-4684-3273-2\_22
- Low, L. A., Bauer, L. C., and Klaunberg, B. A. (2016a). Comparing the effects of isoflurane and  $\alpha$  chloralose upon mouse physiology. *PLoS One* 11:e0154936. doi: 10.1371/journal.pone.0154936

- Low, L. A., Bauer, L. C., Pitcher, M. H., and Bushnell, M. C. (2016b). Restraint training for awake functional brain scanning of rodents can cause long-lasting changes in pain and stress responses. *Pain* 157, 1761–1772. doi: 10.1097/j.pain.0000000000000579
- Lu, H., Zou, Q., Gu, H., Raichle, M. E., Stein, E. A., and Yang, Y. (2012). Rat brains also have a default mode network. *Proc. Natl. Acad. Sci. U S A* 109, 3979–3984. doi: 10.1073/pnas.1200506109
- Luczak, A., Barthó, P., Marguet, S. L., Buzsáki, G., and Harris, K. D. (2007). Sequential structure of neocortical spontaneous activity in vivo. *Proc. Natl. Acad. Sci. U S A* 104, 347–352. doi: 10.1073/pnas.0605643104
- Luczak, A., McNaughton, B. L., and Harris, K. D. (2015). Packet-based communication in the cortex. *Nat. Rev. Neurosci.* 16, 745–755. doi: 10.1038/nrn4026
- Lurie, J. D., Kessler, D., Bassett, D., Betzel, R., Breakspear, M., Keilholz, S., et al. (2018). On the nature of resting fMRI and time-varying functional connectivity. *PsyArXiv* [Preprint].
- Lüthi, A. (2014). Sleep spindles: where they come from, what they do. *Neuroscientist* 20, 243–256. doi: 10.1177/1073858413500854
- Ma, S., Allocca, G., Ong-Palsson, E. K., Singleton, C. E., Hawkes, D., McDougall, S. J., et al. (2017). Nucleus incertus promotes cortical desynchronization and behavioral arousal. *Brain Struct. Funct.* 222, 515–537. doi: 10.1007/s00429-016-1230-0
- Ma, S., Hangya, B., Leonard, C. S., Wisden, W., and Gundlach, A. L. (2018). Dual-transmitter systems regulating arousal, attention, learning and memory. *Neurosci. Biobehav. Rev.* 85, 21–33. doi: 10.1016/j.neubiorev.2017.07.009
- Madularu, D., Mathieu, A. P., Kumaragamage, C., Reynolds, L. M., Near, J., Flores, C., et al. (2017). A non-invasive restraining system for awake mouse imaging. *J. Neurosci. Methods* 287, 53–57. doi: 10.1016/j.jneumeth.2017.06.008
- Mahdavi, A., Qin, Y., Aubry, A.-S., Cornec, D., Kulikova, S., and Pinault, D. (2019). Sleep-related thalamocortical spindles and delta oscillations are reduced during a ketamine-induced psychosis-relevant transition state. *bioRxiv* [Preprint]. doi: 10.1101/833459
- Majeed, W., Magnuson, M., Hasenkamp, W., Schwarb, H., Schumacher, E. H., Barsalou, L., et al. (2011). Spatiotemporal dynamics of low frequency BOLD fluctuations in rats and humans. *NeuroImage* 54, 1140–1150. doi: 10.1016/j.neuroimage.2010.08.030
- Mandino, F., Cerri, D. H., Garin, C. M., Straathof, M., van Tilborg, G. A. F., Chakravarty, M. M., et al. (2020). Animal functional magnetic resonance imaging: trends and path toward standardization. *Front. Neuroinform.* 13:78. doi: 10.3389/fninf.2019.00078
- Marlinski, V., Sirota, M. G., and Beloozerova, I. N. (2012). Differential gating of thalamocortical signals by reticular nucleus of thalamus during locomotion. *J. Neurosci.* 32, 15823–15836. doi: 10.1523/JNEUROSCI.0782-12.2012
- Martins, I., and Tavares, I. (2017). Reticular formation and pain: the past and the future. *Front. Neuroanat.* 11:51. doi: 10.3389/fnana.2017.00051
- Masamoto, K., and Kanno, I. (2012). Anesthesia and the quantitative evaluation of neurovascular coupling. *J. Cereb. Blood Flow Metab.* 32, 1233–1247. doi: 10.1038/jcbfm.2012.50
- Mashour, G. A. (2014). Top-down mechanisms of anesthetic-induced unconsciousness. *Front. Syst. Neurosci.* 8:115. doi: 10.3389/fnsys.2014.00115
- Mashour, G. A., and Alkire, M. T. (2013). Evolution of consciousness: phylogeny, ontogeny, and emergence from general anesthesia. *Proc. Natl. Acad. Sci. U S A* 110, 10357–10364. doi: 10.1073/pnas.1301188110
- Mashour, G. A., and Hudetz, A. G. (2017). Bottom-up and top-down mechanisms of general anesthetics modulate different dimensions of consciousness. *Front. Neural Circuits* 11:44. doi: 10.3389/fncir.2017.00044
- Mashour, G. A., and Hudetz, A. G. (2018). Neural correlates of unconsciousness in large-scale brain networks. *Trends Neurosci.* 41, 150–160. doi: 10.1016/j.tins.2018.01.003
- Masri, R., Trageser, J. C., Bezdudnaya, T., Li, Y., and Keller, A. (2006). Cholinergic regulation of the posterior medial thalamic nucleus. *J. Neurophysiol.* 96, 2265–2273. doi: 10.1152/jn.00476.2006
- Masri, R., Bezdudnaya, T., Trageser, J. C., and Keller, A. (2008). Encoding of stimulus frequency and sensor motion in the posterior medial thalamic nucleus. *J. Neurophysiol.* 100, 681–689. doi: 10.1152/jn.01322.2007
- Massimini, M., Ferrarelli, F., Esser, S. K., Riedner, B. A., Huber, R., Murphy, M., et al. (2007). Triggering sleep slow waves by transcranial magnetic stimulation. *Proc. Natl. Acad. Sci. U S A* 104, 8496–8501. doi: 10.1073/pnas.0702495104
- Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. doi: 10.1126/science.1117256
- Massimini, M., Huber, R., Ferrarelli, F., Hill, S., and Tononi, G. (2004). The sleep slow oscillation as a traveling wave. *J. Neurosci.* 24, 6862–6870. doi: 10.1523/JNEUROSCI.1318-04.2004
- Mateo, C., Knutsen, P. M., Tsai, P. S., Shih, A. Y., and Kleinfeld, D. (2017). Entrainment of arteriole vasomotor fluctuations by neural activity is a basis of blood-oxygenation-level-dependent “resting-state” connectivity. *Neuron* 96, 936.e3–948.e3. doi: 10.1016/j.neuron.2017.10.012
- Matsui, T., Murakami, T., and Ohki, K. (2018). Neuronal origin of the temporal dynamics of spontaneous BOLD activity correlation. *Cereb. Cortex* 29, 1496–1508. doi: 10.1093/cercor/bhy045
- Matsuura, T., and Kanno, I. (2002). Effect of nitric oxide synthase inhibitor on the local cerebral blood flow evoked by rat somatosensory stimulation under hyperoxia. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 131, 267–274. doi: 10.1016/s1095-6433(01)00450-0
- Matsuura, T., Kashikura, K., and Kanno, I. (2001). Hemodynamics of local cerebral blood flow induced by somatosensory stimulation under normoxia and hyperoxia in rats. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 129, 363–372. doi: 10.1016/s1095-6433(00)00354-8
- Mattia, M., and Sanchez-Vives, M. V. (2012). Exploring the spectrum of dynamical regimes and timescales in spontaneous cortical activity. *Cogn. Neurodyn.* 6, 239–250. doi: 10.1007/s11571-011-9179-4
- Maxwell, C. R., Ehrlichman, R. S., Liang, Y., Trief, D., Kanes, S. J., Karp, J., et al. (2006). Ketamine produces lasting disruptions in encoding of sensory stimuli. *J. Pharmacol. Exp. Ther.* 316, 315–324. doi: 10.1124/jpet.105.091199
- McCallum, J. B., Boban, N., Hogan, Q., Schmeling, W. T., Kampine, J. P., and Bosnjak, Z. J. (1998). The mechanism of  $\alpha$ 2-adrenergic inhibition of sympathetic ganglionic transmission. *Anesth. Analg.* 87, 503–510. doi: 10.1097/0000539-199809000-00001
- McCarthy, M. M., Brown, E. N., and Kopell, N. (2008). Potential network mechanisms mediating electroencephalographic  $\beta$  rhythm changes during propofol-induced paradoxical excitation. *J. Neurosci.* 28, 13488–13504. doi: 10.1523/JNEUROSCI.3536-08.2008
- McCormick, D. A. (1992). Neurotransmitter actions in the thalamus and cerebral cortex and their role in neuromodulation of thalamocortical activity. *Prog. Neurobiol.* 39, 337–388. doi: 10.1016/0301-0082(92)90012-4
- McCormick, D. A., and Bal, T. (1997). Sleep and arousal: thalamocortical mechanisms. *Ann. Rev. Neurosci.* 20, 185–215. doi: 10.1146/annurev.neuro.20.1.185
- McCormick, D. A., and Pape, H. C. (1990). Properties of a hyperpolarization-activated cation current and its role in rhythmic oscillation in thalamocortical neurones. *J. Physiol.* 431, 291–318. doi: 10.1113/jphysiol.1990.sp018331
- McGinley, M. J., David, S. V., and McCormick, D. A. (2015a). Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron* 87, 179–192. doi: 10.1016/j.neuron.2015.05.038
- McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zgha, E., Cadwell, C. R., et al. (2015b). Waking state: rapid variations modulate neural and behavioral responses. *Neuron* 87, 1143–1161. doi: 10.1016/j.neuron.2015.09.012
- McIlhonne, A. E., Beausoleil, N. J., Kells, N. J., Johnson, C. B., and Mellor, D. J. (2018). Effects of halothane on the electroencephalogram of the chicken. *Vet. Med.* 113, 98–105. doi: 10.1002/vms3.91
- McIntyre, C. C., and Anderson, R. W. (2016). Deep brain stimulation mechanisms: the control of network activity via neurochemistry modulation. *J. Neurochem.* 139, 338–345. doi: 10.1111/jnc.13649
- McKinstry-Wu, A., and Kelz, M. B. (2019). *Neural Mechanisms of Anaesthetics. Oxford Textbook of Neuroscience and Anaesthesiology*. New York, NY: Oxford University Press.
- McVea, D. A., Murphy, T. H., and Mohajerani, M. H. (2016). Large scale cortical functional networks associated with slow-wave and spindle-burst-related spontaneous activity. *Front. Neural Circuits* 10:103. doi: 10.3389/fncir.2016.00103

- Mechling, A. E., Hubner, N. S., Lee, H. L., Hennig, J., von Elverfeldt, D., and Harsan, L. A. (2014). Fine-grained mapping of mouse brain functional connectivity with resting-state fMRI. *NeuroImage* 96, 203–215. doi: 10.1016/j.neuroimage.2014.03.078
- Melzack, R. (1989). Labat lecture. Phantom limbs. *Reg. Anesth.* 14, 208–211.
- Melzack, R. (2005). Evolution of the neuromatrix theory of pain. The Prithvi Raj Lecture: presented at the third World Congress of World Institute of Pain, Barcelona 2004. *Pain Pract.* 5, 85–94. doi: 10.1111/j.1533-2500.2005.05203.x
- Meuth, S. G., Budde, T., Kanyshkova, T., Broicher, T., Munsch, T., and Pape, H. C. (2003). Contribution of TWIK-related acid-sensitive K<sup>+</sup> channel 1 (TASK1) and TASK3 channels to the control of activity modes in thalamocortical neurons. *J. Neurosci.* 23, 6460–6469. doi: 10.1523/JNEUROSCI.23-16-06.460.2003
- Meuth, S. G., Kanyshkova, T., Meuth, P., Landgraf, P., Munsch, T., Ludwig, A., et al. (2006). Membrane resting potential of thalamocortical relay neurons is shaped by the interaction among TASK3 and HCN2 channels. *J. Neurophysiol.* 96, 1517–1529. doi: 10.1152/jn.01212.2005
- Meyer, H. S., Egger, R., Guest, J. M., Foerster, R., Reissl, S., and Oberlaender, M. (2013). Cellular organization of cortical barrel columns is whisker-specific. *Proc. Natl. Acad. Sci. U S A* 110, 19113–19118. doi: 10.1073/pnas.1312691110
- Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J. M., Kennedy, H., and Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* 89, 384–397. doi: 10.1016/j.neuron.2015.12.018
- Michelson, N. J., and Kozai, T. D. Y. (2018). Isoflurane and ketamine differentially influence spontaneous and evoked laminar electrophysiology in mouse V1. *J. Neurophysiol.* 120, 2232–2245. doi: 10.1152/jn.00299.2018
- Milenkovic, N., Zhao, W.-J., Walcher, J., Albert, T., Siemens, J., Lewin, G. R., et al. (2014). A somatosensory circuit for cooling perception in mice. *Nat. Neurosci.* 17, 1560–1566. doi: 10.1038/nn.3828
- Mizobe, T. (2019). The halothane hepatitis that was not. *Br. J. Anaesth.* 124, e2–e3. doi: 10.1016/j.bja.2019.09.018
- Mo, C., and Sherman, S. M. (2019). A sensorimotor pathway via higher-order thalamus. *J. Neurosci.* 39, 692–704. doi: 10.1523/JNEUROSCI.1467-18.2018
- Moghaddam, B., Adams, B., Verma, A., and Daly, D. (1997). Activation of glutamatergic neurotransmission by ketamine: a novel step in the pathway from NMDA receptor blockade to dopaminergic and cognitive disruptions associated with the prefrontal cortex. *J. Neurosci.* 17, 2921–2927. doi: 10.1523/JNEUROSCI.17-08-02921.1997
- Mohajerani, M. H., Chan, A. W., Mohsenvand, M., LeDue, J., Liu, R., McVea, D. A., et al. (2013). Spontaneous cortical activity alternates between motifs defined by regional axonal projections. *Nat. Neurosci.* 16, 1426–1435. doi: 10.1038/nn.3499
- Mölle, M., Marshall, L., Gais, S., and Born, J. (2002). Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. *J. Neurosci.* 22, 10941–10947. doi: 10.1523/JNEUROSCI.22-24-10941.2002
- Montemurro, M. A., Rasch, M. J., Murayama, Y., Logothetis, N. K., and Panzeri, S. (2008). Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Curr. Biol.* 18, 375–380. doi: 10.1016/j.cub.2008.02.023
- Monti, M. M., Lutkenhoff, E. S., Rubinov, M., Boveroux, P., Vanhaudenhuyse, A., Gosseries, O., et al. (2013). Dynamic change of global and local information processing in propofol-induced loss and recovery of consciousness. *PLoS Comput. Biol.* 9:e1003271. doi: 10.1371/journal.pcbi.1003271
- Montijn, J. S., Meijer, G. T., Lansink, C. S., and Pennartz, C. M. A. (2016). Population-level neural codes are robust to single-neuron variability from a multidimensional coding perspective. *Cell Rep.* 16, 2486–2498. doi: 10.1016/j.celrep.2016.07.065
- Moon, J. Y., Lee, U., Blain-Moraes, S., and Mashour, G. A. (2015). General relationship of global topology, local dynamics, and directionality in large-scale brain networks. *PLoS Comput. Biol.* 11:e1004225. doi: 10.1371/journal.pcbi.1004225
- Mouraux, A., Diukova, A., Lee, M. C., Wise, R. G., and Iannetti, G. D. (2011). A multisensory investigation of the functional significance of the “pain matrix”. *NeuroImage* 54, 2237–2249. doi: 10.1016/j.neuroimage.2010.09.084
- Mouraux, A., and Iannetti, G. D. (2018). The search for pain biomarkers in the human brain. *Brain* 141, 3290–3307. doi: 10.1093/brain/awy281
- Mueggler, T., Baumann, D., Rausch, M., Staufenbiel, M., and Rudin, M. (2003). Age-dependent impairment of somatosensory response in the amyloid precursor protein 23 transgenic mouse model of Alzheimer's disease. *J. Neurosci.* 23, 8231–8236. doi: 10.1523/JNEUROSCI.23-23-08231.2003
- Mukamel, E. A., Pirondini, E., Babadi, B., Wong, K. F., Pierce, E. T., Harrell, P. G., et al. (2014). A transition in brain state during propofol-induced unconsciousness. *J. Neurosci.* 34, 839–845. doi: 10.1523/JNEUROSCI.5813-12.2014
- Muller, L., Chavane, F., Reynolds, J., and Sejnowski, T. J. (2018). Cortical travelling waves: mechanisms and computational principles. *Nat. Rev. Neurosci.* 19, 255–268. doi: 10.1038/nrn.2018.20
- Munting, L. P., Derieppe, M. P. P., Suidgeest, E., Denis de Senneville, B., Wells, J., and van der Weerd, L. (2019). Influence of different isoflurane anesthesia protocols on murine cerebral hemodynamics measured with pseudo-continuous arterial spin labeling. *NMR Biomed.* 32:e4105. doi: 10.1002/nbm.4105
- Muthukumaraswamy, S. D., Shaw, A. D., Jackson, L. E., Hall, J., Moran, R., and Saxena, N. (2015). Evidence that subanesthetic doses of ketamine cause sustained disruptions of NMDA and AMPA-mediated frontoparietal connectivity in humans. *J. Neurosci.* 35, 11694–11706. doi: 10.1523/JNEUROSCI.0903-15.2015
- Murphy, M. C., Chan, K. C., Kim, S.-G., and Vazquez, A. L. (2017). Macroscale variation in resting-state neuronal activity and connectivity assessed by simultaneous calcium imaging, hemodynamic imaging and electrophysiology. *NeuroImage* 169, 352–362. doi: 10.1016/j.neuroimage.2017.12.070
- Murphy, C., Krause, B., and Banks, M. (2019). Selective effects of isoflurane on cortico-cortical feedback afferent responses in murine non-primary neocortex. *Br. J. Anaesth.* 123, 488–496. doi: 10.1016/j.bja.2019.06.018
- Murrell, J. C., Waters, D., and Johnson, C. B. (2008). Comparative effects of halothane, isoflurane, sevoflurane and desflurane on the electroencephalogram of the rat. *Lab. Anim.* 42, 161–170. doi: 10.1258/la.2007.06019e
- Nasrallah, F. A., Lew, S. K., Low, A. S., and Chuang, K. H. (2014a). Neural correlate of resting-state functional connectivity under  $\alpha 2$  adrenergic receptor agonist, medetomidine. *NeuroImage* 84, 27–34. doi: 10.1016/j.neuroimage.2013.08.004
- Nasrallah, F. A., Tay, H. C., and Chuang, K. H. (2014b). Detection of functional connectivity in the resting mouse brain. *NeuroImage* 86, 417–424. doi: 10.1016/j.neuroimage.2013.10.025
- Nasrallah, F. A., Tan, J., and Chuang, K. H. (2012). Pharmacological modulation of functional connectivity:  $\alpha 2$ -adrenergic receptor agonist alters synchrony but not neural activation. *NeuroImage* 60, 436–446. doi: 10.1016/j.neuroimage.2011.12.026
- Nelson, L. E., Lu, J., Guo, T., Saper, C. B., Franks, N. P., and Maze, M. (2003). The  $\alpha 2$ -adrenoceptor agonist dexmedetomidine converges on an endogenous sleep-promoting pathway to exert its sedative effects. *Anesthesiology* 98, 428–436. doi: 10.1097/0000542-200302000-00024
- Nelson, C. L., Sarter, M., and Bruno, J. P. (2005). Prefrontal cortical modulation of acetylcholine release in posterior parietal cortex. *Neuroscience* 132, 347–359. doi: 10.1016/j.neuroscience.2004.12.007
- Neske, G. T. (2015). The slow oscillation in cortical and thalamic networks: mechanisms and functions. *Front. Neural Circuits* 9:88. doi: 10.3389/fncir.2015.00088
- Ní Huircheartaigh, R., Warnaby, C., Rogers, R., Jbabdi, S., and Tracey, I. (2013). Slow-wave activity saturation and thalamocortical isolation during propofol anesthesia in humans. *Sci. Transl. Med.* 5:208ra148. doi: 10.1126/scitranslmed.3006007
- Niedermeyer, E., Sherman, D. L., Geocadin, R. J., Hansen, H. C., and Hanley, D. F. (1999). The burst-suppression electroencephalogram. *Clin. Electroencephalogr.* 30, 99–105. doi: 10.1177/155005949903000305
- Niendorf, T., Pohlmann, A., Reimann, H. M., Waiczies, H., Peper, E., Huelnhagen, T., et al. (2015). Advancing cardiovascular, neurovascular and renal magnetic resonance imaging in small rodents using cryogenic radiofrequency coil technology. *Front. Pharmacol.* 6:255. doi: 10.3389/fphar.2015.00255
- Nir, Y., Staba, R. J., Andrillon, T., Vyazovskiy, V. V., Cirelli, C., Fried, I., et al. (2011). Regional slow waves and spindles in human sleep. *Neuron* 70, 153–169. doi: 10.1016/j.neuron.2011.02.043



- Niranjan, A., Christie, I. N., Solomon, S. G., Wells, J. A., and Lythgoe, M. F. (2016). fMRI mapping of the visual system in the mouse brain with interleaved snapshot GE-EPI. *NeuroImage* 139, 337–345. doi: 10.1016/j.neuroimage.2016.06.015
- Noreika, V., Jylhänkangas, L., Móró, L., Valli, K., Kaskinoro, K., Aantaa, R., et al. (2011). Consciousness lost and found: subjective experiences in an unresponsive state. *Brain Cogn.* 77, 327–334. doi: 10.1016/j.bandc.2011.09.002
- Nuñez, A., Amzica, F., and Steriade, M. (1992). Voltage-dependent fast (20–40 Hz) oscillations in long-axonated neocortical neurons. *Neuroscience* 51, 7–10. doi: 10.1016/0306-4522(92)90464-d
- Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. U S A* 87, 9868–9872. doi: 10.1073/pnas.87.24.9868
- Ojemann, G. A., Ramsey, N. F., and Ojemann, J. (2013). Relation between functional magnetic resonance imaging (fMRI) and single neuron, local field potential (LFP) and electrocorticography (ECoG) activity in human cortex. *Front. Hum. Neurosci.* 7:34. doi: 10.3389/fnhum.2013.00034
- Ojemann, G. A., Schoenfield-McNeill, J., and Corina, D. (2002). Anatomic subdivisions in human temporal cortical neuronal activity related to recent verbal memory. *Nat. Neurosci.* 5, 64–71. doi: 10.1038/nn785
- Olcese, U., Oude Lohuis, M. N., and Pennartz, C. M. A. (2018). Sensory processing across conscious and nonconscious brain states: from single neurons to distributed networks for inferential representation. *Front. Syst. Neurosci.* 12:49. doi: 10.3389/fnsys.2018.00049
- Olufsen, M., Whittington, M. A., Camperi, M., and Kopell, N. (2003). New functions for the gamma rhythm: population tuning and preprocessing for the beta rhythm. *J. Comput. Neurosci.* 14, 33–54. doi: 10.1023/a:1021124317706
- Oye, I., Paulsen, O., and Maurset, A. (1992). Effects of ketamine on sensory perception: evidence for a role of N-methyl-D-aspartate receptors. *J. Pharmacol. Exp. Ther.* 260, 1209–1213.
- Paasonen, J., Laakso, H., Pirttimäki, T., Stenroos, P., Salo, R. A., Zhurakovskaya, E., et al. (2020). Multi-band SWIFT enables quiet and artefact-free EEG-fMRI and awake fMRI studies in rat. *NeuroImage* 206:116338. doi: 10.1016/j.neuroimage.2019.116338
- Paasonen, J., Stenroos, P., Salo, R. A., Kiviniemi, V., and Grohn, O. (2018). Functional connectivity under six anesthesia protocols and the awake condition in rat brain. *NeuroImage* 172, 9–20. doi: 10.1016/j.neuroimage.2018.01.014
- Pachitariu, M., Lyamzin, D. R., Sahani, M., and Lesica, N. A. (2015). State-dependent population coding in primary auditory cortex. *J. Neurosci.* 35, 2058–2073. doi: 10.1523/JNEUROSCI.3318-14.2015
- Pack, C. C., Berezovskii, V. K., and Born, R. T. (2001). Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys. *Nature* 414, 905–908. doi: 10.1038/414905a
- Pais-Roldán, P., Biswal, B., Scheffler, K., and Yu, X. (2018). Identifying respiration-related aliasing artifacts in the rodent resting-state fMRI. *Front. Neurosci.* 12:788. doi: 10.3389/fnins.2018.00788
- Pal, D., Dean, J. G., Liu, T., Li, D., Watson, C. J., Hudetz, A. G., et al. (2018). Differential role of prefrontal and parietal cortices in controlling level of consciousness. *Curr. Biol.* 28, 2145.e5–2152.e5. doi: 10.1016/j.cub.2018.05.025
- Pal, D., Hambrecht-Wiedbusch, V. S., Silverstein, B. H., and Mashour, G. A. (2015). Electroencephalographic coherence and cortical acetylcholine during ketamine-induced unconsciousness. *Br. J. Anaesth.* 114, 979–989. doi: 10.1093/bja/aev095
- Palanca, B. J., Mitra, A., Larson-Prior, L., Snyder, A. Z., Avidan, M. S., and Raichle, M. E. (2015). Resting-state functional magnetic resonance imaging correlates of sevoflurane-induced unconsciousness. *Anesthesiology* 123, 346–356. doi: 10.1097/aln.0000000000000731
- Palva, S., Linkenkaer-Hansen, K., Näätänen, R., and Palva, J. M. (2005). Early neural correlates of conscious somatosensory perception. *J. Neurosci.* 25, 5248–5258. doi: 10.1523/jneurosci.0141-05.2005
- Palva, S., and Palva, J. M. (2007). New vistas for  $\alpha$ -frequency band oscillations. *Trends Neurosci.* 30, 150–158. doi: 10.1016/j.tins.2007.02.001
- Pan, W. J., Billings, J. C., Grooms, J. K., Shakil, S., and Keilholz, S. D. (2015). Considerations for resting state functional MRI and functional connectivity studies in rodents. *Front. Neurosci.* 9:269. doi: 10.3389/fnins.2015.00269
- Pan, W. J., Billings, J., Nezafati, M., Abbas, A., and Keilholz, S. (2018). Resting state fMRI in rodents. *Curr. Protoc. Neurosci.* 83:e45. doi: 10.1002/cpns.45
- Paquette, T., Jeffrey-Gauthier, R., Leblond, H., and Pich, E. M. (2018). Functional neuroimaging of nociceptive and pain-related activity in the spinal cord and brain: insights from neurovascular coupling studies. *Anat. Rec.* 301, 1585–1595. doi: 10.1002/ar.23854
- Paquette, T., Tokunaga, R., Touj, S., Leblond, H., and Piché, M. (2019). Regulation of cortical blood flow responses by the nucleus basalis of Meynert during nociceptive processing. *Neurosci. Res.* 149, 22–28. doi: 10.1016/j.neures.2019.01.008
- Pavone, K. J., Su, L., Gao, L., Eromo, E., Vazquez, R., Rhee, J., et al. (2017). Lack of responsiveness during the onset and offset of sevoflurane anesthesia is associated with decreased awake- $\alpha$  oscillation power. *Front. Syst. Neurosci.* 11:38. doi: 10.3389/fnsys.2017.00038
- Pawela, C. P., Biswal, B. B., Cho, Y. R., Kao, D. S., Li, R., Jones, S. R., et al. (2008). Resting-state functional connectivity of the rat brain. *Magn. Reson. Med.* 59, 1021–1029. doi: 10.1002/mrm.21524
- Pennartz, C. M. A., Dora, S., Muckli, L., and Lorteije, J. A. M. (2019). Towards a Unified View on Pathways and Functions of Neural Recurrent Processing. *Trends Neurosci.* 42. doi: 10.1016/j.tins.2019.07.005
- Petersen, C. C. H., Hahn, T. T. G., Mehta, M., Grinvald, A., and Sakmann, B. (2003). Interaction of sensory responses with spontaneous depolarization in layer 2/3 barrel cortex. *Proc. Natl. Acad. Sci. U S A* 100, 13638–13643. doi: 10.1073/pnas.2235811100
- Petersen, S. E., and Sporns, O. (2015). Brain networks and cognitive architectures. *Neuron* 88, 207–219. doi: 10.1016/j.neuron.2015.09.027
- Petrinovic, M. M., Hankov, G., Schroeter, A., Bruns, A., Rudin, M., Von Kienlin, M., et al. (2016). A novel anesthesia regime enables neurofunctional studies and imaging genetics across mouse strains. *Sci. Rep.* 6:24523. doi: 10.1038/srep24523
- Peyrache, A., Battaglia, F. P., and Destexhe, A. (2011). Inhibition recruitment in prefrontal cortex during sleep spindles and gating of hippocampal inputs. *Proc. Natl. Acad. Sci. U S A* 108, 17207–17212. doi: 10.1073/pnas.1103612108
- Pfaff, D. (2005). *Brain Arousal and Information Theory: Neural and Genetic Mechanisms*. Cambridge, MA: Harvard University Press. page 71–72, 123.
- Phillips, W. A., Bachmann, T., and Storm, J. F. (2018). Apical function in neocortical pyramidal cells: a common pathway by which general anesthetics can affect mental state. *Front. Neural Circuits* 12:50. doi: 10.3389/fncir.2018.00050
- Picard, N., Takesian, A. E., Fagiolini, M., and Hensch, T. K. (2019). NMDA 2A receptors in parvalbumin cells mediate sex-specific rapid ketamine response on cortical activity. *Mol. Psychiatry* 24, 828–838. doi: 10.1038/s41380-018-0341-9
- Picchioni, D., Reith, R. M., Nadel, J. L., and Smith, C. B. (2014). Sleep, plasticity and the pathophysiology of neurodevelopmental disorders: the potential roles of protein synthesis and other cellular processes. *Brain Sci.* 4, 150–201. doi: 10.3390/brainsci4010150
- Polack, P.-O., Friedman, J., and Golshani, P. (2013). Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *Nat. Neurosci.* 16, 1331–1339. doi: 10.1038/nn.3464
- Porro, C. A., Cavazzuti, M., Giuliani, D., Vellani, V., Lui, F., and Baraldi, P. (2004). Effects of ketamine anesthesia on central nociceptive processing in the rat: a 2-deoxyglucose study. *Neuroscience* 125, 485–494. doi: 10.1016/j.neuroscience.2004.01.039
- Poulet, J. F. A., and Crochet, S. (2019). The cortical states of wakefulness. *Front. Syst. Neurosci.* 12:64. doi: 10.3389/fnsys.2018.00064
- Prerau, M. J., Brown, R. E., Bianchi, M. T., Ellenbogen, J. M., and Purdon, P. L. (2017). Sleep neurophysiological dynamics through the lens of multitaper spectral analysis. *Physiology* 32, 60–92. doi: 10.1152/physiol.00062.2015
- Prerau, M. J., Hartnack, K. E., Obregon-Henao, G., Sampson, A., Merlino, M., Gannon, K., et al. (2014). Tracking the sleep onset process: an empirical model of behavioral and physiological dynamics. *PLoS Comput. Biol.* 10:e1003866. doi: 10.1371/journal.pcbi.1003866



- Preti, M. G., Bolton, T. A., and Van De Ville, D. (2017). The dynamic functional connectome: state-of-the-art and perspectives. *NeuroImage* 160, 41–54. doi: 10.1016/j.neuroimage.2016.12.061
- Price, D. D. (2000). Psychological and neural mechanisms of the affective dimension of pain. *Science* 288, 1769–1772. doi: 10.1126/science.288.5472.1769
- Pries, A. R., Secomb, T. W., and Gaetgens, P. (1995). Design principles of vascular beds. *Circ. Res.* 77, 1017–1023. doi: 10.1161/01.res.77.5.1017
- Proekt, A., and Hudson, A. E. (2018). A stochastic basis for neural inertia in emergence from general anaesthesia. *Br. J. Anaesth.* 121, 86–94. doi: 10.1016/j.bja.2018.02.035
- Proekt, A., and Kelz, M. (2018). Schrödinger's cat: anaesthetised and not! *Br. J. Anaesth.* 120, 424–428. doi: 10.1016/j.bja.2017.11.068
- Purdon, P. L., Pierce, E. T., Bonmassar, G., Walsh, J., Harrell, P. G., Kwo, J., et al. (2009). Simultaneous electroencephalography and functional magnetic resonance imaging of general anesthesia. *Ann. N Y Acad. Sci.* 1157, 61–70. doi: 10.1111/j.1749-6632.2008.04119.x
- Purdon, P. L., Pierce, E. T., Mukamel, E. A., Prerau, M. J., Walsh, J. L., Wong, K. F., et al. (2013). Electroencephalogram signatures of loss and recovery of consciousness from propofol. *Proc. Natl. Acad. Sci. U S A* 110, E1142–E1151. doi: 10.1073/pnas.1221180110
- Purdon, P. L., Sampson, A., Pavone, K. J., and Brown, E. N. (2015). Clinical electroencephalography for anesthesiologists: part I: background and basic signatures. *Anesthesiology* 123, 937–960. doi: 10.1097/ALN.0000000000000841
- Putnam, R. W., Filosa, J. A., and Ritucci, N. A. (2004). Cellular mechanisms involved in CO<sub>2</sub> and acid signaling in chemosensitive neurons. *Am. J. Physiol. Cell Physiol.* 287, C1493–C1526. doi: 10.1152/ajpcell.00282.2004
- Quasha, A. L., Eger, E. I. II., and Tinker, J. H. (1980). Determination and applications of MAC. *Anesthesiology* 53, 315–334. doi: 10.1097/00000542-198010000-00008
- Raichle, M. E. (2015). The brain's default mode network. *Annu. Rev. Neurosci.* 38, 433–447. doi: 10.1146/annurev-neuro-071013-014030
- Raichle, M. E., Macleod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U S A* 98, 676–682. doi: 10.1073/pnas.98.2.676
- Ranft, A., Golkowski, D., Kiel, T., Riedl, V., Kohl, P., Rohrer, G., et al. (2016). Neural correlates of sevoflurane-induced unconsciousness identified by simultaneous functional magnetic resonance imaging and electroencephalography. *Anesthesiology* 125, 861–872. doi: 10.1097/ALN.0000000000001322
- Raper, A. J., and Levasseur, J. E. (1971). Accurate sustained measurement of intra-luminal pressure from the microvasculature. *Cardiovasc. Res.* 5, 589–593. doi: 10.1093/cvr/5.4.589
- Raz, A., Grady, S. M., Krause, B. M., Uhrlich, D. J., Manning, K. A., and Banks, M. I. (2014). Preferential effect of isoflurane on top-down vs. bottom-up pathways in sensory cortex. *Front. Syst. Neurosci.* 8:191. doi: 10.3389/fnsys.2014.00191
- Redinbaugh, M. J., Phillips, J. M., Kambi, N. A., Mohanta, S., Andryk, S., Dooley, G. L., et al. (2020). Thalamus modulates consciousness via layer-specific control of cortex. *Neuron* doi: 10.1016/j.neuron.2020.01.005 [Epub ahead of print].
- Rees, G., Kreiman, G., and Koch, C. (2002). Neural correlates of consciousness in humans. *Nat. Rev. Neurosci.* 3, 261–270. doi: 10.1038/nrn783
- Reig, R., Mattia, M., Compte, A., Belmonte, C., and Sanchez-Vives, M. V. (2010). Temperature modulation of slow and fast cortical rhythms. *J. Neurophysiol.* 103, 1253–1261. doi: 10.1152/jn.00890.2009
- Reimann, H. M., Hentschel, J., Marek, J., Huelnhagen, T., Todiras, M., Kox, S., et al. (2016). Normothermic mouse functional MRI of acute focal thermostimulation for probing nociception. *Sci. Rep.* 6:17230. doi: 10.1038/srep17230
- Reimann, H. M., Todiras, M., Hodge, R., Huelnhagen, T., Millward, J. M., Turner, R., et al. (2018). Somatosensory BOLD fMRI reveals close link between salient blood pressure changes and the murine neuromatrix. *NeuroImage* 172, 562–574. doi: 10.1016/j.neuroimage.2018.02.002
- Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., et al. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat. Commun.* 7:13289. doi: 10.1038/ncomms13289
- Roberts, J. A., Gollo, L. L., Abeyesuriya, R. G., Roberts, G., Mitchell, P. B., Woolrich, M. W., et al. (2019). Metastable brain waves. *Nat. Commun.* 10:1056. doi: 10.1038/s41467-019-08999-0
- Roth, B. L. (2016). DREADDs for neuroscientists. *Neuron* 89, 683–694. doi: 10.1016/j.neuron.2016.01.040
- Rubin, R., Abbott, L. F., and Sompolinsky, H. (2017). Balanced excitation and inhibition are required for high-capacity, noise-robust neuronal selectivity. *Proc. Natl. Acad. Sci. U S A* 114, E9366–E9375. doi: 10.1073/pnas.1705841114
- Ruiz-Mejias, M., Ciria-Suarez, L., Mattia, M., and Sanchez-Vives, M. V. (2011). Slow and fast rhythms generated in the cerebral cortex of the anesthetized mouse. *J. Neurophysiol.* 106, 2910–2921. doi: 10.1152/jn.00440.2011
- Ryali, S., Shih, Y.-Y. I., Chen, T., Kochalka, J., Albaugh, D., Fang, Z., et al. (2016). Combining optogenetic stimulation and fMRI to validate a multivariate dynamical systems model for estimating causal brain interactions. *NeuroImage* 132, 398–405. doi: 10.1016/j.neuroimage.2016.02.067
- Saalmann, Y. B. (2014). Intralaminar and medial thalamic influence on cortical synchrony, information transmission and cognition. *Front. Syst. Neurosci.* 8:83. doi: 10.3389/fnsys.2014.00083
- Saalmann, Y. B., Pinsk, M. A., Wang, L., Li, X., and Kastner, S. (2012). The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 337, 753–756. doi: 10.1126/science.1223082
- Sachidhanandam, S., Sreenivasan, V., Kyriakatos, A., Kremer, Y., and Petersen, C. C. H. (2013). Membrane potential correlates of sensory perception in mouse barrel cortex. *Nat. Neurosci.* 16, 1671–1677. doi: 10.1038/nn.3532
- Sadaghiani, S., Scheeringa, R., Lehongre, K., Morillon, B., Giraud, A. L., D'Esposito, M., et al. (2012).  $\alpha$ -band phase synchrony is related to activity in the fronto-parietal adaptive control network. *J. Neurosci.* 32, 14305–14310. doi: 10.1523/jneurosci.1358-12.2012
- Samuels, E. R., and Szabadi, E. (2008a). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part I: principles of functional organisation. *Curr. Neuropharmacol.* 6, 235–253. doi: 10.2174/157015908785777229
- Samuels, E. R., and Szabadi, E. (2008b). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part II: physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. *Curr. Neuropharmacol.* 6, 254–285. doi: 10.2174/157015908785777193
- Sanchez-Vives, M. V., Massimini, M., and Mattia, M. (2017). Shaping the default activity pattern of the cortical network. *Neuron* 94, 993–1001. doi: 10.1016/j.neuron.2017.05.015
- Sanders, T. H. (2017). Stimulation of cortico-subthalamic projections amplifies resting motor circuit activity and leads to increased locomotion in dopamine-depleted mice. *Front. Integr. Neurosci.* 11:24. doi: 10.3389/fnint.2017.00024
- Sanders, R. D., Banks, M. I., Darracq, M., Moran, R., Sleight, J., Gosseries, O., et al. (2018). Propofol-induced unresponsiveness is associated with impaired feedforward connectivity in cortical hierarchy. *Br. J. Anaesth.* 121, 1084–1096. doi: 10.1016/j.bja.2018.07.006
- Sanders, R. D., Degos, V., and Young, W. L. (2011). Cerebral perfusion under pressure: is the autoregulatory “plateau” a level playing field for all? *Anaesthesia* 66, 968–972. doi: 10.1111/j.1365-2044.2011.06915.x
- Sanders, R., Tononi, G., Laureys, S., and Sleight, J. (2012). Unresponsiveness  $\neq$  unconsciousness. *Anesthesiology* 116, 946–959. doi: 10.1097/ALN.0b013e318249d0a7
- Saper, C. B., Fuller, P. M., Pedersen, N. P., Lu, J., and Scammell, T. E. (2010). Sleep state switching. *Neuron* 68, 1023–1042. doi: 10.1016/j.neuron.2010.11.032
- Saper, C. B., Scammell, T. E., and Lu, J. (2005). Hypothalamic regulation of sleep and circadian rhythms. *Nature* 437, 1257–1263. doi: 10.1038/nature04284
- Sarasso, S., Boly, M., Napolitani, M., Gosseries, O., Charland-Verville, V., Casarotto, S., et al. (2015). Consciousness and complexity during unresponsiveness induced by propofol, xenon, and ketamine. *Curr. Biol.* 25, 3099–3105. doi: 10.1016/j.cub.2015.10.014
- Scammell, T. E., Arrigoni, E., and Lipton, J. O. (2017). Neural circuitry of wakefulness and sleep. *Neuron* 93, 747–765. doi: 10.1016/j.neuron.2017.01.014

- Schlegel, F. (2017). *Improving Mouse FMRI and Our Understanding of Its Underpinnings*. [Dissertation] Zürich, Switzerland: ETH Zurich.
- Schlegel, F., Schroeter, A., and Rudin, M. (2015). The hemodynamic response to somatosensory stimulation in mice depends on the anesthetic used: implications on analysis of mouse fMRI data. *NeuroImage* 116, 40–49. doi: 10.1016/j.neuroimage.2015.05.013
- Schlegel, F., Sych, Y., Schroeter, A., Stobart, J., Weber, B., Helmchen, F., et al. (2018). Fiber-optic implant for simultaneous fluorescence-based calcium recordings and BOLD fMRI in mice. *Nat. Protoc.* 13, 840–855. doi: 10.1038/nprot.2018.003
- Schmid, F., Wachsmuth, L., Schwalm, M., Prouvot, P. H., Jubal, E. R., Fois, C., et al. (2016). Assessing sensory versus optogenetic network activation by combining (o)fMRI with optical  $\text{Ca}^{2+}$  recordings. *J. Cereb. Blood Flow Metab.* 36, 1885–1900. doi: 10.1177/0271678x15619428
- Schneider, D. W., and Logan, G. D. (2014). Modelling response selection in task switching: testing the contingent encoding assumption. *Q. J. Exp. Psychol.* 67, 1074–1095. doi: 10.1080/17470218.2013.843009
- Schölvinck, M. L., Saleem, A. B., Benucci, A., Harris, K. D., and Carandini, M. (2015). Cortical state determines global variability and correlations in visual cortex. *J. Neurosci.* 35, 170–178. doi: 10.1523/jneurosci.4994-13.2015
- Schroeder, C. E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosci.* 32, 9–18. doi: 10.1016/j.tins.2008.09.012
- Schroeder, K. E., Irwin, Z. T., Gaidica, M., Nicole Bentley, J., Patil, P. G., Mashour, G. A., et al. (2016). Disruption of corticocortical information transfer during ketamine anesthesia in the primate brain. *NeuroImage* 134, 459–465. doi: 10.1016/j.neuroimage.2016.04.039
- Schroeter, A., Grandjean, J., Schlegel, F., Saab, B. J., and Rudin, M. (2017). Contributions of structural connectivity and cerebrovascular parameters to functional magnetic resonance imaging signals in mice at rest and during sensory paw stimulation. *J. Cereb. Blood Flow Metab.* 37, 2368–2382. doi: 10.1177/0271678x16666292
- Schroeter, A., Schlegel, F., Seuwen, A., Grandjean, J., and Rudin, M. (2014). Specificity of stimulus-evoked fMRI responses in the mouse: the influence of systemic physiological changes associated with innocuous stimulation under four different anesthetics. *NeuroImage* 94, 372–384. doi: 10.1016/j.neuroimage.2014.01.046
- Schröter, M. S., Spoormaker, V. I., Schorer, A., Wohlschläger, A., Czisch, M., Kochs, E. F., et al. (2012). Spatiotemporal reconfiguration of large-scale brain functional networks during propofol-induced loss of consciousness. *J. Neurosci.* 32, 12832–12840. doi: 10.1523/jneurosci.6046-11.2012
- Schumacher, J. W., Schneider, D. M., and Woolley, S. M. N. (2011). Anesthetic state modulates excitability but not spectral tuning or neural discrimination in single auditory midbrain neurons. *Neuroscience* 106, 500–514. doi: 10.1152/jn.01072.2010
- Schwalm, M., and Easton, C. (2016). Cortical temperature change: a tool for modulating brain states? *eNeuro* 3:ENEURO.0096-16.2016. doi: 10.1523/eneuro.0096-16.2016
- Schwalm, M., Schmid, F., Wachsmuth, L., Backhaus, H., Kronfeld, A., Jury, F. A., et al. (2017). Cortex-wide BOLD fMRI activity reflects locally-recorded slow oscillation-associated calcium waves. *Elife* 6:e27602. doi: 10.7554/elifesciences.27602
- Schwertner, A., Zortea, M., Torres, F. V., and Caumo, W. (2018). Effects of subanesthetic ketamine administration on visual and auditory event-related potentials (ERP) in humans: a systematic review. *Front. Behav. Neurosci.* 12:70. doi: 10.3389/fnbeh.2018.00070
- Seamans, J. (2008). Losing inhibition with ketamine. *Nat. Chem. Biol.* 4, 91–93. doi: 10.1038/nchembio0208-91
- Sellers, K. K., Bennett, D. V., Hutt, A., Williams, J. H., and Fröhlich, F. (2015). Awake vs. anesthetized: layer-specific sensory processing in visual cortex and functional connectivity between cortical areas. *J. Neurophysiol.* 113, 3798–3815. doi: 10.1152/jn.00923.2014
- Senzai, Y., Fernandez-Ruiz, A., and Buzsáki, G. (2019). Layer-specific physiological features and interlaminar interactions in the primary visual cortex of the mouse. *Neuron* 101, 500.e5–513.e5. doi: 10.1016/j.neuron.2018.12.009
- Seth, A. K., Baars, B. J., and Edelman, D. B. (2005). Criteria for consciousness in humans and other mammals. *Conscious. Cogn.* 14, 119–139. doi: 10.1016/j.concog.2004.08.006
- Seth, A. K., and Tsakiris, M. (2018). Being a beast machine: the somatic basis of selfhood. *Trends Cogn. Sci.* 22, 969–981. doi: 10.1016/j.tics.2018.08.008
- Sforzazzini, F., Schwarz, A. J., Galbusera, A., Bifone, A., and Gozzi, A. (2014). Distributed BOLD and CBV-weighted resting-state networks in the mouse brain. *NeuroImage* 87, 403–415. doi: 10.1016/j.neuroimage.2013.09.050
- Shah, D., Blockx, I., Guns, P., Paul, P., Deyn, D., Van Dam, D., et al. (2015). Acute modulation of the cholinergic system in the mouse brain detected by pharmacological resting-state functional MRI. *NeuroImage* 109, 151–159. doi: 10.1016/j.neuroimage.2015.01.009
- Shah, D., Jonckers, E., Praet, J., Vanhoutte, G., Delgado, Y. P., Bigot, C., et al. (2013). Resting state fMRI reveals diminished functional connectivity in a mouse model of amyloidosis. *PLoS One* 8:e84241. doi: 10.1371/journal.pone.0084241
- Sharp, P. S., Shaw, K., Boorman, L., Harris, S., Kennerley, A. J., Azzouz, M., et al. (2015). Comparison of stimulus-evoked cerebral hemodynamics in the awake mouse and under a novel anesthetic regime. *Sci. Rep.* 5:14890. doi: 10.1038/srep14890
- Shein-Idelson, M., Ondracek, J. M., Liaw, H. P., Reiter, S., and Laurent, G. (2016). Slow waves, sharp waves, ripples, and REM in sleeping dragons. *Science* 352, 590–595. doi: 10.1126/science.aaf3621
- Shelton, L., Becerra, L., and Borsook, D. (2012a). Unmasking the mysteries of the habenula in pain and analgesia. *Prog. Neurobiol.* 96, 208–219. doi: 10.1016/j.pneurobio.2012.01.004
- Shelton, L., Pendse, G., Maleki, N., Moulton, E. A., Lebel, A., Becerra, L., et al. (2012b). Mapping pain activation and connectivity of the human habenula. *J. Neurophysiol.* 107, 2633–2648. doi: 10.1152/jn.00012.2012
- Sherin, J. E., Elmquist, J. K., Torrealba, F., and Saper, C. B. (1998). Innervation of histaminergic tuberomammillary neurons by GABAergic and galaninergic neurons in the ventrolateral preoptic nucleus of the rat. *J. Neurosci.* 18, 4705–4721. doi: 10.1523/jneurosci.18-12-04705.1998
- Sherman, S. M. (2017). Functioning of circuits connecting thalamus and cortex. *Compr. Physiol.* 7, 713–739. doi: 10.1002/cphy.c160032
- Sheroziya, M., and Timofeev, I. (2014). Global intracellular slow-wave dynamics of the thalamocortical system. *J. Neurosci.* 34, 8875–8893. doi: 10.1523/jneurosci.4460-13.2014
- Sheroziya, M., and Timofeev, I. (2015). Moderate cortical cooling eliminates thalamocortical silent states during slow oscillation. *J. Neurosci.* 35, 13006–13019. doi: 10.1523/jneurosci.1359-15.2015
- Shim, H.-J., Jung, W. B., Schlegel, F., Lee, J., Kim, S., Lee, J., et al. (2018). Mouse fMRI under ketamine and xylazine anesthesia: robust contralateral somatosensory cortex activation in response to forepaw stimulation. *NeuroImage* 177, 30–44. doi: 10.1016/j.neuroimage.2018.04.062
- Shimaoka, D., Harris, K. D., and Carandini, M. (2018). Effects of arousal on mouse sensory cortex depend on modality. *Cell Rep.* 22, 3160–3167. doi: 10.1016/j.celrep.2018.02.092
- Shimosegawa, E., Kanno, I., Hatazawa, J., Fujita, H., Iida, H., Miura, S., et al. (1995). Photostimulation study of changing the arterial partial pressure level of carbon dioxide. *J. Cereb. Blood Flow Metab.* 15, 111–114. doi: 10.1038/jcbfm.1995.12
- Shofty, B., Bergmann, E., Zur, G., Asleh, J., Bosak, N., Kavushansky, A., et al. (2019). Autism-associated Nf1 deficiency disrupts corticocortical and corticostriatal functional connectivity in human and mouse. *Neurobiol. Dis.* 130:104479. doi: 10.1016/j.nbd.2019.104479
- Siklens, T., Bosman, C. A., and Olcese, U. (2019). The role of top-down modulation in shaping sensory processing across brain states: implications for consciousness. *Front. Syst. Neurosci.* 13:31. doi: 10.3389/fnsys.2019.00031
- Sinclair, M. D. (2003). A review of the physiological effects of  $\alpha$ 2-agonists related to the clinical use of medetomidine in small animal practice. *Can. Vet. J.* 44, 885–897.
- Sinclair, D. R., Chung, F., and Smiley, A. (2003). General anesthesia does not impair simulator driving skills in volunteers in the immediate recovery period—a pilot study. *Can. J. Anaesth.* 50, 238–245. doi: 10.1007/bf03017791

- Sinner, B., and Graf, B. M. (2008). Ketamine. *Handb. Exp. Pharmacol.* 182, 313–333. doi: 10.1007/978-3-540-74806-9\_15
- Sinning, A., and Hübner, C. A. (2013). Minireview: pH and synaptic transmission. *FEBS Lett.* 587, 1923–1928. doi: 10.1016/j.febslet.2013.04.045
- Sirmipilatz, N., Baudewig, J., and Boretius, S. (2019). Temporal stability of fMRI in medetomidine-anesthetized rats. *Sci. Rep.* 9:16673. doi: 10.1038/s41598-019-53144-y
- Sirota, A., Csicsvari, J., Buhl, D., and Buzsáki, G. (2003). Communication between neocortex and hippocampus during sleep in rodents. *Proc. Natl. Acad. Sci. U S A* 100, 2065–2069. doi: 10.1073/pnas.0437938100
- Sleigh, J., Warnaby, C., and Tracey, I. (2018). General anaesthesia as fragmentation of selfhood: insights from electroencephalography and neuroimaging. *Br. J. Anaesth.* 121, 233–240. doi: 10.1016/j.bja.2017.12.038
- Slézia, A., Hangya, B., Ulbert, I., and Acsády, L. (2011). Phase advancement and nucleus-specific timing of thalamocortical activity during slow cortical oscillation. *J. Neurosci.* 31, 607–617. doi: 10.1523/jneurosci.3375-10.2011
- Solt, K., Van Dort, C. J., Chemali, J. J., Taylor, N. E., Kenny, J. D., and Brown, E. N. (2014). Electrical stimulation of the ventral tegmental area induces reanimation from general anesthesia. *Anesthesiology* 121, 311–319. doi: 10.1097/ALN.0000000000000117
- Song, A. H., Kucyi, A., Napadow, V., Brown, E. N., Loggia, M. L., and Akeju, O. (2017). Pharmacological modulation of noradrenergic arousal circuitry disrupts functional connectivity of the locus ceruleus in humans. *J. Neurosci.* 37, 6938–6945. doi: 10.1523/jneurosci.0446-17.2017
- Song, B., Ma, N., Liu, G., Zhang, H., Yu, L., Liu, L., et al. (2019). Maximal flexibility in dynamic functional connectivity with critical dynamics revealed by fMRI data analysis and brain network modelling. *J. Neural Eng.* 16:056002. doi: 10.1088/1741-2552/ab20bc
- Soltesz, I., Lightowler, S., Leresche, N., Jassik-Gerschenfeld, D., Pollard, C. E., and Crunelli, V. (1991). Two inward currents and the transformation of low frequency oscillations of rat and cat thalamocortical cells. *J. Physiol.* 441, 175–197. doi: 10.1113/jphysiol.1991.sp018745
- Stafford, J. M., Jarrett, B. R., Miranda-Dominguez, O., Mills, B. D., Cain, N., Mihalas, S., et al. (2014). Large-scale topology and the default mode network in the mouse connectome. *Proc. Natl. Acad. Sci. U S A* 111, 18745–18750. doi: 10.1073/pnas.1404346111
- Steffey, E. P. (2017). Methodology for determining minimum alveolar concentration: a critical appraisal. *Vet. Anaesth. Analg.* 44, 2–6. doi: 10.1016/j.vaa.2016.12.001
- Steriade, M., Curró Dossi, R., and Nunez, A. (1991). Network modulation of a slow intrinsic oscillation of cat thalamocortical neurons implicated in sleep delta waves: cortical potentiation and brainstem cholinergic suppression. *J. Neurosci.* 11, 3200–3217. doi: 10.1523/jneurosci.11-10-03200.1991
- Steriade, M. (2000). Corticothalamic resonance, states of vigilance and mentation. *Neuroscience* 101, 243–276. doi: 10.1016/s0306-4522(00)00353-5
- Steriade, M. (2001). Impact of network activities on neuronal properties in corticothalamic systems. *J. Neurophysiol.* 86, 1–39. doi: 10.1152/jn.2001.86.1.1
- Steriade, M., Amzica, F., and Contreras, D. (1994). Cortical and thalamic cellular correlates of electroencephalographic burst-suppression. *Electroencephalogr. Clin. Neurophysiol.* 90, 1–16. doi: 10.1016/0013-4694(94)90108-2
- Steriade, M., Amzica, F., and Nufiez, A. (1993a). Cholinergic and noradrenergic modulation of the slow (approximately 0.3 Hz) oscillation in neocortical cells. *J. Neurophysiol.* 70, 1385–1400. doi: 10.1152/jn.1993.70.4.1385
- Steriade, M., Contreras, D., Curró Dossi, R., and Nuñez, A. (1993b). The slow (<1 Hz) oscillation in reticular thalamic and thalamocortical neurons: scenario of sleep rhythm generation in interacting thalamic and neocortical networks. *J. Neurosci.* 13, 3284–3299. doi: 10.1523/jneurosci.13-08-03284.1993
- Steriade, M., Nuñez, A., and Amzica, F. (1993c). Intracellular analysis of relations between the slow (<1 Hz) neocortical oscillation and other sleep rhythms of the electroencephalogram. *J. Neurosci.* 13, 3266–3283. doi: 10.1523/jneurosci.13-08-03266.1993
- Steriade, M., Nuñez, A., and Amzica, F. (1993d). A novel slow (<1 Hz) oscillation of neocortical neurons *in vivo*: depolarizing and hyperpolarizing components. *J. Neurosci.* 13, 3252–3265. doi: 10.1523/jneurosci.13-08-03252.1993
- Steriade, M., and Contreras, D. (1995). Relations between cortical and thalamic cellular events during transition from sleep pattern to paroxysmal activity. *J. Neurosci.* 15, 623–642. doi: 10.1523/jneurosci.15-01-00623.1995
- Steriade, M., Deschenes, M., Domich, L., and Mulle, C. (1985). Abolition of spindle oscillations in thalamic neurons disconnected from nucleus reticularis thalami. *J. Neurophysiol.* 54, 1473–1497. doi: 10.1152/jn.1985.54.6.1473
- Steriade, M., Domich, L., Oakson, G., and Deschênes, M. (1987). The deafferented reticularis thalami nucleus generates spindle rhythmicity. *J. Neurophysiol.* 57, 260–273. doi: 10.1152/jn.1987.57.1.260
- Steriade, M., Timofeev, I., and Grenier, F. (2001). Natural waking and sleep states: a view from inside neocortical neurons. *J. Neurophysiol.* 85, 1969–1985. doi: 10.1152/jn.2001.85.5.1969
- Storm, J. F., Boly, M., Casali, A. G., Massimini, M., Olcese, U., Pennartz, C. M. A., et al. (2017). Consciousness regained: disentangling mechanisms, brain systems and behavioral responses. *J. Neurosci.* 37, 10882–10893. doi: 10.1523/jneurosci.1838-17.2017
- Straathof, M., Sinke, M. R., Dijkhuizen, R. M., and Otte, W. M. (2019). A systematic review on the quantitative relationship between structural and functional network connectivity strength in mammalian brains. *J. Cereb. Blood Flow Metab.* 39, 189–209. doi: 10.1177/0271678x18809547
- Stroh, A., Adelsberger, H., Groh, A., Rühlmann, C., Fischer, S., Schierloh, A., et al. (2013). Making waves: initiation and propagation of corticothalamic  $Ca^{2+}$  waves *in vivo*. *Neuron* 77, 1136–1150. doi: 10.1016/j.neuron.2013.01.031
- Sumiyoshi, A., Keeley, R. J., and Lu, H. (2019). Physiological considerations of functional magnetic resonance imaging in animal models. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 4, 522–532. doi: 10.1016/j.bpsc.2018.08.002
- Sun, Y. M., Favre, I., Schild, L., and Moczydlowski, E. (1997). On the structural basis for size-selective permeation of organic cations through the voltage-gated sodium channel. Effect of alanine mutations at the DEKA locus on selectivity, inhibition by  $Ca^{2+}$  and  $H^{+}$  and molecular sieving. *J. Gen. Physiol.* 110, 693–715. doi: 10.1085/jgp.110.6.693
- Sun, W., McConnell, E., Pare, J. F., Xu, Q., Chen, M., Peng, W., et al. (2013). Glutamate-dependent neuroglial calcium signaling differs between young and adult brain. *Science* 339, 197–200. doi: 10.1126/science.1226740
- Sun, Y. G., Wu, C. S., Renger, J. J., Uebele, V. N., Lu, H. C., and Beierlein, M. (2012). GABAergic synaptic transmission triggers action potentials in thalamic reticular nucleus neurons. *J. Neurosci.* 32, 7782–7790. doi: 10.1523/jneurosci.0839-12.2012
- Sunkin, S. M., Ng, L., Lau, C., Dolbeare, T., Gilbert, T. L., Thompson, C. L., et al. (2013). Allen brain atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res.* 41, D996–D1008. doi: 10.1093/nar/gks1042
- Suzuki, M., and Larkum, M. E. (2020). General anesthesia decouples cortical pyramidal neurons. *Cell* 180, 666.e13–676.e13. doi: 10.1016/j.cell.2020.01.024
- Takuwa, H., Matsuura, T., Obata, T., Kawaguchi, H., Kanno, I., and Ito, H. (2012). Hemodynamic changes during somatosensory stimulation in awake and isoflurane-anesthetized mice measured by laser-Doppler flowmetry. *Brain Res.* 1472, 107–112. doi: 10.1016/j.brainres.2012.06.049
- Taylor, N. E., Van Dort, C. J., Kenny, J. D., Pei, J., Guidera, J. A., Vlasov, K. Y., et al. (2016). Optogenetic activation of dopamine neurons in the ventral tegmental area induces reanimation from general anesthesia. *Proc. Natl. Acad. Sci. U S A* 113, 12826–12831. doi: 10.1073/pnas.1614340113
- Theyel, B. B., Llano, D. A., and Sherman, S. M. (2010). The corticothalamic circuit drives higher-order cortex in the mouse. *Nat. Neurosci.* 13, 84–88. doi: 10.1038/nn.2449
- Thompson, G. J., Merritt, M. D., Pan, W. J., Magnuson, M. E., Grooms, J. K., Jaeger, D., et al. (2013). Neural correlates of time-varying functional connectivity in the rat. *NeuroImage* 83, 826–836. doi: 10.1016/j.neuroimage.2013.07.036
- Timofeev, I., Grenier, F., Bazhenov, M., Sejnowski, T. J., and Steriade, M. (2000). Origin of slow cortical oscillations in deafferented cortical slabs. *Cereb. Cortex* 10, 1185–1199. doi: 10.1093/cercor/10.12.1185



- Tombaugh, G. C., and Somjen, G. G. (1996). Effects of extracellular pH on voltage-gated  $\text{Na}^+$ ,  $\text{K}^+$  and  $\text{Ca}^{2+}$  currents in isolated rat CA1 neurons. *J. Physiol.* 493, 719–732. doi: 10.1113/jphysiol.1996.sp021417
- Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461. doi: 10.1038/nrn.2016.44
- Toussay, X., Basu, K., Lacoste, B., and Hamel, E. (2013). Locus coeruleus stimulation recruits a broad cortical neuronal network and increases cortical perfusion. *J. Neurosci.* 33, 3390–3401. doi: 10.1523/jneurosci.3346-12.2013
- Tracey, I. (2017). Neuroimaging mechanisms in pain: from discovery to translation. *Pain* 158, S115–S122. doi: 10.1097/j.pain.0000000000000863
- Trageser, J. C., Burke, K. A., Masri, R., Li, Y., Sellers, L., and Keller, A. (2006). State-dependent gating of sensory inputs by zona incerta. *J. Neurophysiol.* 96, 1456–1463. doi: 10.1152/jn.00423.2006
- Uchida, S., and Kagitani, F. (2018). Effect of basal forebrain stimulation on extracellular acetylcholine release and blood flow in the olfactory bulb. *J. Physiol. Sci.* 68, 415–423. doi: 10.1007/s12576-017-0542-z
- Uhrig, L., Sitt, J. D., Jacob, A., Tasserie, J., Barttfeld, P., Dupont, M., et al. (2018). Resting-state dynamics as a cortical signature of anesthesia in monkeys. *Anesthesiology* 129, 942–958. doi: 10.1097/aln.0000000000002336
- Ulrich-Lai, Y. M., and Herman, J. P. (2009). Neural regulation of endocrine and autonomic stress responses. *Nat. Rev. Neurosci.* 10, 397–409. doi: 10.1038/nrn2647
- van Alst, T. M., Wachsmuth, L., Datunashvili, M., Albers, F., Just, N., Budde, T., et al. (2019). Anesthesia differentially modulates neuronal and vascular contributions to the BOLD signal. *NeuroImage* 195, 89–103. doi: 10.1016/j.neuroimage.2019.03.057
- Van Bockstaele, E. J., Saunders, A., Telegan, P., and Page, M. E. (1999). Localization of mu-opioid receptors to locus coeruleus-projecting neurons in the rostral medulla: morphological substrates and synaptic organization. *Synapse* 34, 154–167. doi: 10.1002/(SICI)1098-2396(199911)34:2<154::AID-SYN8>3.0.CO;2-C
- Van de Ville, D., Britz, J., and Michel, C. M. (2010). EEG microstate sequences in healthy humans at rest reveal scale-free dynamics. *Proc. Natl. Acad. Sci. U S A* 107, 18179–18184. doi: 10.1073/pnas.1007841107
- van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M. A., Poort, J., van der Togt, C., et al. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. U S A* 111, 14332–14341. doi: 10.1073/pnas.1402773111
- Varela, C. (2014). Thalamic neuromodulation and its implications for executive networks. *Front. Neural Circuits* 8:69. doi: 10.3389/fncir.2014.00069
- Vazey, E. M., and Aston-Jones, G. (2014). Designer receptor manipulations reveal a role of the locus coeruleus noradrenergic system in isoflurane general anesthesia. *Proc. Natl. Acad. Sci. U S A* 111, 3859–3864. doi: 10.1073/pnas.1310025111
- Venn, R. M., and Grounds, R. M. (2001). Comparison between dexmedetomidine and propofol for sedation in the intensive care unit: patient and clinician perceptions. *Br. J. Anaesth.* 87, 684–690. doi: 10.1093/bja/87.5.684
- Ventura-Antunes, L., Mota, B., and Herculano-Houzel, S. (2013). Different scaling of white matter volume, cortical connectivity and gyrification across rodent and primate brains. *Front. Neuroanat.* 7:3. doi: 10.3389/fnana.2013.00003
- Veselis, R. A., Feshchenko, V. A., Reinsel, R. A., Beattie, B., and Akhurst, T. J. (2005). Propofol and thiopental do not interfere with regional cerebral blood flow response at sedative concentrations. *Anesthesiology* 102, 26–34. doi: 10.1097/0000542-200501000-00008
- Vetter, I., Touska, F., Hess, A., Hinsbey, R., Sattler, S., Lampert, A., et al. (2012). Ciguatoxins activate specific cold pain pathways to elicit burning pain from cooling. *EMBO J.* 3, 3795–3808. doi: 10.1038/emboj.2012.207
- Vijayan, S., Ching, S., Purdon, P. L., Brown, E. N., and Kopell, N. J. (2013). Thalamocortical mechanisms for the anteriorization of  $\alpha$  rhythms during propofol-induced unconsciousness. *J. Neurosci.* 33, 11070–11075. doi: 10.1523/jneurosci.5670-12.2013
- Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Van Essen, D. C., et al. (2007). Intrinsic functional architecture in the anesthetized monkey brain. *Nature* 447, 83–86. doi: 10.1038/nature05758
- Vizuete, J. A., Pillay, S., Ropella, K. M., and Hudetz, A. G. (2014). Graded defragmentation of cortical neuronal firing during recovery of consciousness in rats. *Neuroscience* 275, 340–351. doi: 10.1016/j.neuroscience.2014.06.018
- Vlissides, P. E., Bel-Bahar, T., Lee, U., Li, D., Kim, H., Janke, E., et al. (2017). Neurophysiologic correlates of ketamine sedation and anesthesia: a high-density electroencephalography study in healthy volunteers. *Anesthesiology* 127, 58–69. doi: 10.1097/aln.0000000000001671
- Vogt, B. A., Hof, P. R., Friedman, D. P., Sikes, R. W., and Vogt, L. J. (2008). Norepinephrine afferents and cytology of the macaque monkey midline, mediodorsal and intralaminar thalamic nuclei. *Brain Struct. Funct.* 212, 465–479. doi: 10.1007/s00429-008-0178-0
- Vogt, B. A., and Pandya, D. N. (1987). Cingulate cortex of the rhesus monkey: II. Cortical afferents. *J. Comp. Neurol.* 262, 271–289. doi: 10.1002/cne.902620208
- Volgushev, M., Vidyasagar, T. R., Chistiakova, M., Yousef, T., and Eysel, U. T. (2000). Membrane properties and spike generation in rat visual cortical cells during reversible cooling. *J. Physiol.* 522, 59–76. doi: 10.1111/j.1469-7793.2000.0059m.x
- Voss, B., Thienel, R., Reske, M., Kellermann, T., Sheldrick, A. J., Halfter, S., et al. (2012). Cholinergic blockade under working memory demands encountered by increased rehearsal strategies: evidence from fMRI in healthy subjects. *Eur. Arch. Psychiatry Clin. Neurosci.* 262, 329–339. doi: 10.1007/s00406-011-0267-6
- Vyazovskiy, V. V., Olcese, U., Hanlon, E. C., Nir, Y., Cirelli, C., and Tononi, G. (2011). Local sleep in awake rats. *Nature* 472, 443–447. doi: 10.1038/nature10009
- Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C. W., and Kross, E. (2013). An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* 368, 1388–1397. doi: 10.1056/NEJMoa1204471
- Wang, R., Foniok, T., Wamsteeker, J. I., Qiao, M., Tomanek, B., Vivanco, R. A., et al. (2006). Transient blood pressure changes affect the functional magnetic resonance imaging detection of cerebral activation. *NeuroImage* 31, 1–11. doi: 10.1016/j.neuroimage.2005.12.004
- Wang, R., Macmillan, L. B., Freneau, R. T. Jr., Magnuson, M. A., Lindner, J., and Limbird, L. E. (1996). Expression of  $\alpha$ 2-adrenergic receptor subtypes in the mouse brain: evaluation of spatial and temporal information imparted by 3 kb of 5' regulatory sequence for the  $\alpha$  2A AR-receptor gene in transgenic animals. *Neuroscience* 74, 199–218. doi: 10.1016/0306-4522(96)00116-9
- Warnaby, C. E., Seretny, M., Ni Mhuircheartaigh, R., Rogers, R., Jbabdi, S., Sleight, J., et al. (2016). Anesthesia-induced suppression of human dorsal anterior insula responsivity at loss of volitional behavioral response. *Anesthesiology* 124, 766–778. doi: 10.1097/aln.0000000000001027
- Weber, F., and Dan, Y. (2016). Circuit-based interrogation of sleep control. *Nature* 538, 51–59. doi: 10.1038/nature19773
- Wilding, L. A., Hampel, J. A., Khoury, B. M., Kang, S., Machado-Aranda, D., Raghavendran, K., et al. (2017). Benefits of 21% oxygen compared with 100% oxygen for delivery of isoflurane to mice (*Mus musculus*) and rats (*Rattus norvegicus*). *J. Am. Assoc. Lab. Anim. Sci.* 56, 148–154.
- Williams, R. H., Jensen, L. T., Verkhatsky, A., Fugger, L., and Burdakov, D. (2007). Control of hypothalamic orexin neurons by acid and CO<sub>2</sub>. *Proc. Natl. Acad. Sci. U S A* 104, 10685–10690. doi: 10.1073/pnas.0702676104
- Wimmer, R. D., Astori, S., Bond, C. T., Rovó, Z., Chatton, J. Y., Adelman, J. P., et al. (2012). Sustaining sleep spindles through enhanced SK2-channel activity consolidates sleep and elevates arousal threshold. *J. Neurosci.* 32, 13917–13928. doi: 10.1523/jneurosci.2313-12.2012
- Winkler, S. A., Schmitt, F., Landes, H., de Bever, J., Wade, T., Alejski, A., et al. (2018). Gradient and shim technologies for ultra high field MRI. *NeuroImage* 168, 59–70. doi: 10.1016/j.neuroimage.2016.11.033
- Woo, C. W., Schmidt, L., Krishnan, A., Jepma, M., Roy, M., Lindquist, M. A., et al. (2017). Quantifying cerebral contributions to pain beyond nociception. *Nat. Commun.* 8:14211. doi: 10.1038/ncomms14211
- Wu, T., Grandjean, J., Bosshard, S. C., Rudin, M., Reutens, D., and Jiang, T. (2017). Altered regional connectivity reflecting effects of different anaesthesia protocols in the mouse brain. *NeuroImage* 149, 190–199. doi: 10.1016/j.neuroimage.2017.01.074
- Xu, F., Liu, N., Kida, I., Rothman, D. L., Hyder, F., and Shepherd, G. M. (2003). Odor maps of aldehydes and esters revealed by functional MRI in the



- glomerular layer of the mouse olfactory bulb. *Proc. Natl. Acad. Sci. U S A* 100, 11029–11034. doi: 10.1073/pnas.1832864100
- Xu, F., Schaefer, M., Kida, I., Schafer, J., Liu, N., Rothman, D. L., et al. (2005). Simultaneous activation of mouse main and accessory olfactory bulbs by odors or pheromones. *J. Comp. Neurol.* 489, 491–500. doi: 10.1002/cne.20652
- Yli-Hankala, A., Lindgren, L., Porkkala, T., and Jantti, V. (1993). Nitrous oxide-mediated activation of the EEG during isoflurane anaesthesia in patients. *Br. J. Anaesth.* 70, 54–57. doi: 10.1093/bja/70.1.54
- Yoshida, K., Mimura, Y., Ishihara, R., Nishida, H., Komaki, Y., Minakuchi, T., et al. (2016). Physiological effects of a habituation procedure for functional MRI in awake mice using a cryogenic radiofrequency probe. *J. Neurosci. Methods* 274, 38–48. doi: 10.1016/j.jneumeth.2016.09.013
- Yuste, R. (2015). From the neuron doctrine to neural networks. *Nat. Rev. Neurosci.* 16, 487–497. doi: 10.1038/nrn3962
- Záborszky, L., Gombkoto, P., Varsanyi, P., Gielow, M. R., Poe, G., Role, L. W., et al. (2018). Specific basal forebrain-cortical cholinergic circuits coordinate cognitive operations. *J. Neurosci.* 38, 9446–9458. doi: 10.1523/jneurosci.1676-18.2018
- Zaman, T., Lee, K., Park, C., Paydar, A., Choi, J. H., Cheong, E., et al. (2011). Cav2.3 channels are critical for oscillatory burst discharges in the reticular thalamus and absence epilepsy. *Neuron* 70, 95–108. doi: 10.1016/j.neuron.2011.02.042
- Zandieh, S., Hopf, R., Redl, H., and Schlag, M. G. (2003). The effect of ketamine/xylazine anesthesia on sensory and motor evoked potentials in the rat. *Spinal Cord* 41, 16–22. doi: 10.1038/sj.sc.3101400
- Zeman, A. (2004). Theories of visual awareness. *Prog. Brain Res.* 144, 321–329. doi: 10.1016/s0079-6123(03)14422-6
- Zerbi, V., Floriou-Servou, A., Markicevic, M., Vermeiren, Y., Sturman, O., Privitera, M., et al. (2019). Rapid reconfiguration of the functional connectome after chemogenetic locus coeruleus activation. *Neuron* 103, 702.e5–718.e5. doi: 10.1016/j.neuron.2019.05.034
- Zerbi, V., Grandjean, J., Rudin, M., and Wenderoth, N. (2015). Mapping the mouse brain with rs-fMRI: an optimized pipeline for functional network identification. *NeuroImage* 123, 11–21. doi: 10.1016/j.neuroimage.2015.07.090
- Zhang, R. V., Featherstone, R. E., Melynenko, O., Gifford, R., Weger, R., Liang, Y., et al. (2019). High-beta/low-gamma frequency activity reflects top-down predictive coding during a spatial working memory test. *Exp. Brain Res.* 237, 1881–1888. doi: 10.1007/s00221-019-05558-3
- Zhang, S., Hu, S., Hu, J., Wu, P. L., Chao, H. H., and Li, C. S. (2015). Barratt impulsivity and neural regulation of physiological arousal. *PLoS One* 10:e0129139. doi: 10.1371/journal.pone.0129139
- Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J. P., Chang, W.-C., Jenvey, S., et al. (2014). Long-range and local circuits for top-down modulation of visual cortex processing. *Science* 345, 660–665. doi: 10.1126/science.1254126
- Zhang, Y., Yoshida, T., Katz, D. B., and Lisman, J. E. (2012). NMDAR antagonist action in thalamus imposes delta oscillations on the hippocampus. *J. Neurophysiol.* 107, 3181–3189. doi: 10.1152/jn.00072.2012
- Zhao, F., Zhao, T., Zhou, L., Wu, Q., and Hu, X. (2008). BOLD study of stimulation-induced neural activity and resting-state connectivity in medetomidine-sedated rat. *NeuroImage* 39, 248–260. doi: 10.1016/j.neuroimage.2007.07.063
- Zhou, H., Zhang, Q., Martinez, E., Dale, J., Hu, S., Zhang, E., et al. (2018). Ketamine reduces aversion in rodent pain models by suppressing hyperactivity of the anterior cingulate cortex. *Nat. Commun.* 9:3751. doi: 10.1038/s41467-018-06295-x

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Reimann and Niendorf. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Late Positivity Does Not Meet the Criteria to be Considered a Proper Neural Correlate of Perceptual Awareness

Chiara Mazzi\*, Gaetano Mazzeo and Silvia Savazzi

*Perception and Awareness (PandA) Laboratory, Department of Neuroscience, Biomedicine and Movement Sciences, University of Verona, Verona, Italy*

## OPEN ACCESS

### Edited by:

Cyriel Pennartz,  
University of Amsterdam,  
Netherlands

### Reviewed by:

Mazyar Fallah,  
York University, Canada  
Mario Rosanova,  
University of Milan, Italy

### \*Correspondence:

Chiara Mazzi  
chiara.mazzi@univr.it

**Received:** 29 October 2019

**Accepted:** 18 May 2020

**Published:** 07 July 2020

### Citation:

Mazzi C, Mazzeo G and Savazzi S  
(2020) Late Positivity Does Not Meet  
the Criteria to be Considered a  
Proper Neural Correlate of Perceptual  
Awareness.  
*Front. Syst. Neurosci.* 14:36.  
doi: 10.3389/fnsys.2020.00036

Contrastive analysis has been widely employed in the search for the electrophysiological neural correlates of consciousness. However, despite its clear logic, it has been argued that it may not succeed in isolating neural processes solely involved in the emergence of perceptual awareness. In fact, data from contrastive analysis would be contaminated by potential confounding factors reflecting distinct, though related, processes either preceding or following the conscious perception. At present, the ERP components representing the proper correlates of perceptual awareness still remain to be identified among those correlating with awareness (i.e., Visual Awareness Negativity, VAN and Late Positivity, LP). In order to dissociate visual awareness from post-perceptual confounds specifically related to decision making, we manipulated the response criterion, which affects how a percept is translated into a decision. In particular, while performing an orientation discrimination task, participants were asked to shift their response criterion across sessions. As a consequence, the resulting modulation should concern the ERP component(s) not exclusively reflecting mechanisms regulating the subjective conscious experience itself but rather the processes accompanying it. Electrophysiological results showed that N1 and P3 were sensitive to the response criterion adopted by participants. Additionally, the more the participants shifted their response criterion, the bigger the ERP modulation was; this was consequently indicative of the critical role of these components in the decision-making processes regardless of awareness level. When considering data independently from the response criterion, the aware vs. unaware contrast showed that both VAN and LP were significant. Crucially, the LP component was also modulated by the interaction of awareness and response criterion, while VAN results to be unaffected. In agreement with previous literature, these findings provide evidence supporting the hypothesis that VAN tracks the emergence of visual awareness by encoding the conscious percept, whereas LP reflects the contribution from post-perceptual processes related to response requirements. This excludes a direct functional role of this later component in giving rise to perceptual awareness.

**Keywords:** neural correlates of consciousness (NCC), perceptual awareness, event-related potentials (ERP), EEG, contrastive analysis, visual awareness negativity (VAN), late positivity (LP), response criterion

## INTRODUCTION

Neural correlates of consciousness (hereafter called NCCs), according to an accepted operationalization, are defined as the minimal subset of neural activations sufficient to trigger a conscious experience (Chalmers, 2000; Crick and Koch, 2003; Fink, 2016). A considerable amount of EEG studies devoted to disclosing the temporal dimension of NCCs has identified two possible candidates for electrophysiological markers of perceptual conscious experience: the Visual Awareness Negativity (VAN), occurring at occipito-temporal sites in the N2 latency range, and the Late Positivity (LP), which is maximally distributed over centro-parietal sites in the P3 time window (at about 300–500 ms after the stimulus onset; Koivisto and Revonsuo, 2003, 2010). These components, though opposite in polarity, are both characterized by a greater amplitude for aware than unaware condition and for this reason are computed as difference waves of the N2 and P3 components. While VAN has been associated with perceptual awareness quite recently and attempts to localize its source are still sparse, a study published in 1971 already reported a larger P3 amplitude in response to detected near-threshold auditory compared to the missed ones (Hillyard et al., 1971). So far, P3 has been shown to be modulated by a variety of factors and observed across multiple experimental tasks. Therefore, different cognitive mechanisms have been ascribed to this component family, and its generator is still a matter of extensive debate (Polich, 2007; Volpe et al., 2007; Verleger, 2020).

The two components were principally pinpointed by means of the so-called “contrastive analysis” (Baars, 1988), which, in perceptual awareness literature, has been applied to several paradigms, such as simple visual detection or discrimination tasks, backward masking, binocular rivalry, or inattention blindness (for a review see Kim and Blake, 2005), and combined with brain imaging techniques of a different nature. According to this approach, the physical properties of near-threshold visual stimuli are kept constant across repeated trials, while conscious experience fluctuates as a result of experimental manipulation. Then, aware and unaware trials are *post hoc* sorted on the basis of participants’ subjective reports and contrasted against each other.

Critically, the contrastive approach may have led to an oversimplification. Indeed, despite its clear logic, this approach does not take into account any cognitive process systematically occurring in association with perceptual awareness. As such, the contrastive method may not succeed in segregating the solely “proper” NCCs for the reason that the contrasted conditions would not exclusively differ in terms of awareness. Results, indeed, would include related, but conceptually distinct, processes either preceding or following the actual conscious perception (Melloni et al., 2011; Aru et al., 2012; Pitts et al., 2014; Sandberg et al., 2014).

Specifically, the function of pre-conscious processes is ensuring the occurrence and/or boosting the subsequent perceptual experience. Reduced pre-stimulus alpha oscillations responsible for increased cortical excitability, a certain level of arousal, and, in general, attention-based processes represent possible examples of awareness prerequisites that can selectively

enhance relevant processing and/or inhibit irrelevant ones. Conversely, post-perceptual processes involve, among others, working memory and decision-making aspects, allowing the access and the maintenance of the perceptual information needed for a decision. This represents a critical issue since contrastive experimental designs mostly rely upon trial-by-trial subjective reports. As exemplified by the analogy of the “refrigerator door problem” reported in Pitts et al. (2014), if you are interested in identifying the minimal sufficient conditions (i.e., neural correlates of consciousness) under which the refrigerator lights up (where the light represents the conscious percept), you necessarily need to open the door (i.e., asking observers to openly rate their conscious experience). However, we cannot know what would have occurred if the door had been kept closed: would the same processes have arisen without direct access to conscious experience? To what extent can we disentangle the neural basis of conscious perception *per se* from cognitive access mechanisms? Accessibility seems, thus, to be a constitutive part (though not causally related) of perceptual awareness, further increasing the level of complexity to assess and investigate the phenomenon of perception itself.

Another point is that subjective measure protocols are systematically and unavoidably influenced by the response criterion setting (van Gaal and Fahrenfort, 2008; Irvine, 2013), as already highlighted within both signal detection theory (Macmillan and Creelman, 1990) and drift-diffusion model frameworks (White et al., 2012). Decision-related aspects represent a bridge between sensory and motor processes necessary to perform a discrimination task. The discrimination process would, indeed, be the result of subprocesses, such as a sensory process that encodes the physical stimulation and a decision process that determines the response. Within this perspective, the behavioral performance is characterized by the sensitivity (how well the participants perceive the stimulus) and the criterion (how the participants choose to respond). Since these two factors are potentially independent, the criterion can be experimentally shifted without affecting the perceptual sensitivity. Specifically, when presenting near-threshold stimuli where participants are required to report trial-by-trial if they were aware of them, an internal criterion is set; according to this criterion, a given threshold of accumulated sensory evidence is needed to prompt an “aware response” whereas below this level, an “unaware rating” is given. From a behavioral point of view, this threshold can be shifted, leading to a liberal criterion, which is characterized by a higher amount of aware responses (increased hit trials) and minimized missed detections, or to a conservative criterion, which is, instead, associated with a reduction of aware responses (decreased hit trials) and an increased exposure to false alarms (Lynn and Barrett, 2014; Kloosterman et al., 2019). Such pattern results even more exacerbated in the presence of perceptual uncertainty due to weak sensory stimuli, as in the case of contrastive analysis with stimuli at the threshold. The criterion is naturally different across participants as a function of several factors, including personality traits, cognitive styles, and stimulus expectation (Kantner and Lindsay, 2012; Bang and Rahnev, 2017). Moreover, a significant shift in criterion can be driven by the experimental design

as well (e.g., amount of catch trials, visibility of the target, task instructions or asymmetries in stimulus-response reward, etc.) since the decision strategy is flexibly tailored depending on the experimental context (Lynn and Barrett, 2014; Bang and Rahnev, 2017; Kloosterman et al., 2019). The response criterion has also been of substantial interest within parallel lines of research regarding recognition memory (Hockley, 2011) and decision making (Gold and Shadlen, 2001; Wyart and Tallon-Baudry, 2009; Kloosterman et al., 2019), where analogous simple perceptual decision tasks requiring subjective reports are commonly employed. In this respect, however, the temporal relation between the subjective perceptual awareness of the experimental stimuli and the decision completion has only poorly been explored.

So far, no converging evidence has unveiled the temporal dynamics of the groundbreaking taxonomy put forward by Aru et al. (2012): the exact temporal windows of pre-conscious and post-perceptual processes should still be clearly identified, leaving open the intense debate about the best ERP component, among those that correlate to awareness (i.e., VAN and LP), representing the proper NCCs. In addition, it remains unclear whether preceding and following processes are fully dissociated from perceptual experience. Within this fragmented framework of perspectives, we tried to turn the aforementioned weakness of the contrastive analysis to our advantage. Since this approach might not isolate the solely NCCs due to contaminations from other cognitive processes characterizing the aware condition only, we reasoned that we could intentionally modulate a possible source of contamination instead of focusing on avoiding it. In so doing, we held constant everything else apart from the response criterion, which is known to affect how a percept is converted into decision evidence, and it is thereby involved in a post-perceptual stage (Green and Swets, 1966; White and Poldrack, 2014). The resulting modulation should affect the ERP component(s) not exclusively reflecting the conscious perception itself. This manipulation can provide a clearer picture of NCCs, which can be disentangled from their consequences and shed novel light on the close relationship between electrophysiological correlates of awareness and sensory evidence accumulation leading to the decision stage.

## EXPERIMENT 1

### Materials and Methods

#### Participants

Fifty-nine right-handed participants with normal or corrected-to-normal vision were recruited for the study. After the initial threshold assessment performed in order to select the stimuli to be used in the main EEG experiment (see “Experimental Procedure” section for details), 21 participants were discarded: six failed to modulate the response criterion (conservative | liberal) as required in the task (i.e., showing overlapped psychometric functions, e.g., **Figure 1C** on the left). A total of 12 showed the percentage of the aware trials was not modulated across levels of stimulation (i.e., flat psychometric function in at least one of the two experimental conditions, e.g.,

**Figure 1C**). Three were discarded because 50% threshold value could not be determined in at least one condition since all the stimuli reached higher aware trial percentages (i.e., the entire psychometric function above the 50% level, e.g., **Figure 1C** on the right). Additionally, data from four participants were excluded from the EEG analysis due to either technical issues during recording or incomplete data acquisition. Of the remaining 34 participants, one was further excluded because he adopted the opposite response bias shift, and nine were also further excluded because the liberal-conservative difference in awareness was less than 10%. Hence, the final sample comprised 24 participants (13 females, mean age = 21.83, SD = 1.61) who were naïve as to the purpose of the study. It is worth noting that such a high participant exclusion rate is not unusual in this kind of experiment (for examples see Wilenius and Revonsuo, 2007; Salti et al., 2012; Tagliabue et al., 2016; Travis et al., 2019; Ye et al., 2019).

All the participants were recruited from the student community of the University of Verona and were paid for their participation. Written informed consent was obtained from all participants prior to study enrollment. The experimental procedure was approved by the local ethics committee and conducted in accordance with the 2013 Declaration of Helsinki.

#### Apparatus and Stimuli

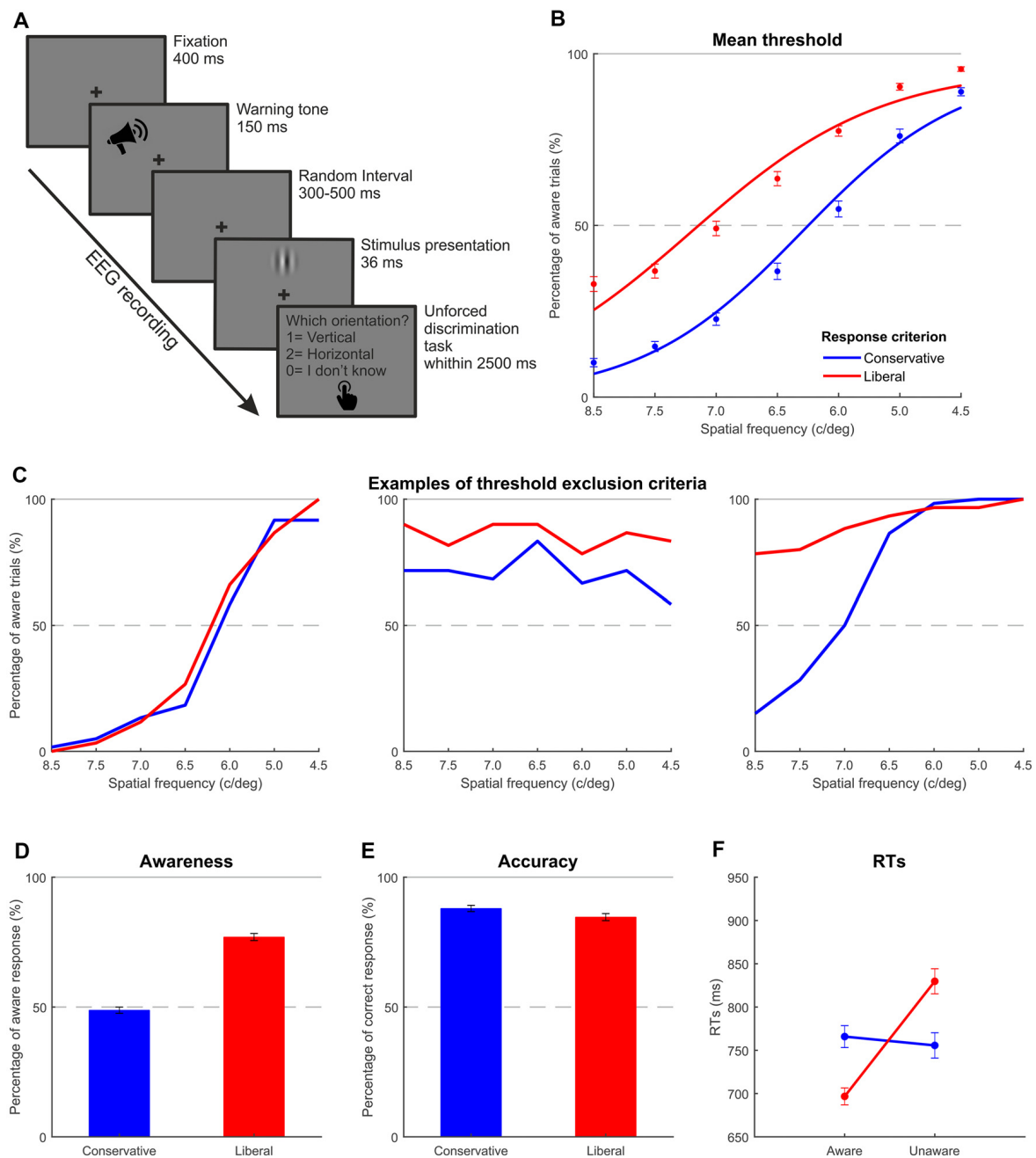
Visual stimuli consisted of vertically or horizontally oriented sinusoidal Gabor patches of about 4° of visual angle (Michelson contrast 0.50). Spatial frequencies employed in the experiment were individually calibrated by means of a threshold procedure performed prior to the main experiment. During this phase, seven different preselected vertical and horizontal spatial frequencies (4.5, 5.0, 6.0, 6.5, 7.0, 7.5, and 8.5 c/deg) were presented to the participants. In catch trials, the stimulus consisted of a patch characterized by a plaid pattern resulting from the overlap of two orthogonally oriented Gabor patches with a spatial frequency of 4.5 c/deg.

Stimuli appeared for 36 ms at the top center of the screen (11° above the fixation cross along the vertical meridian) against a gray background (3.9 cd/m<sup>2</sup>). A 17-inch CRT monitor (LG L1753HM, resolution 1,024 × 768 pixels, refresh rate 85 Hz) placed 57 cm away from the observer's head was used to display the stimuli which were previously generated by MATLAB custom code (The Mathworks Inc., Natick, MA, USA). E-prime 2.0 software (E-Prime Psychology Software Tools Inc., Pittsburgh, PA, USA) running on a Windows PC was used both to present stimuli and to collect behavioral data. Before starting the data collection, the timing of the event markers and the effective duration of the stimuli were verified by means of a photodiode and an oscilloscope.

#### Experimental Procedure

Experiment 1 was designed to manipulate participants' response criterion in a simple perceptual discrimination task by inducing a conservative or liberal bias through different verbal instructions given prior to the experiment. In the conservative session, participants were instructed to report the Gabor orientation only if they clearly perceived it, while, in the liberal session, participants were asked to report the orientation of the stimulus





**FIGURE 1 |** Experiment 1: thresholding and behavioral results. **(A)** Trial structure of the discrimination task carried out in Experiments 1 and 2. **(B)** Mean threshold functions for the two response criteria. The blue line represents the fitted function obtained adopting a conservative criterion, while the red line shows the fitted function obtained adopting a liberal criterion. Each dot indicates the mean performance across the participants included in the experimental sample corresponding to each Gabor spatial frequency. **(C)** Three possible examples of participant exclusion criteria through the thresholding phase. On the left, the two functions are overlapped indicating a failing modulation of the response criterion. In the middle, one of the two experimental conditions elicited a flat pattern and highlighted about the same awareness rate regardless of the spatial frequency. On the right, the whole psychometric function of one of the two experimental conditions is above the 50% level, making impossible to determine the threshold value. **(D,E)** Average awareness and accuracy rates as a function of response criterion. **(F)** Average reaction times per awareness and response criterion. Error bars represent standard errors of the mean (SEM).

whenever they had a minimal impression. Each participant consecutively completed the two phases of the experiment (threshold assessment and EEG experiment) on the same

day. In total, the entire experimental procedure lasted about 4 h including the set-up of the EEG cap and short breaks between blocks.

### Thresholding

First, each participant underwent a behavioral threshold assessment procedure with the aim of selecting individually calibrated spatial frequencies (one for horizontal and one for vertical Gabor patches) for the main experimental session. The threshold was estimated using the method of constant stimuli with a three-alternative unforced-choice task (i.e., the possible responses were represented by the two orientations and the additional option “do not know”; Kaernbach, 2001). According to this method, a set of vertical and a set of horizontal spatial frequencies were randomly presented to the observers, and the observers were asked to report, *via* a button press, the orientation of the presented stimulus (vertical or horizontal, respectively keys 1 and 2) or that they could not discriminate it (“do not know” answer, key 0). This allowed participants to modulate their response criterion and, at the same time, to keep the working memory load as low as possible due to the employment of a single task.

Participants were seated in a dimly lit room with their head supported by a chinrest in order to keep the distance from the monitor constant and minimize head movements. Two separate sessions of threshold assessment were consecutively carried out, and they were identical except for the induced response criterion bias (conservative | liberal). The order of the two sessions was counterbalanced across participants. Each session included six blocks of 77 trials each for a total of 462 trials. Each (vertical or horizontal) spatial frequency was presented 30 times, while there were 42 catch trials. The trial sequence started with a central fixation cross followed by a 1,000 Hz warning tone. After a random interval (ranging from 300 to 500 ms) one of the preselected stimuli was displayed, and participants were asked to make a button press response using their right hand, after which no feedback was provided (see **Figure 1A**). At the end of the threshold procedure, one horizontal frequency value and one vertical frequency value, both yielding 50% of aware responses (i.e., all responses excluding the “do not know” choices, which represent unaware condition) were visually identified for each of the two sessions (conservative | liberal) on the basis of the obtained psychometric functions. Then, the mean between the conservative and liberal horizontal values and the mean between the conservative and liberal vertical values were calculated separately (intermediate levels of spatial frequency other than those tested in the threshold assessment procedure could also be chosen) and used as spatial frequency of the target stimuli employed during the subsequent main phase of the experiment so that, in each participant, the amount of aware responses could be comparable for horizontal and vertical Gabors.

### Main Experiment

The main experiment consisted of two sessions with a short break of 15 min between them. As in the threshold assessment phase, the two sessions were identical, but a different response criterion shift was induced in the participants: in one session they were asked to adopt a conservative criterion while in the other one they were encouraged to use a liberal criterion. Again, the order of the two sessions was counterbalanced across participants. The task was the same performed in the threshold phase, where

participants were requested to discriminate the orientation of a Gabor (**Figure 1A**). However, in this context, just two kinds of near-threshold vertical or horizontal Gabor patches (whose spatial frequency values were determined during the previous phase) were presented and the EEG signal was concomitantly recorded. Each experimental session—composed of six blocks of 77 trials—yielded a total of 462 trials: 210 horizontal Gabor patches, 210 vertical Gabor patches, and 42 catch trials (10% of critical trials).

### EEG Recording and Pre-processing

A continuous EEG signal was recorded through a BrainAmp system (Brain Products GmbH, Munich, Germany—Brain Vision Recorder) using a Fast'n Easy cap (EasyCap, GmbH, Herrshing, Germany) with 59 Ag/AgCl electrodes arranged according to the 10-10 International System. Additional electrodes were placed around the eyes (left and right canthi and above and below the right eye) for monitoring the electro-oculogram activity (EOG, blinks, and saccades in particular), while electrode AFz served as the ground and the right mastoid (RM) as the online reference. Electrode impedance was constantly kept below 5 K $\Omega$  throughout the experiment. Data were recorded at a sampling rate of 1,000 Hz with a time constant of 10 s as a low cut-off and a high cut-off of 250 Hz.

The EEG signal was first imported to the EEGLAB toolbox (v2019\_0, Swartz Center for Computational Neuroscience, University of California at San Diego, Delorme and Makeig, 2004), resampled to 250 Hz, and high-pass filtered at 0.1 Hz. Power line noise (50 Hz and its harmonics) was attenuated by means of adaptive multitaper regression implemented in the CleanLine EEGLAB plugin. All scalp channels were then offline re-referenced to the mean of the two mastoids (RM-LM) prior to data segmentation into 3-s epochs (from  $-1,000$  ms to  $2,000$  ms with respect to the stimulus onset). An independent component analysis (ICA) using the extended InfoMax algorithm (Bell and Sejnowski, 1995) was performed. The 63 resulting independent components were visually inspected and removed when identified as artifactual due to blinks, eye movements, and muscle activity. Subsequently, a low-pass filter at 40 Hz was applied. The epoch window was shortened to 1,300 ms (starting from 300 ms before the stimulus), and, thereafter, baseline correction was performed based on the pre-stimulus period. Before averaging, the epochs showing activity contaminated by extreme values ( $\pm 75$   $\mu$ V) in any of the 59 electrodes were automatically rejected (on average, 13% of the epochs were discarded). Finally, stimulus-locked grand-average for aware (trials in which the participants reported to have discriminated the stimulus) and unaware (“do not know” responses) critical trials were computed separately for each response criterion bias (conservative and liberal).

### Data Analysis

#### Behavioral Data

Data were processed using MATLAB 2019a and analyzed with IBM SPSS Statistics for Windows, version 22. For each participant, trials with reaction times lower than 150 ms and higher than 1,500 ms, as well as trials with no response, were not

included in the analyses (about 3% of the data). Horizontal and vertical trials were systematically collapsed for (behavioral and EEG) data analysis. Paired-samples *t*-test was applied to compare mean percentages of aware responses for conservative criterion with those for liberal criterion. The same statistical test was used to assess whether the mean percentages of correct responses were significantly different between the two response criteria. Reaction times were submitted to a repeated-measures ANOVA with awareness (Aware | Unaware) and criterion (Conservative | Liberal) as within-subject factors.

### EEG Data

Time windows and electrodes chosen for statistical analyses were selected both according to previous literature (Koivisto and Revonsuo, 2010; Tagliabue et al., 2016) and on the basis of visual inspection of the grand-average ERPs (i.e., the electrodes with the highest peak amplitude in the component of interest when looking at the difference waves between conditions).

Grand-average ERPs were submitted to repeated-measure ANOVA with Awareness and Criterion as within-subject factors. The main effects were computed using parametric statistical routines with a statistical threshold of 0.05 implemented in EEGLAB study, while the Mass Univariate ERP Toolbox (Groppe et al., 2011) was used to further explore the interactions.

## Results

### Behavioral Results

#### Thresholding

As expected, the awareness rate increased as the spatial frequency decreased, and it was overall higher in the liberal than in the conservative condition. When a catch stimulus was presented, on average, participants chose the “do not know” response category 86.00% of the time, as they could not fittingly discriminate the orientation. Moreover, no differences were found between the two criteria (conservative catch 86.93% vs. liberal catch 85.06%;  $t_{(23)} = 0.594$ ;  $p = 0.594$ ).

At the end of the threshold procedure, which aimed at individually selecting the stimulus yielding 50% of aware responses in both the conditions (conservative and liberal) and in both the orientation (vertical and horizontal), the mean spatial frequency chosen was 6.9 c/deg for vertical stimuli and 6.2 c/deg for horizontal stimuli. Mean psychometric functions for the two response criteria, computed by collapsing data across stimulus orientation (vertical and horizontal), are shown in **Figure 1B**. For illustrative purposes only, data averaged across participants were fitted with a logistic psychometric function (lapse rate 4%) with a maximum likelihood criterion using Palamedes toolbox for Matlab<sup>1</sup> (Mazzi et al., 2017a,b).

### Main Experiment

On average, participants reported a higher awareness rate (computed taking into consideration responses “vertical” and “horizontal” of the three-alternative choices over the total amount of target stimuli) in the liberal session (76.96%) than in the conservative session (48.78%;  $t_{(23)} = -11.729$ ;  $p < 0.001$ ), indicating that participants performed the task as instructed

and could modulate their response bias (**Figure 1D**). Around 96% of the time, catch stimuli were correctly categorized as unaware with respect to the orientation, thereby indicating the reliability of performance. No differences were found between the two criteria (conservative catch 94.79% vs. liberal catch 96.90%;  $t_{(23)} = -1.529$ ;  $p = 0.140$ ). Participants were significantly more accurate in the conservative session (87.97%) than in the liberal session (84.65%; **Figure 1E**;  $t_{(23)} = 2.336$ ;  $p < 0.05$ ). Due to the low amount of catch trials representing only 10% of target stimuli and their easily detectable nature (low spatial frequency plaid pattern) resulting in a low false alarm rate, signal-detection measures such as sensitivity and criterion (Green and Swets, 1966) were not computed.

The two-way repeated-measures ANOVA conducted on reaction time data with awareness and criterion as within-subject factors revealed a main effect of awareness ( $F_{(1,23)} = 7.912$ ;  $p < 0.05$ ,  $\eta_p^2 = 0.26$ ): aware trials (731 ms) were faster than unaware trials (792 ms; **Figure 1F**), while the main effect of criterion was not found to be significant (761 ms for conservative vs. 763 ms for liberal condition;  $F_{(1,23)} = 0.039$ ;  $p = 0.845$ ,  $\eta_p^2 = 0.00$ ). Importantly, the interaction between awareness and criterion was significant ( $F_{(1,23)} = 34.159$ ;  $p < 0.001$ ,  $\eta_p^2 = 0.60$ ), highlighting that awareness differently affects the two criteria: while conservative criterion RTs for aware trials did not differ from RTs of trials classified as unaware in the same condition (766 ms conservative aware vs. 756 ms conservative unaware;  $t_{(23)} = 0.419$ ;  $p = 0.679$ ), a significant difference was reported pairwise comparing aware and unaware RTs of the liberal condition ( $t_{(23)} = -5.170$ ;  $p < 0.001$ ) with shorter RTs for aware trials (follow-up *t*-tests with Bonferroni correction applied).

### EEG Results

Across participants, the mean trial number per condition considered in the analyses after EEG pre-processing was as follows: 197 for conservative aware, 209 for conservative unaware, 304 for liberal aware, and 94 for liberal unaware.

Consistent with previous studies (e.g., Koivisto and Revonsuo, 2010), a significant main effect of Awareness was found for VAN and LP, respectively, in the 204–224 time window (electrodes O1, Oz, O2, PO7, PO8, and P7, all  $ps < 0.05$ ) and in the 328–676 time window (all electrodes except for Fp1, Fp2, AF7, AF3, AF8, F7, F5, F3, and FT7, all  $ps < 0.05$ ). **Figures 2A,B** show the difference waves computed by subtracting the unaware from the aware condition and highlights an early negative deflection (VAN) at posterior temporal electrodes (Oz in **Figure 2A**) followed by a positive enhancement at centro-parietal sites (CPz in **Figure 2B**).

A main effect of Criterion was also found to be significant. In particular, N1 showed a greater amplitude in the 152–180 time window (in electrodes Cz, C2, CP5, CP3, CP1, CPz, CP2, CP4, CP6, P3, P1, Pz, P2, P4, P6, PO3, POz, PO4, and Oz, all  $ps < 0.05$ ) when participants were requested to adopt a conservative criterion compared to the session adopting the liberal criterion (**Figure 2C**). The P3 component was also significantly modulated as a function of the criterion adopted: the liberal session elicited, indeed, a greater amplitude respect to the conservative session in the time windows 332–380 ms (electrodes C2, T8, CPz, CP2, CP4, CP6, TP8, P1, Pz, P2, P4, P6, P8, POz,

<sup>1</sup><http://www.palamedestoolbox.org>

PO4, PO8, Oz, and O2, all  $ps < 0.05$ ), 432–468 ms (electrodes CP4, P1, Pz, P2, P4, P6, PO3, POz, and PO4, all  $ps < 0.05$ ), and 512–560 ms (electrodes P3, P1, Pz, P2, P4, and POz, all  $ps < 0.05$ ; **Figure 2D**). There was no evidence of a significant amplitude difference at the N2 latency.

Critically, a significant interaction between the factors Awareness and Criterion was found for LP in the 500–800 time windows in electrodes O1, Oz, PO7, PO3, POz, PO4, PO8, P7, P5, P3, P1, Pz, P2, P4, P6, CP5, and CP4 (all  $ps < 0.05$ ), with a more pronounced LP in the conservative condition than in the liberal one (**Figure 2F**). The interaction was not significant for VAN (**Figure 2E**).

To further explore the effect of response criterion shift on ERP components, we conducted Pearson's correlation analysis between the individual behavioral difference of awareness rate (liberal minus conservative) in order to have a measure of how much each participant could shift her/his response criterion, and the difference in amplitude (liberal minus conservative) of each ERP component considered in the previous ANOVA. The aim of this analysis was to determine whether participants who could modulate to a greater extent their response criterion, also showed increased ERP differences in the different components' amplitude. Correlations were conducted on the entire sample ( $N = 33$ ), including the nine participants whose liberal-conservative difference in awareness was less than 10% and thus excluded from the ERPs analysis. The maximum peak latency of N1, N2, and P3 components was respectively identified from the grand-average of all trials (liberal and conservative sessions collapsed) in three different electrodes (PO3 for N1, PO8 for N2, and Pz for P3) and used to extract the corresponding individual amplitudes. N1 ( $r = 0.475, p < 0.01$ ) and P3 ( $r = 0.567, p < 0.001$ ) components positively correlated with criterion, while N2 did not show any correlation effect ( $r = 0.224, p = 0.21$ ; **Figure 2G**). Crucially, the same correlation analysis was performed for all the critical electrodes according to the specific topography of each component. Results were always consistent with those reported above: N1 and P3 components remained significant in all cases, while N2 never reached the significance threshold.

## EXPERIMENT 2

In Experiment 1, we manipulated the participant's response criterion by inducing either a liberal or a conservative bias in two different experimental sessions. Results showed no evidence for a difference between the two conditions in the N2 window, which is thought to trigger the process of sensory evidence accumulation (Loughnane et al., 2016). Building upon this observation, in Experiment 2, we sought to further explore dynamics underlying visual awareness by asking participants to follow their natural response criterion with the aim of comparing behavioral and electrophysiological data from the two experiments.

## Materials and Methods

### Participants

Sixteen out of 24 participants who were included in Experiment 1 (nine females, mean age = 21.88, SD = 1.50) agreed to

perform an additional experimental session and were paid for their participation.

### Stimuli, Apparatus, and Experimental Procedure

The stimuli, apparatus, and task were the same as in Experiment 1 except that participants completed one session only (for a total of 462 trials) in which they were asked to adopt their own natural response criterion (hereafter called "Own criterion session"). Importantly, the Gabor spatial frequencies employed were the same as in the previous sessions (i.e., those chosen individually by means of the threshold assessment in Experiment 1) so that comparison among sessions could be performed. The session lasted about 1.5 h including the EEG cap set-up and short breaks when needed.

### Data Analysis

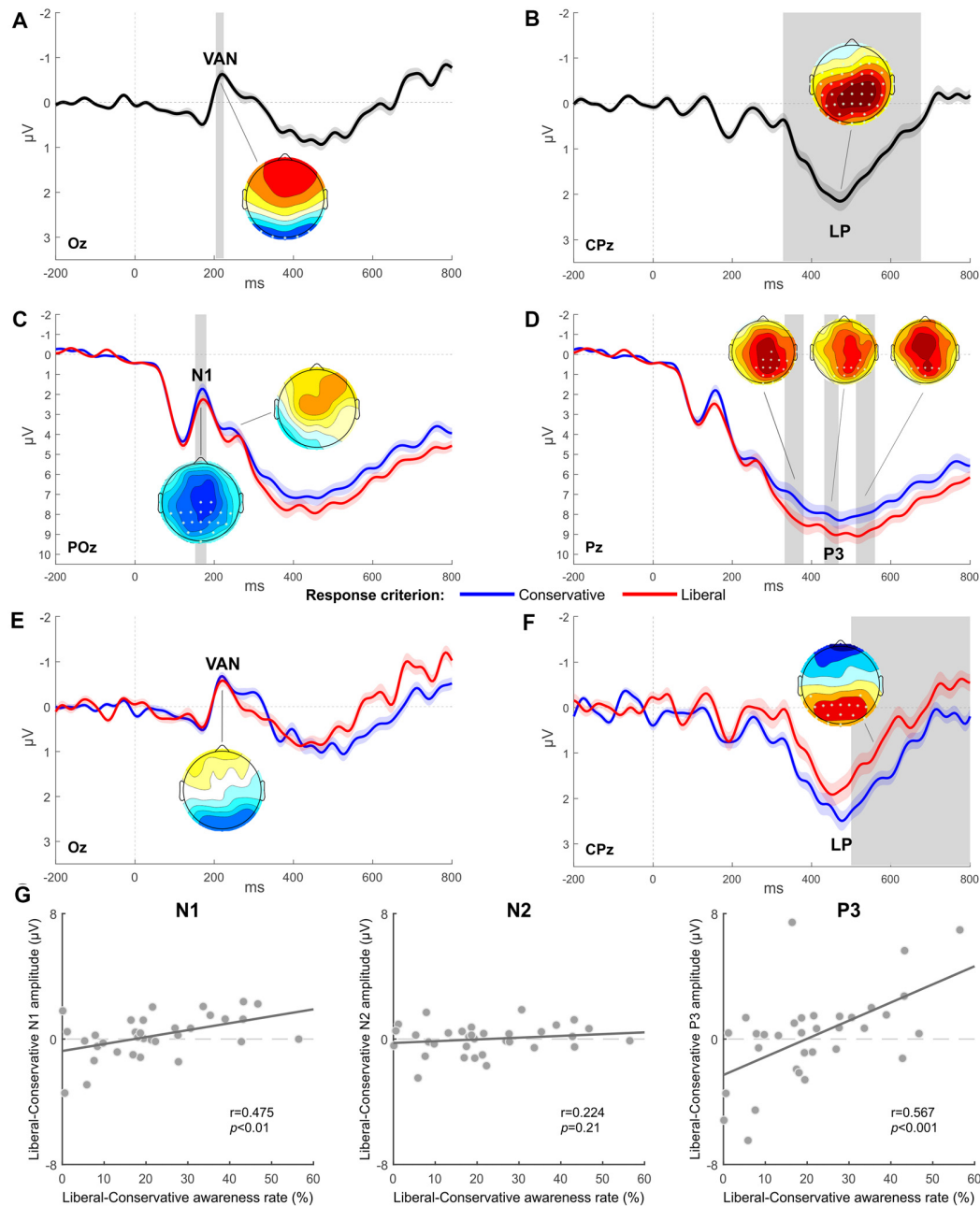
Since EEG recording settings and data pre-processing were the same as in Experiment 1, both behavioral and EEG data from the own criterion session were compared with those of the two previous sessions (conservative and liberal), considering selectively the participants who also performed Experiment 2. Specifically, awareness and accuracy percentages collected in Experiment 2 (own session) were compared to conservative and liberal sessions by means of two one-way ANOVAs at three levels (i.e., conservative, liberal, and own). Moreover, a  $2 \times 3$  repeated measures ANOVA assessed the effect of the different criteria on the awareness in terms of reaction times. As *post hoc* analysis, multiple paired *t*-test with Bonferroni correction was applied. Moreover, grand average ERP waveforms were submitted to a two-way repeated-measure ANOVA with Criterion (Conservative, Liberal, and Own) and Awareness (Aware and Unaware) as within-subject factors. The Greenhouse-Geisser correction was applied where data violated the sphericity assumption. As in Experiment 1, time windows and electrodes for the analysis were chosen based on the inspection of grand-averaged ERPs across conditions.

## Results

### Behavioral Results

For the natural criterion session, 77.92% of the 420 critical trials was rated as aware. The one-way ANOVA with Criterion as factor (with three levels: conservative, liberal, and own) disclosed a significant difference across sessions adopting different response criteria ( $F_{(2,30)} = 44.557; p < 0.001, \eta_p^2 = 0.75$ ; see **Figure 3A**). Pairwise comparisons (Bonferroni correction applied) indicated a lower awareness rate for the conservative session (44.97%) compared both to the liberal (72.95%,  $p < 0.001$ ) and own ( $p < 0.001$ ). On average, participants reported to be unaware of catch trial orientation (plaid pattern) in 95.37% of the cases, and no differences were revealed across criteria (conservative = 93.52%, liberal = 95.95%;  $F_{(1,098,16.475)} = 1.119; p = 0.312, \eta_p^2 = 0.07$ ). As with accuracy data (own = 85.27%), we found a main effect of criterion ( $F_{(1,447,21.706)} = 6.686; p < 0.05, \eta_p^2 = 0.31$ ) where the only significant pairwise comparison indicated the conservative criterion accuracy (89.46%) as higher than liberal (84.31%,  $p < 0.001$ ; **Figure 3B**).

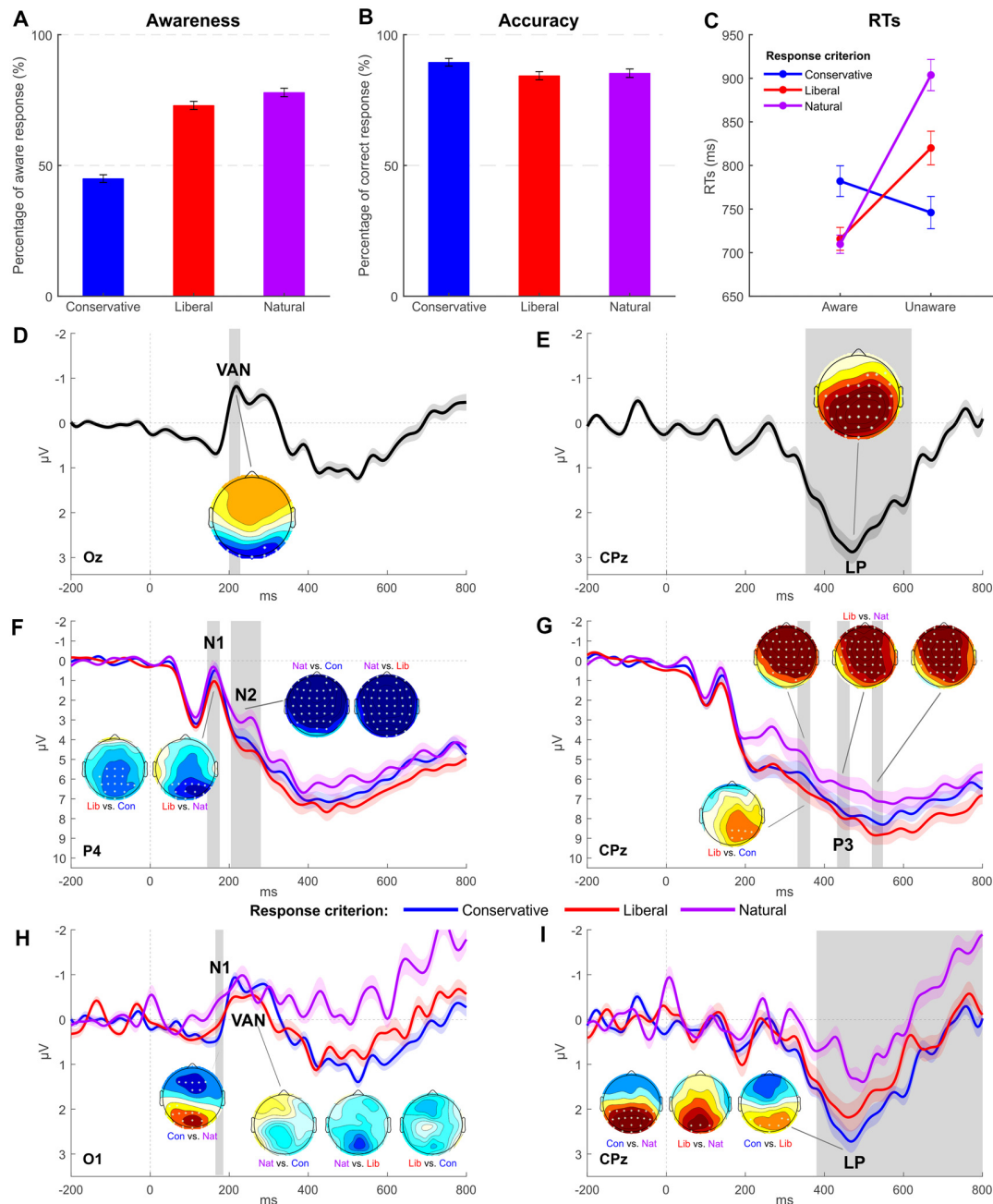




**FIGURE 2 |** Experiment 1: EEG results. (A,B) Grand-average waveform computed as the difference between the aware and unaware condition and respectively highlighting a significant Visual Awareness Negativity (VAN; electrode Oz) followed by a significant Late Positivity (LP; electrode Pz). (C,D) Grand-average of ERPs obtained for each response criterion (electrode POz for N1 and N2 components, electrode Oz for VAN, electrode CPz for P3 component). (E,F) Grand-average difference waves (aware minus unaware condition) obtained for each response criterion (electrode Oz for VAN, electrode CPz for LP). The most representative electrodes are respectively chosen for each ERP component. The shaded area of the waveforms represents SEM at each time point. Significant time windows are vertically highlighted in gray. Scalp distribution maps represent the voltage difference between conditions, with asterisks indicating statistically significant electrodes ( $p < 0.05$ ) in the given contrast and time window. The amplitude scale of topographical maps is set at  $\pm 1 \mu V$ . (G) Scatter plots showing Pearson correlations between the magnitude of response criterion shift (represented on the x-axis) and the difference in amplitude between the two response criterion conditions for respectively N1, N2, or P3 components (represented on the y-axis). Each circle represents a participant.

To further examine the relationship between awareness and criterion, RTs were analyzed with a two-way repeated measures ANOVA (Figure 3C) that revealed a significant effect of

Awareness ( $F_{(1,15)} = 21.879$ ;  $p < 0.001$ ,  $\eta_p^2 = 0.59$ ) with aware RTs (735 ms) overall shorter than unaware (823 ms). No differential RT modulations were found across criteria ( $F_{(2,30)} = 2.773$ ;



**FIGURE 3 |** Experiment 2: behavioral and EEG results. **(A,B)** Average awareness and accuracy rates across the three conditions of interest. **(C)** Average reaction times per awareness and response criterion. Error bars represent SEM. **(D,E)** Grand-average waveform computed as the difference between the aware and unaware condition and respectively highlighting a significant VAN (electrode Oz) followed by a significant LP (electrode CPz). **(F,G)** Grand-average of ERPs obtained for each response criterion (electrode P4 for N1 and N2 components, electrode CPz for P3 component). **(H,I)** Grand-average difference waves (aware minus unaware condition) obtained for each response criterion (electrode Oz for VAN, electrode CPz for LP). The most representative electrodes are respectively chosen for each ERP component. The shaded area of the waveforms represents SEM at each time point. Significant time windows are vertically highlighted in gray. Scalp distribution maps represent the voltage difference between conditions, with asterisks indicating statistically significant electrodes ( $p < 0.05$ ) in the given contrast and time window. The amplitude scale of topographical maps is set at  $\pm 1 \mu\text{V}$ .

$p = 0.078$ ,  $\eta_p^2 = 0.16$ ; conservative = 764 ms, liberal = 768 ms, own = 807 ms). The Awareness by Criterion interaction was statistically significant ( $F_{(2,30)} = 18.222$ ;  $p < 0.001$ ,  $\eta_p^2 = 0.55$ ).

*Post hoc* pairwise comparisons were then carried out with the aim of exploring this interaction. According to the Bonferroni multiple testing correction, the nine possible comparisons were

considered to be significant if  $p \leq 0.0055$  (0.05/9). In the conservative session, RTs of aware trials were not different from those of unaware trials (782 ms vs. 746 ms, respectively;  $t_{(15)} = 1.285$ ;  $p = 0.218$ ). By contrast, in liberal ( $t_{(15)} = -3.674$ ;  $p < 0.01$ ; 716 ms vs. 820 ms) and own ( $t_{(15)} = -6.349$ ;  $p < 0.001$ ; 710 ms vs. 904 ms) sessions, aware trials were shorter than unaware, thereby confirming that participants reacted differently as a function of awareness when adopting the different response criteria. Comparing the three criteria either within aware or unaware responses, the only significant contrasts were the ones comparing conservative to liberal unaware trials ( $t_{(15)} = -3.651$ ;  $p < 0.01$ ; 746 ms vs. 820 ms) and conservative to natural unaware trials ( $t_{(15)} = -4.510$ ;  $p < 0.001$ ; 746 ms vs. 904 ms), while no differences were detected between liberal and natural unaware trials ( $t_{(15)} = -2.572$ ;  $p = 0.021$ ; 820 ms vs. 904 ms) and within aware trials [Con. Aw. vs. Lib. Aw. ( $t_{(15)} = -2.741$ ;  $p = 0.015$ ; 782 ms vs. 716 ms), Con. Aw. – Nat. Aw. ( $t_{(15)} = -2.546$ ;  $p = 0.022$ ; 782 ms vs. 710 ms) and Lib. Aw. – Nat. Aw. ( $t_{(15)} = 0.263$ ;  $p = 0.796$ ; 716 ms vs. 710 ms)].

## EEG Results

After artifact rejection, the mean number of epochs used for statistical analyses was 183 for conservative aware, 223 for conservative unaware, 288 for liberal aware, 109 for liberal unaware, 314 for own aware, and 93 for own unaware condition.

A main effect of awareness was observed for VAN and LP. In particular, VAN was significant between 200 and 228 ms at posterior electrodes (O1, Oz, O2, PO7, PO4, PO8, P7, and P8, all  $ps < 0.05$ ; **Figure 3D**), while LP was significant between 352 and 620 ms at centro-parietal sites (all electrodes except for Fp1, Fp2, AF7, AF3, AF4, AF8, F7, F5, F3, F1, Fz, FT7, FC5, FC3, O2, PO8, and P8, all  $ps < 0.05$ ; **Figure 3E**).

Furthermore, the analysis revealed a main effect of criterion for components N1, N2, and P3. A significant N1 modulation was found in the 144–176 time window (**Figure 3F**), and follow-up comparisons showed a greater amplitude for own and conservative conditions as respect to the liberal condition in the following electrodes: P3, P1, Pz, P2, P4, P6, P8, PO3, POz, PO4, O1, Oz, O2, and PO8 (all  $ps < 0.05$ ). Differently from Experiment 1, the N2 component was found to be significant in the 204–280 time window in all the electrodes except for F8, FT7, FC6, FT8, T7, T8, TP7, TP8, P7, P8, and PO7 (**Figure 3F**). Follow-up paired  $t$ -tests showed a clear amplitude enhancement of the N2 component for own session trials, while conservative and liberal did not differ from each other.

P3 component was significant in several time windows: 332–364 ms (all electrodes  $ps < 0.05$  except for FC6, FT8, T7, C6, T8, TP7, CP5, CP6, TP8, P7, P5, P3, Pz, P8, PO7, PO3, POz, PO8, O1, Oz, and O2), 432–464 ms (electrodes FP1, FP2, AF7, AF3, AF4, AF8, F5, F3, F1, Fz, F2, FC3, FC1, FCz, FC2, C3, C1, Cz, C2, CP3, CP1, CPz, P3, P1, POz, and PO4), and 520–548 ms (significant electrodes FP1, FP2, AF7, AF3, F5, F1, Fz, F2, FC3, FC1, FCz, FC2, C5, C3, C1, Cz, C2, CP3, CP1, and CPz; **Figure 3G**). Follow-up  $t$ -tests revealed that the liberal criterion session elicited the greatest P3 in amplitude compared to conservative and own sessions, which were not different from each other.

The two-way interaction was significant for LP in the 380–800 time window in electrodes O1, Oz, O2, PO7, PO3, POz, PO4, PO8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, CP5, CP3, CP1, and CPz (all  $ps < 0.05$ ; **Figure 3I**). *Post hoc* two-tailed paired  $t$ -tests revealed that LP was significantly more pronounced in conservative condition than in liberal and own conditions. A significant difference was found also between liberal and own conditions.

The interaction resulted to be significant also for the N1 component in the 165–185 time window (electrodes O1, Oz, O2, PO7, PO3, POz, PO4, and P5). *Post hoc* two-tailed paired  $t$ -tests revealed that N1 was significantly greater in amplitude in the conservative condition when compared to the own condition (**Figure 3H**).

## DISCUSSION

The present article sought to dissociate the electrophysiological NCCs from potential post-perceptual confounds. Indeed, as claimed by Aru et al. (2012), contrastive analysis results would usually overestimate NCCs in report-based paradigms by including processes not strictly corresponding to the proper NCCs. Our manipulation intended to induce a response bias shift through different task instructions, in order to probe, from an electrophysiological point of view, how conscious perception interacts with decision processes to identify which ERP component cannot reflect perception *per se* since it is modulated by such post-sensory decision processes.

Behavioral results revealed that participants could successfully adopt the expected response criterion across conditions, as shown by a reduced awareness rate along with increased accuracy for the conservative criterion. This confirms that participants were less likely to report the stimulus orientation in the conservative session as compared to the liberal session. Reaction times were modulated as well, showing a different pattern as a function of awareness when adopting the conservative criterion compared to the other sessions: while in both liberal and natural sessions participants reacted faster to aware than unaware stimuli, no differences were found within the conservative session. We can speculate that the speed-up of unaware responses could be due to the adopted conservative criterion since participants were encouraged to choose the unaware option when not clearly perceiving the orientation of a stimulus at threshold. Moreover, they may have inhibited aware responses even though they were aware of the stimuli. Overall, no behavioral differences were found between liberal and natural criteria. This might depend on the spontaneous participants' tendency to try to be as accurate as possible rather than "accurately biased." This therefore relies on sensory information instead of adopting the expected decision rule (Kantner et al., 2015). From this point of view, implementing a conservative strategy may be easier than adopting a liberal criterion where participants are more likely to be inaccurate since aware answers are driven by minimal perceptual impressions. Moreover, the high perceptual uncertainty caused by the presentation of near-threshold stimuli could have also played a role in this process, making participants less prone to guess the stimuli

orientation and thus reducing the magnitude of liberal bias shift if compared to natural session. Nevertheless, the lack of a clear behavioral shift for the liberal session in comparison with the natural criterion session does not preclude that electrophysiological modulations have occurred.

As to the electrophysiological effect of response criterion manipulation regardless of the awareness level, the main findings were the decrease of N1 and, conversely, the increase of P3 amplitude for the liberal condition over (parieto-occipital electrodes for the former and over centro-posterior electrodes for the latter. Importantly, the more the participants shifted their criterion across sessions, the bigger was the difference in amplitude of the modulated components, and this suggests that neural mechanisms underlying decision making occurred in this time window. In keeping with that, P3 has been suggested to resemble the Centro-Parietal Positivity (CPP) by showing comparable latency and topography (O'Connell et al., 2012; Kelly and O'Connell, 2013; Twomey et al., 2015). This prominent component has been found to correlate with the accumulation of sensory evidence and is modulated by the timing and the accuracy of the subjective report, thereby tracing the evolving neural dynamics leading to perceptual decisions. Indeed, according to the literature on decision-making, simple perceptual decisions, such as those requested in our paradigm, are conceptualized as the result of sequential processing stages ensuring the translation of task-related physical stimulation into a behavioral decision (Sternberg, 1969). Specifically, the three steps would reflect functionally distinct processes: the sensory encoding phase, where the physical stimulus is encoded by the visual cortex; the decision formation (i.e., CPP), which is based on evidence accumulation over time until a decision threshold is reached; and, finally, the corresponding motor planning and execution. Unfortunately, to date, the temporal dimension of these steps remains still poorly explored (but see Mostert et al., 2016).

With regard to N1, it is commonly thought to be involved in early sensory and perceptual processing, as well as in the orienting of attention toward task-relevant stimuli (Luck et al., 1990; Posner and Dehaene, 1994; Hillyard et al., 1998). Given that participants did not shift their response criterion on a trial-by-trial basis but across sessions, one possibility is that the N1 decrease associated with the liberal criterion reflects an inhibitory effect of perceptual processes due to a sort of task-set “less engaging” from an attentional point of view, which is implemented to facilitate the task performance. Indeed, in the liberal session, a higher response rate is requested *a priori*, independently from accuracy. An alternative interpretative hypothesis, which does not exclude the former one, can be represented by the confidence degree: in a previous study investigating NCCs (Ye et al., 2019), N1 was modulated by the confidence level, showing an enhanced amplitude in correspondence to high-confidence trials when compared to slight-confidence trials. In this context, it seems plausible to assume a poorer confidence level in the liberal criterion session because at least a portion of the answers is based on faint perceptual impressions.

Furthermore, a significant enhanced N2 amplitude was found for the natural criterion as compared to the experimentally biased conditions (conservative and liberal sessions), and this was irrespective of the awareness level. This finding appears to be consistent with previous evidence showing that N2 might play a role in processes preceding the accumulation of evidence, specifically devoted to selecting task-relevant sensory information and enhancing their relative processing (Loughnane et al., 2016). This was a crucial step in the natural criterion session since participants were asked to make orientation judgments without following any predetermined decision rule but relying solely on the sensory evidence even though this led to uncertainty. Intriguingly, the N2 amplitude in our dataset did not correlate with the magnitude of the criterion shift adopted by the participants, suggesting that it would be perceptual in nature rather than related to decision processing.

In keeping with ERP awareness literature employing the classical contrastive design, when considering data independently from the response criterion, the typical pattern is replicated where VAN and LP correlate with visual awareness. As previously mentioned, at present, there is no unanimous consensus on the functional significance of VAN and LP (e.g., Rutiku and Bachmann, 2017), and different theories have been developed in this respect. Mainly, they can be classified as “early” or “late” theories of visual awareness depending on the temporal window considered to be critical for the emergence of awareness. Some studies have identified VAN as the earliest neural marker of visual awareness, fueling the idea that LP is more related to further post-perceptual processing rather than to conscious perception *per se* (Koivisto and Revonsuo, 2003, 2010; Wilenius-Emet et al., 2004; Pitts et al., 2014; Koivisto and Grassini, 2016; Koivisto et al., 2016; Eklund and Wiens, 2018; Mazzi et al., 2019; Ye and Lyu, 2019). Other evidence flowing into the late theories of visual awareness, usually endorsing the global workspace theory as well, have indicated that LP, instead, could represent the ERP component giving rise to conscious perception (Sergent et al., 2005; Babiloni et al., 2006; Del Cul et al., 2007; Lamy et al., 2009; Dehaene and Changeux, 2011; Salti et al., 2012; Boncompagni and Cosmelli, 2018). According to this point of view, VAN would reflect a pre-conscious stage. Besides, others pointed out a twofold contribution of VAN and LP in triggering perceptual experience (Rutiku et al., 2015, 2016; Tagliabue et al., 2016; Derda et al., 2019; Ye et al., 2019). Our approach has the added benefit of going beyond these findings, dissociating the processes that are not directly related to awareness and may act as confounds.

Interestingly, our main result is that the LP amplitude is modulated as a function of both awareness and response criterion, while VAN is unaffected by this interaction. This finding supports the idea that VAN tracks visual awareness, whereas LP does not represent a pure index of awareness, though it would be involved in post-perceptual stages of processing. This observation corroborates a growing body of results obtained employing an approach analogous to that used here, intentionally designed to rule out potential confounds associated with awareness experience. Pitts et al. (2014), by means



of a modified inattention blindness paradigm, orthogonally manipulated visual awareness and task relevance, including a condition in which participants were aware of the experimental stimulus, whereas the access of perceptual information for the subjective report was not needed. Results showed that VAN was consistently observed regardless of the task-relevance of the stimuli, whereas P3b (i.e., LP) did not show the same pattern since it was absent when the stimuli were irrelevant to the task. This effect confirmed our findings suggesting that LP is crucial for reporting requirements rather than indexing the conscious perception itself. Similar conclusions were also drawn by Koivisto et al. (2016) in a study comparing different response requirements and participants were asked to report subjective experience in the GO condition, while they had to withhold responding in the NO-GO condition. Most recently, LP has been reported to be also modulated as a function of the temporal window in which the subjective report is requested (i.e., right after the stimulus presentation or after a 2 s delay; Ye and Lyu, 2019). In this case, even though the core finding does not differ from the two previously reported studies, the effect seems to be ascribable to working memory aspects since, in the delayed condition, perceptual decisions should be kept in mind.

In accordance with our findings indicating that LP reflects post-perceptual processing, it has been also suggested that LP is related to decision making, arguing that it is in many ways similar to CPP (Koivisto et al., 2016; Tagliabue et al., 2019). Indeed, it has recently been shown that CPP is mainly modulated by the perceived stimulus clarity rather than the physical stimulus intensity (Tagliabue et al., 2019). This evidence, on one hand, upholds the undeniable relationship between subjective perceptual experience and the evidence accumulation process (Wyart and Tallon-Baudry, 2009; Gregori-Grgič et al., 2011; de Lange et al., 2011). On the other hand, however, it further corroborates the idea that LP could not be strictly involved in the very early emergence of perceptual awareness but would reflect later stages.

Taken together, the present data along with previous evidence in literature support the idea that the modulation of the LP component is not exclusively driven by perceptual processes, but it is likely the result of a combination of different processes such as confidence and evidence accumulation that need to be disentangled with further data in order to pinpoint the

corresponding neural signatures. Importantly, we can thus claim that LP is not causally involved in perceptual awareness but, instead, reflects also post-perceptual processes. For this reason, we endorse the view that there may be an earlier critical component responsible for the emergence of awareness. This could be represented by VAN, even though the account that it is an awareness prerequisite cannot be definitively ruled out at present. Moreover, we provided novel insight in disentangling perceptual awareness from decision making, especially from a temporal point of view.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Verona. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SS conceived the study. CM and SS contributed to the design of the experiment. GM carried out the data acquisition. CM and GM analyzed the data and performed the statistical analysis. CM and SS discussed the results and their interpretation. CM wrote the manuscript and SS provided critical revisions. All authors read and approved the submitted version.

## FUNDING

This research was partially supported by European Research Council (ERC) Grant 339939 “Perceptual Awareness” by Ministero dell’Istruzione, dell’Università e della Ricerca (MIUR, PRIN) 2015NA4S55\_002 grant “Visual awareness lost and found: neural and cognitive mechanisms” and by MIUR PRIN 2017TBA4KS\_002 grant “From brain dynamics to restoration of visual awareness after damage to the visual cortex.”

## REFERENCES

- Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746. doi: 10.1016/j.neubiorev.2011.12.003
- Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Available online at: <https://philpapers.org/rec/BAAACT>. Accessed July 16, 2019.
- Babiloni, C., Vecchio, F., Miriello, M., Romani, G. L., and Rossini, P. M. (2006). Visuo-spatial consciousness and parieto-occipital areas: a high-resolution EEG study. *Cereb. Cortex* 16, 37–46. doi: 10.1093/cercor/bhi082
- Bang, J. W., and Rahnev, D. (2017). Stimulus expectation alters decision criterion but not sensory signal in perceptual decision making. *Sci. Rep.* 7:17072. doi: 10.1038/s41598-017-16885-2
- Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129
- Boncompagni, G., and Cosmelli, D. (2018). Neural correlates of conscious motion perception. *Front. Hum. Neurosci.* 12:355. doi: 10.3389/fnhum.2018.00355
- Chalmers, D. (2000). “What is a neural correlate of consciousness?,” in *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, ed. T. Metzinger (Cambridge, MA: MIT press), 17–39.
- Crick, F., and Koch, C. (2003). A framework for consciousness. *Nat. Neurosci.* 6, 119–126. doi: 10.1038/nn0203-119
- Dehaene, S., and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018

- Del Cul, A., Baillet, S., and Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol.* 5:e260. doi: 10.1371/journal.pbio.0050260
- de Lange, F. P., van Gaal, S., Lamme, V. A. F., and Dehaene, S. (2011). How awareness changes the relative weights of evidence during human decision-making. *PLoS Biol.* 9:e1001203. doi: 10.1371/journal.pbio.1001203
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Derda, M., Koculak, M., Windey, B., Gociewicz, K., Wierchoń, M., Cleeremans, A., et al. (2019). The role of levels of processing in disentangling the ERP signatures of conscious visual processing. *Conscious. Cogn.* 73:102767. doi: 10.1016/j.concog.2019.102767
- Eklund, R., and Wiens, S. (2018). Visual awareness negativity is an early neural correlate of awareness: a preregistered study with two Gabor sizes. *Cogn. Affect. Behav. Neurosci.* 18, 176–188. doi: 10.3758/s13415-018-0562-z
- Fink, S. B. (2016). A deeper look at the “neural correlate of consciousness”. *Front. Psychol.* 7:1044. doi: 10.3389/fpsyg.2016.01044
- Gold, J. I., and Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* 5, 10–16. doi: 10.1016/s1364-6613(00)01567-9
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Oxford, England: John Wiley.
- Gregori-Grgič, R., Balderi, M., and de'Sperati, C. (2011). Delayed perceptual awareness in rapid perceptual decisions. *PLoS One* 6:e17079. doi: 10.1371/journal.pone.0017079
- Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x
- Hillyard, S. A., Squires, K. C., Bauer, J. W., and Lindsay, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science* 172, 1357–1360. doi: 10.1126/science.172.3990.1357
- Hillyard, S. A., Vogel, E. K., and Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1257–1270. doi: 10.1098/rstb.1998.0281
- Hockley, W. E. (2011). “Criterion changes: how flexible are recognition decision processes?” in *Constructions of Remembering and Metacognition* (London: Palgrave MacMillan UK), 155–166.
- Irvine, E. (2013). *Consciousness as a Scientific Concept: A Philosophy of Science Perspective* (Vol. 5). Dordrecht: Springer Science & Business Media.
- Kaernbach, C. (2001). Adaptive threshold estimation with unforced-choice tasks. *Percept. Psychophys.* 63, 1377–1388. doi: 10.3758/bf03194549
- Kantner, J., and Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Mem. Cognit.* 40, 1163–1177. doi: 10.3758/s13421-012-0226-0
- Kantner, J., Vettel, J. M., and Miller, M. B. (2015). Dubious decision evidence and criterion flexibility in recognition memory. *Front. Psychol.* 6:1320. doi: 10.3389/fpsyg.2015.01320
- Kelly, S. P., and O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *J. Neurosci.* 33, 19434–19441. doi: 10.1523/JNEUROSCI.3355-13.2013
- Kim, C.-Y., and Blake, R. (2005). Psychophysical magic: rendering the visible “invisible”. *Trends Cogn. Sci.* 9, 381–388. doi: 10.1016/j.tics.2005.06.012
- Kloosterman, N. A., de Gee, J. W., Werkle-Bergner, M., Lindenberger, U., Garrett, D. D., and Fahrenfort, J. J. (2019). Humans strategically shift decision bias by flexibly adjusting sensory evidence accumulation. *Elife* 8:e37321. doi: 10.7554/eLife.37321.023
- Koivisto, M., and Grassini, S. (2016). Neural processing around 200 ms after stimulus-onset correlates with subjective visual awareness. *Neuropsychologia* 84, 235–243. doi: 10.1016/j.neuropsychologia.2016.02.024
- Koivisto, M., and Revonsuo, A. (2003). An ERP study of change detection, change blindness and visual awareness. *Psychophysiology* 40, 423–429. doi: 10.1111/1469-8986.00044
- Koivisto, M., and Revonsuo, A. (2010). Event-related brain potential correlates of visual awareness. *Neurosci. Biobehav. Rev.* 34, 922–934. doi: 10.1016/j.neubiorev.2009.12.002
- Koivisto, M., Salminen-Vaparanta, N., Grassini, S., and Revonsuo, A. (2016). Subjective visual awareness emerges prior to P3. *Eur. J. Neurosci.* 43, 1601–1611. doi: 10.1111/ejn.13264
- Lamy, D., Salti, M., and Bar-Haim, Y. (2009). Neural correlates of subjective awareness and unconscious processing: an ERP study. *J. Cogn. Neurosci.* 21, 1435–1446. doi: 10.1162/jocn.2009.21064
- Loughnane, G. M., Newman, D. P., Bellgrove, M. A., Lalor, E. C., Kelly, S. P., and O'Connell, R. G. (2016). Target selection signals influence perceptual decisions by modulating the onset and rate of evidence accumulation. *Curr. Biol.* 26, 496–502. doi: 10.1016/j.cub.2015.12.049
- Luck, S. J., Heinze, H. J., Mangun, G. R., and Hillyard, S. A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. *Electroencephalogr. Clin. Neurophysiol.* 75, 528–542. doi: 10.1016/0013-4694(90)90139-b
- Lynn, S. K., and Barrett, L. F. (2014). “Utilizing” signal detection theory. *Psychol. Sci.* 25, 1663–1673. doi: 10.1177/0956797614541991
- Macmillan, N. A., and Creelman, C. D. (1990). Response bias: characteristics of detection theory, threshold theory and “nonparametric” indexes. *Psychol. Bull.* 107, 401–413. doi: 10.1037/0033-2909.107.3.401
- Mazzi, C., Mazzeo, G., and Savazzi, S. (2017a). Markers of TMS-evoked visual conscious experience in a patient with altitudinal hemianopia. *Conscious. Cogn.* 54, 143–154. doi: 10.1016/j.concog.2017.01.022
- Mazzi, C., Savazzi, S., Abrahamyan, A., and Ruzzoli, M. (2017b). Reliability of TMS phosphene threshold estimation: toward a standardized protocol. *Brain Stimul.* 10, 609–617. doi: 10.1016/j.brs.2017.01.582
- Mazzi, C., Tagliabue, C. F., Mazzeo, G., and Savazzi, S. (2019). Reliability in reporting perceptual experience: behaviour and electrophysiology in hemianopic patients. *Neuropsychologia* 128, 119–126. doi: 10.1016/j.neuropsychologia.2018.01.017
- Melloni, L., Schwiedrzik, C. M., Muller, N., Rodriguez, E., and Singer, W. (2011). Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. *J. Neurosci.* 31, 1386–1396. doi: 10.1523/JNEUROSCI.4570-10.2011
- Mostert, P., Kok, P., and de Lange, F. P. (2016). Dissociating sensory from decision processes in human perceptual decision making. *Sci. Rep.* 5:18253. doi: 10.1038/srep18253
- O'Connell, R. G., Dockree, P. M., and Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat. Neurosci.* 15, 1729–1735. doi: 10.1038/nn.3248
- Pitts, M. a., Metzler, S., and Hillyard, S. A. (2014). Isolating neural correlates of conscious perception from neural correlates of reporting one's perception. *Front. Psychol.* 5:1078. doi: 10.3389/fpsyg.2014.01078
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019
- Posner, M. I., and Dehaene, S. (1994). Attentional networks. *Trends Neurosci.* 17, 75–79. doi: 10.1016/0166-2236(94)90078-7
- Rutiku, R., Aru, J., and Bachmann, T. (2016). General markers of conscious visual perception and their timing. *Front. Hum. Neurosci.* 10:23. doi: 10.3389/fnhum.2016.00023
- Rutiku, R., and Bachmann, T. (2017). Juxtaposing the real-time unfolding of subjective experience and ERP neuromarker dynamics. *Conscious. Cogn.* 54, 3–19. doi: 10.1016/j.concog.2017.05.003
- Rutiku, R., Martin, M., Bachmann, T., and Aru, J. (2015). Does the P300 reflect conscious perception or its consequences? *Neuroscience* 298, 180–189. doi: 10.1016/j.neuroscience.2015.04.029
- Salti, M., Bar-Haim, Y., and Lamy, D. (2012). The P3 component of the ERP reflects conscious perception, not confidence. *Conscious. Cogn.* 21, 961–968. doi: 10.1016/j.concog.2012.01.012
- Sandberg, K., Andersen, L. M., and Overgaard, M. (2014). Using multivariate decoding to go beyond contrastive analyses in consciousness research. *Front. Psychol.* 5:1250. doi: 10.3389/fpsyg.2014.01250
- Sergent, C., Baillet, S., and Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nat. Neurosci.* 8, 1391–1400. doi: 10.1038/nn1549

- Sternberg, S. (1969). The discovery of processing stages: extensions of Donders' method. *Acta Psychol.* 30, 276–315. doi: 10.1016/0001-6918(69)90055-9
- Tagliabue, C. F., Mazzi, C., Bagattini, C., and Savazzi, S. (2016). Early local activity in temporal areas reflects graded content of visual perception. *Front. Psychol.* 7:572. doi: 10.3389/fpsyg.2016.00572
- Tagliabue, C. F., Veniero, D., Benwell, C. S. Y., Cecere, R., Savazzi, S., and Thut, G. (2019). The EEG signature of sensory evidence accumulation during decision formation closely tracks subjective perceptual experience. *Sci. Rep.* 9:4949. doi: 10.1038/s41598-019-41024-4
- Travis, S. L., Dux, P. E., and Mattingley, J. B. (2019). Neural correlates of goal-directed enhancement and suppression of visual stimuli in the absence of conscious perception. *Atten. Percept. Psychophys.* 81, 1346–1364. doi: 10.3758/s13414-018-1615-7
- Twomey, D. M., Murphy, P. R., Kelly, S. P., and O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *Eur. J. Neurosci.* 42, 1636–1643. doi: 10.1111/ejn.12936
- van Gaal, S., and Fahrenfort, J. J. (2008). The relationship between visual awareness, attention and report. *J. Neurosci.* 28, 5401–5412. doi: 10.1523/JNEUROSCI.1208-08.2008
- Verleger, R. (2020). Effects of relevance and response frequency on P3b amplitudes: review of findings and comparison of hypotheses about the process reflected by P3b. *Psychophysiology* 57:e13542. doi: 10.1111/psyp.13542
- Volpe, U., Mucci, A., Bucci, P., Merlotti, E., Galderisi, S., and Maj, M. (2007). The cortical generators of P3a and P3b: a LORETA study. *Brain Res. Bull.* 73, 220–230. doi: 10.1016/j.brainresbull.2007.03.003
- White, C. N., Mumford, J. A., and Poldrack, R. A. (2012). Perceptual criteria in the human brain. *J. Neurosci.* 32, 16716–16724. doi: 10.1523/JNEUROSCI.1744-12.2012
- White, C. N., and Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 385–398. doi: 10.1037/a0034851
- Wilenius, M. E., and Revonsuo, A. T. (2007). Timing of the earliest ERP correlate of visual awareness. *Psychophysiology* 44, 703–710. doi: 10.1111/j.1469-8986.2007.00546.x
- Wilenius-Emet, M., Revonsuo, A., and Ojanen, V. (2004). An electrophysiological correlate of human visual awareness. *Neurosci. Lett.* 354, 38–41. doi: 10.1016/j.neulet.2003.09.060
- Wyart, V., and Tallon-Baudry, C. (2009). How ongoing fluctuations in human visual cortex predict perceptual awareness: baseline shift versus decision bias. *J. Neurosci.* 29, 8715–8725. doi: 10.1523/JNEUROSCI.0962-09.2009
- Ye, M., and Lyu, Y. (2019). Later positivity reflects post-perceptual processes: evidence from immediate detection and delayed detection tasks. *Front. Psychol.* 10:82. doi: 10.3389/fpsyg.2019.00082
- Ye, M., Lyu, Y., Scodnick, B., and Sun, H.-J. (2019). The P3 reflects awareness and can be modulated by confidence. *Front. Neurosci.* 1:510. doi: 10.3389/fnins.2019.00510

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Mazzi, Mazzeo and Savazzi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Effects of Cortical Cooling on Activity Across Layers of the Rat Barrel Cortex

Gulshat Burkhanova<sup>1†</sup>, Kseniya Chernova<sup>1†</sup>, Roustem Khazipov<sup>1,2</sup> and Maxim Sheroziya<sup>1\*</sup>

<sup>1</sup>Laboratory of Neurobiology, Kazan Federal University, Kazan, Russia, <sup>2</sup>Aix Marseille University, INSERM, INMED, Marseille, France

## OPEN ACCESS

### Edited by:

Alain Destexhe,  
FRE3693 Unité de Neuroscience,  
Information et Complexité (UNIC),  
France

### Reviewed by:

Sylvain Crochet,  
École Polytechnique Fédérale de  
Lausanne, Switzerland  
Sylvain Chauvette,  
Laval University, Canada

### \*Correspondence:

Maxim Sheroziya  
mgsheroziya@kpfu.ru

<sup>†</sup>These authors have contributed  
equally to this work

**Received:** 30 April 2020

**Accepted:** 06 July 2020

**Published:** 04 August 2020

### Citation:

Burkhanova G, Chernova K,  
Khazipov R and Sheroziya M (2020)  
*Effects of Cortical Cooling on Activity  
Across Layers of the Rat  
Barrel Cortex.*  
*Front. Syst. Neurosci.* 14:52.  
doi: 10.3389/fnsys.2020.00052

Moderate cortical cooling is known to suppress slow oscillations and to evoke persistent cortical activity. However, the cooling-induced changes in electrical activity across cortical layers remain largely unknown. Here, we performed multi-channel local field potential (LFP) and multi-unit activity (MUA) recordings with linear silicone probes through the layers of single cortical barrel columns in urethane-anesthetized rats under normothermia (38°C) and during local cortical surface cooling (30°C). During cortically generated slow oscillations, moderate cortical cooling decreased delta wave amplitude, delta-wave occurrence, the duration of silent states, and delta wave-locked MUA synchronization. Moderate cortical cooling increased total time spent in the active state and decreased total time spent in the silent state. Cooling-evoked changes in the MUA firing rate in cortical layer 5 (L5) varied from increase to decrease across animals, and the polarity of changes in L5 MUA correlated with changes in total time spent in the active state. The decrease in temperature reduced MUA firing rates in all other cortical layers. Sensory-evoked MUA responses also decreased during cooling through all cortical layers. The cooling-dependent slowdown was detected at the fast time-scale with a decreased frequency of sensory-evoked high-frequency oscillations (HFO). Thus, moderate cortical cooling suppresses slow oscillations and desynchronizes neuronal activity through all cortical layers, and is associated with reduced firing across all cortical layers except L5, where cooling induces variable and non-consistent changes in neuronal firing, which are common features of the transition from slow-wave synchronization to desynchronized activity in the barrel cortex.

**Keywords:** cortical cooling, slow waves, cooling-evoked desynchronization, multi-unit activity, sensory-evoked potential, barrel cortex, cortical layers, urethane

## INTRODUCTION

Temperature strongly affects brain activity and metabolism, and vice versa (Kiyatkin, 2007; Wang et al., 2014). A cortical temperature below 32°C with normal core body temperature is one of the conditions for reporting brain death, and this is associated with a profound depression of EEG activity, although cortical activity *per se* may persist even at lower temperatures



(Goila and Pawar, 2009; Wang et al., 2014). Therapeutic hypothermia is used for the treatment of hypoxic-ischemic encephalopathy in neonates (Gluckman et al., 2005; Shankaran et al., 2005; Thoresen and Whitelaw, 2005; Gunn et al., 2017; Lemyre and Chau, 2018) and for the prevention of hypoxic brain injury during cardiac arrest in cardiac surgery (Hypothermia after Cardiac Arrest Study Group, 2002; Andresen et al., 2015; Sekhon et al., 2017). Local cortical hypothermia is also used to control seizures (Motamedi et al., 2013) and has been investigated as a tool for the treatment of focal brain ischemia (Miyazawa et al., 2003; van der Worp et al., 2010). Monitoring of electrical cortical activity under these conditions is instrumental for an investigation into the brain's survival and decision-making (Abend et al., 2011, 2013; Boylan et al., 2015; Frauscher et al., 2017). Hence the importance of understanding the effects of hypothermia on cortical activity under normal and pathological conditions.

Several neuronal and synaptic mechanisms of cooling have been found *in vitro*. First, a reduction in temperature to below 37°C decreases the activity of potassium leak channels and causes neuronal depolarization and an increase in input resistance (Volgushev et al., 2000; Trevelyan and Jack, 2002). Second, a moderate reduction in temperature diminishes synaptic vesicle release probability and/or desynchronizes neurotransmitter release (Katz and Miledi, 1965; Jasper et al., 1970; Hardingham and Larkman, 1998; Volgushev et al., 2004; Yang et al., 2005). Deep cooling induces strong depolarization, subsequent inactivation of fast sodium channels and depolarization block of the action potentials, and inhibits synaptic release (Jasper et al., 1970; Volgushev et al., 2000). Moderate cooling allows a change in the network balance in terms of neuronal membrane excitability (depolarization and increased input resistance) vs. synaptic excitability (reduced vesicle release probability and asynchronous release). The mechanistic similarity between moderate cortical cooling and the action of arousal neuromodulators was detected and it has been suggested that cooling can manipulate thalamocortical slow-wave activity *in vivo* (Sheroziya and Timofeev, 2015; Schwalm and Easton, 2016).

The cortical network displays bistable behavior during the synchronized states occurring during slow-wave sleep and general anesthesia, which can be detected as slow/delta waves in EEG associated with intermittent active (depolarized or UP) and silent (hyperpolarized or DOWN) states at the cellular level (Steriade et al., 1993, 2001; Lee and Dan, 2012; Neske, 2015; Poulet and Crochet, 2018). During wakefulness and rapid eye movement sleep, which are characterized by EEG desynchronization, neuromodulators such as acetylcholine partially suppress potassium channels and depolarize the membrane potential of cortical neurons, in particular, large L5 pyramidal cells (Krnjević et al., 1971; McCormick, 1992; Steriade et al., 1993; Hasselmo and McGaughy, 2004; Eggermann et al., 2014; Baker et al., 2018). In the cortex, acetylcholine is known to reduce the synaptic release and short-term synaptic depression (Gil et al., 1997; Hasselmo and McGaughy, 2004; Eggermann and Feldmeyer, 2009; Hasselmo and Sarter, 2011). Partial suppression of potassium leakage channels and a reduction of

synaptic vesicle release probability, used to simulate the impact of neuromodulators *in silico*, transformed network slow-wave activity into persistent activation (Bazhenov et al., 2002; Compte et al., 2003; Hill and Tononi, 2005). Along with these findings, moderate cortical cooling eliminated silent states during slow oscillations in lightly anesthetized mice and prevented slow-wave generation in non-anesthetized head-restrained mice (Sheroziya and Timofeev, 2015). However, the changes in neuronal activity across cortical layers and sensory-evoked responses during cooling-evoked desynchronization remain largely unknown. It was shown previously that desynchronization increases or decreases neuronal firing in L5 while superficial layers display a more uniform decrease in firing during the transition from slow-wave to persistent activity (see "Discussion" section).

Here, using multi-channel recordings with linear probes from the barrel cortex, we found that the effects of cortical cooling on electrical activity are complex and differ across cortical layers, and involve an increase in the total time spent in active states with heterogeneous changes in L5 multi-unit activity (MUA) firing rates and reduced firing rates in all other cortical layers. The strongest decline in MUA rates was observed in the layer 4 (L4) and at the putative border between layers 5 and 6, which are the main recipient layers of thalamic sensory inputs. We proposed that hypothermia changes the excitation/inhibition balance in these layers. Overall our findings are in agreement with the hypothesis that the effects of moderate cooling on cortical network activity are associated with a transition of the cortex towards a state which resembles a desynchronized cortical state.

## MATERIALS AND METHODS

### Ethical Approval

Experiments were performed following EU Directive 2010/63/EU for animal experiments, and the animal-use protocols were in line with Kazan Federal University on the use of laboratory animals (ethical approval by the Institutional Animal Care and Use Committee of Kazan State Medical University N9-2013).

### Surgery

Adult (1–2 months old) Wistar rats of both sexes were used. Surgery was performed under isoflurane anesthesia (4% for induction, 2% for maintenance), and urethane (0.7 g/kg, i.p.) was injected by the end of surgery. To reduce blood pressure and the possibility of bleeding, xylazine (10 mg/kg, i.p.) was injected at the beginning of surgery. Animals were placed on a warm thermal pad (37°C). The head was fixed in the stereotaxic frame and the skull of the animal was cleaned of skin. To reduce brain pressure and pulsation, we opened the fourth ventricle. A cranial window (~8 × 5 mm) was drilled out above the left hemisphere and the dura was gently dissected and removed above the barrel cortex (~1 × 0.5 mm). A chlorided silver wire placed in the cerebellum was used as the reference electrode. In the experiments with cortical perfusion, the craniotomy was encircled with a 2–3 mm-high dental cement wall to form a chamber.

## Cortical Cooling

To change the cortical temperature we used either perfusion with artificial cerebrospinal fluid or a U-shaped duralumin plate with a Peltier module (Sheroziya and Timofeev, 2015). We found that both methods were equally efficient. The cortex was perfused with ACSF at a speed of 6–8 ml/min. The ACSF temperature was maintained by a heater controller (Warner Instrument Corp., Hamden, CT, USA), at 38°C in control and 29–30°C during cooling. The ACSF temperature was monitored with a temperature sensor fixed inside the dental cement chamber. The duralumin plate was gently placed on the cortex and the craniotomy with the plate was covered by agar (4%) in saline. To monitor the temperature, the sensor was attached at the end of the duralumin plate by thermal conductive glue. The temperature of the plate was maintained with the Peltier element, at 38°C in control and 29–30°C during cooling. To reduce noise during the recordings the Peltier element was powered by a battery. The experiment lasted ~40 min and composed of three cooling sessions with recovery. The experiment ( $n = 22$ ) was repeated twice in each animal ( $n = 11$  rats) with a 1-h interval.

## Extracellular Recordings

LFPs and MUA were recorded from a barrel column using 16-site linear silicon probes with 100  $\mu\text{m}$  distance between the neighboring recording sites (NeuroNexus Technologies, Ann Arbor, MI, USA). The probe was inserted into the cortex to a depth of 1,600–2,000  $\mu\text{m}$  to cover all cortical layers with the recording sites. The extracellular signals from the probes were acquired using a Digital Lynx amplifier (Neuralynx, United States), digitized at 32 kHz, and saved for *post hoc* analysis. The whiskers were cut to a length of 4–5 mm and attached to a piezoelectric bending actuator (Noliac, Denmark). The principal whisker was identified by the shortest (compared to other whiskers) latency of MUA responses and was stimulated every 5 or 10 s by 200 ms long square pulse deflection in the caudal direction (amplitude of the deflection ~3 mm).

## Data Analysis and Statistics

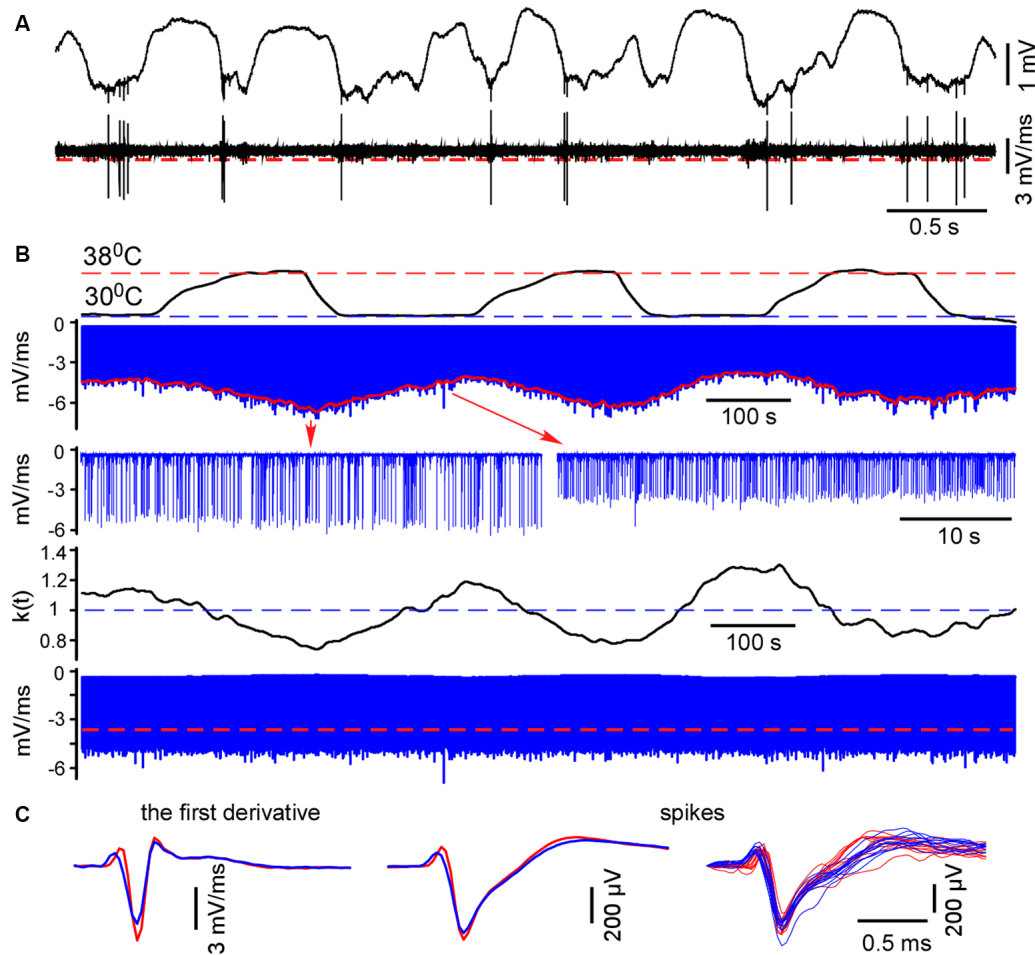
Signals were analyzed using custom-developed MATLAB scripts (MathWorks, United States). For the spontaneous activity analysis, we excluded 1 s of the recordings after the whisker stimulus. For power analysis, we calculated delta (0.5–4 Hz), theta (4–7 Hz), alpha (7–15 Hz), beta (15–30 Hz), gamma (30–90 Hz), fast gamma (90–200 Hz) and “ultrafast” gamma (200–500 Hz) power within a 10 s sliding window.

For spike detection, we differentiated the raw wide-band signal and located negative peaks (minimal slopes). Extracellular spike amplitudes and slopes interdependently decline during cooling (Henze et al., 2000; Volgushev et al., 2000). To compare MUA/spike firing rates and plotting spike histograms in control and during cooling we plotted the envelope (Figure 1) and calculated the correcting coefficient  $k(t) = \text{mean}[\text{envelope}]/\text{envelope}(t)$ . Negative peaks were multiplied by the correcting coefficient  $k(t)$ . Then we redetected negative peaks with the constant threshold = 70% of the mean

maximum value. Extracellular events with slopes below  $||-1||$  mV/ms were discarded. Detected spikes were inspected and averaged. Mean firing rates were calculated for control and hypothermic conditions. We detected spontaneous and sensory-evoked spikes independently because, as a rule, they had different amplitudes. To detect sensory-evoked spikes we consistently collected 40 ms epochs of the recordings after the stimulus and applied the same correction method as above. The MUA grouping index for spontaneous activity was calculated as the mean number of spikes within the  $\pm 10$  ms time window around a spike.

Similarly to spikes, cortical cooling decreased slow-wave amplitude (Figure 3), and using a constant threshold for slow-wave detection might mislead. To analyze changes in slow-wave activity in detail we used “noisy spikes” [all negative peaks above  $||-1||$  mV/ms after differentiation of the raw wide-band local field potential (LFP) signal] in L5 and calculated interspike intervals (ISIs; Figure 4A). For each ISI we detected the maximal L5 LFP value between the pair of spikes and L5 LFP amplitudes corresponding to each spike (Figure 4A). Plotting the indicated factors against each other, 2 clusters representing active and silent states were detected (Figure 4B). For automatic cluster discrimination, k-means clustering was used. Time spent in silent and active states was calculated as a sum of “silent” and “active” ISIs, respectively. The slow-wave frequency was calculated as the number of “silent” ISIs per second. Near the end of each “silent” ISI, the minimal value of the LFP (band-pass 0.5–6 Hz) slope was detected and used for averaging delta waves (zero time in Figure 5A). To calculate a depth profile for slow-wave onset (vertical propagation) and termination, the LFP signals from the recording channels were normalized to the maximal amplitude after the onset and before the silent state respectively (Figure 5E). Then, using threshold 0.1 for onsets and 0.5 for termination we detected threshold-crossing times for each channel and used L5 as a reference channel (zero time, Figures 5E,F). To analyze the slow-wave associated MUA synchrony we combined spikes from all the recording channels and plotted spike histograms. The spike histogram slopes were calculated as maximal values of the first derivatives for ascending phases of the MUA response. Sensory evoked potential (SEP) and slow-wave slopes were calculated as the minimal values of the first derivative (Figures 5, 6).

The whisker-evoked response in the barrel cortex of anesthetized rats (triphasic response) is composed of fast initial excitation, a prolonged inhibition and then delayed excitation (Chapin et al., 1981; Armstrong-James and George, 1988; Zhu and Connors, 1999). We analyzed the sensory-evoked MUA response corresponding to the fast initial excitation, divided into early and late components. In all experiments, the early short-latency MUA response revealed two separate peaks of maximal firing rates, located in layer 4 and at the border between layers 5 and 6 (Figure 2B). The duration of the early response was defined individually in each experiment, starting from the response onset, and, so long as firing rates maintained the same depth profile, composed of two peaks (7–12 ms, Figure 2B). The late evoked MUA response



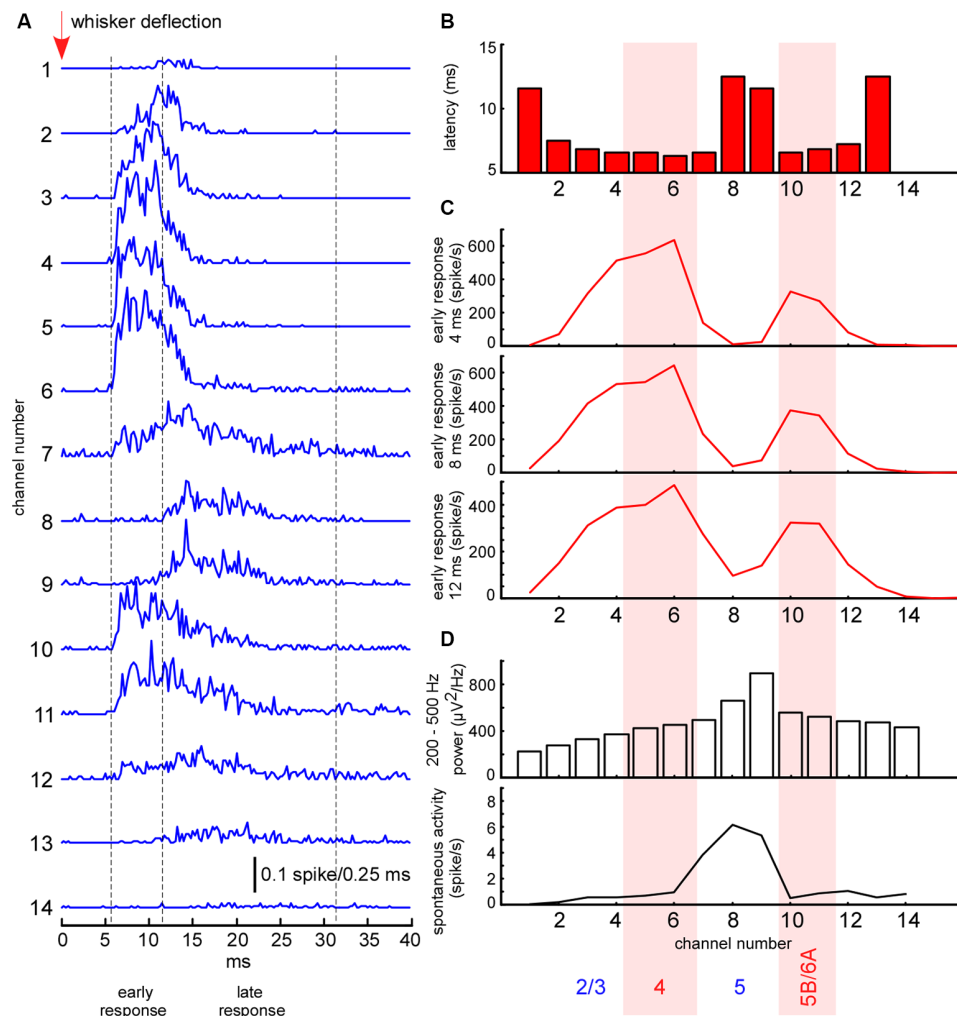
**FIGURE 1 |** Spike detection during cortical cooling. **(A)** The upper trace represents the original wide-band signal recorded in L5. The lower trace is the first derivative of the original signal. **(B)** All negative peaks in the differentiated signal were detected (the upper blue trace, L5). To detect spikes in control and during cooling the envelope was plotted (shown with red). The correcting coefficient  $k(t)$  was calculated and the negative peaks were multiplied by  $k(t)$ . Finally, spikes were detected using constant threshold = 70% of the mean maximal amplitude (the lower trace, shown with a red dashed line). **(C)** Examples of the averaged spike first derivative, averaged spikes and superposition of spikes are shown in control and during cooling (38°C—shown with red, 30°C—blue).

was calculated within the 20 ms time window after the early component. To detect evoked MUA response latency we plotted the averaged spikes histogram and detected the moment they reached 25% of the maximal value during the sensory response.

To detect cortical layers we used known electrophysiological clues (**Figure 2**). L5 thick-tufted pyramidal cells display the highest spontaneous firing rate in the rat primary somatosensory cortex (de Kock et al., 2007; Reyes-Puerta et al., 2015; Fiáth et al., 2016). As compared to other cortical layers, L5 also displays the highest “spike power” (Senzai et al., 2019). For this reason, L5 (representative channel) was detected by the peak in fast gamma power (priority) and by the highest spontaneous MUA firing rate. All types of excitatory neurons in the rat barrel cortex potentially receive direct thalamocortical inputs, but cortical neuron density and thalamocortical VPM bouton density are highest in L4 and approximately at the

border between the layer 5B and 6A (Meyer et al., 2010a,b; Constantinople and Bruno, 2013; Vinokurova et al., 2018), so it is logical to expect high MUA density during early short-latency sensory-evoked response within these layers. L4 and the putative border of 5B/6A (B5B/6A, representative channels) were detected by the frequency peaks in the early sensory-evoked response. L2/3 (representative channel) was formally detected 200–300  $\mu$ m above layer 4. In several experiments, we confirmed the validity of our layer detection methods with histology (not shown).

Statistical analysis was performed using the Matlab Statistics toolbox and OriginPro (OriginLab Corporation, Northampton, MA, USA). Data are expressed as mean  $\pm$  SEM. If the normality and equal variance tests were passed, the two-tailed  $t$ -test was used; otherwise, the nonparametric Wilcoxon signed-rank test was performed (significance level,  $p < 0.05$ ). To detect linear relations between variables we



**FIGURE 2 |** Electrophysiological clues for layer detection. **(A)** Averaged spike histograms for each channel (the last two channels not shown) represent averaged sensory-evoked multi-unit activity (MUA) response in the barrel cortex to the principal whisker stimulation. **(B)** The latency of sensory-evoked MUA response shown in panel **(A)** calculated for each recording channel. **(C)** Mean MUA firing rate of the response during the first 4, 8, and 12 ms after the response onset. **(D)** Fast gamma (200–500 Hz) power and spontaneous MUA firing rate for each recording channel.

employed robust regression and calculated Spearman and Pearson coefficients.

## RESULTS

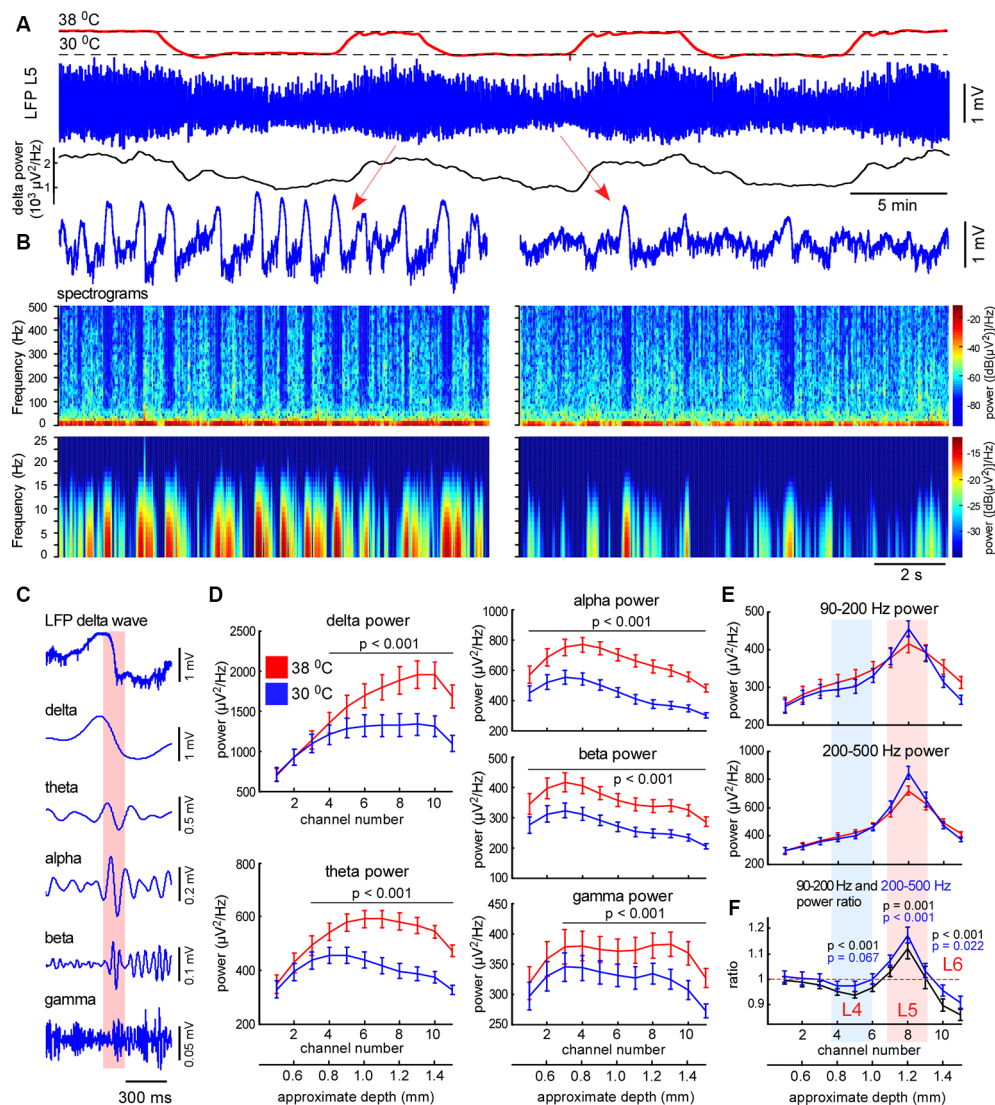
### Moderate Cortical Cooling Decreases Slow-Wave Activity and Promotes Continuity of Neuronal Network Activity

To explore cooling effects on the slow-wave activity we recorded LFP and MUA with linear probes in normothermia and during moderate cooling in the rat barrel cortex. A drop in temperature reversibly decreased slow-wave activity (**Figure 3A**), that decrease was accompanied by reduced silent state occurrence, as revealed with spectrograms (**Figure 3B**). Layer-specific power analysis showed significant decreases in the delta, theta, alpha, beta, and gamma power during cooling

through all layers (**Figure 3D**). It is important to note here, that fast transitions from silence to activity during delta wave generation strongly contribute to a broad range of frequencies (**Figure 3C**), thus, cooling-evoked suppression of slow-wave activity was followed by a drop in theta, alpha, and beta power. Cortical cooling significantly increased fast gamma power in L5, but significantly reduced it in L4 and 6 (**Figures 3E,F**).

Next, we analyzed changes in slow-wave activity in more detail (**Figure 4**). With this aim, we performed k-means clustering for ISIs and LFP activity in L5 (see “Materials and Methods” section). Slow waves are an alternation of the active (depth-negative phase in L5 LFP) and silent (depth-positive phase in L5 LFP) states. Two clusters, composed of short (and negative LFP value) and longer-duration (with positive LFP) ISIs (**Figure 4B**), represented the active and silent phases of cortical delta waves, respectively. The sum of the corresponding ISIs forming two clusters allowed us to

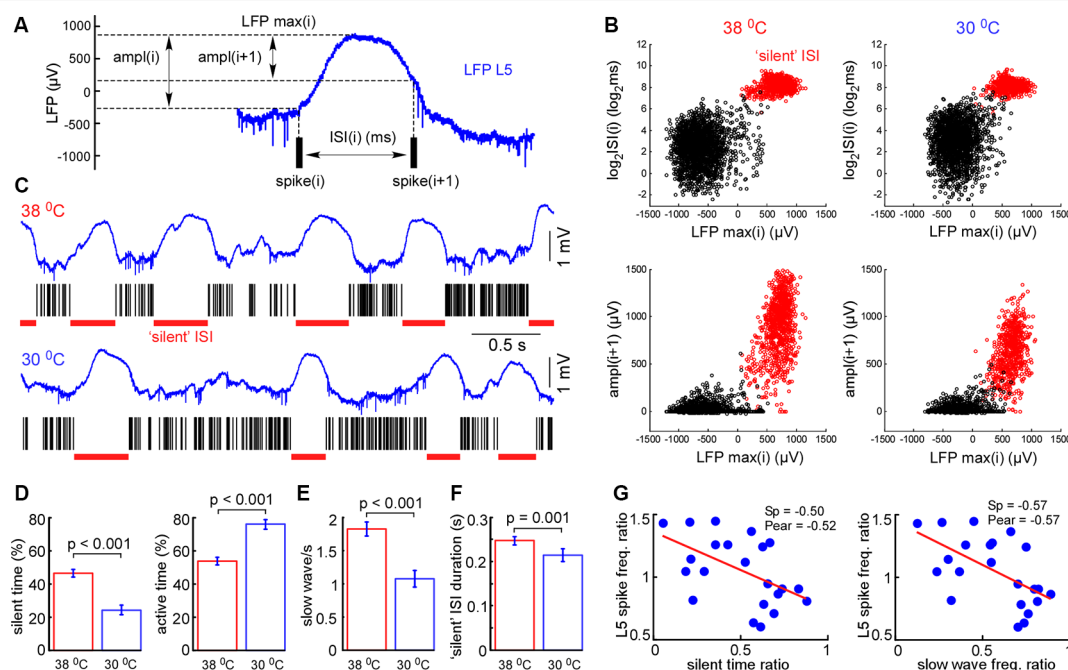




**FIGURE 3 |** Moderate cortical cooling decreased slow-wave activity and increased fast gamma power in L5. **(A)** Moderate cortical cooling reversibly decreased the frequency and amplitude of the slow waves. The red, blue, and black traces represent the cortical surface temperature (38°C—control, 30°C—cooling), the L5 local field potential (LFP) signal, and delta power, respectively. **(B)** Examples of L5 LFP slow-wave activity with spectrograms in control and during cooling. **(C)** Example of an original L5 LFP signal (one delta wave is shown) and the corresponding band-passed signals within the indicated frequency ranges. In the shown example, the transition from the silent to the active state (highlighted with pink) contributes to theta, alpha, and beta power. **(D)** Averaged delta, theta, alpha, beta and gamma power for each recording channel of the silicon probe in control and during cooling ( $n = 22$ ). Activities were averaged around the fast gamma (200–500 Hz) peak in L5 (reference channel, indicated as channel 8). **(E)** Fast gamma power (90–200 Hz and 200–500 Hz) averaged around the fast gamma peak in L5 (reference channel, indicated as channel 8). **(F)** The ratio of fast gamma power (cooling/control) was significant  $>1$  in L5 and  $<1$  in L4 and 6. The 90–200 Hz power and 200–500 Hz power ratio are shown with black and blue respectively.  $P$ -values are shown for channel 5 (L4), 8 (L5), and 10 (L6).

estimate total time spent in the active and silent states. We found that cooling significantly decreased total silent time, and increased total active time, as well as decreased slow-wave occurrence (Figures 4D,E). Cooling lightly, but significantly decreased the duration of silent states estimated as the duration of “silent” ISIs (Figure 4F). Accordingly, the strongest decrease in silent time and slow-wave occurrence was associated with the strongest increase in L5 MUA mean firing ( $n = 22$ , Figure 4G), though, the grand average for L5 MUA rate was not changed by cooling (see below). Averaged delta waves

under normothermic and hypothermic conditions are shown in Figure 5. We found that cooling significantly reduced delta wave amplitude and MUA modulation (synchronization) by delta waves (Figures 5B,G). This manifested as a decrease in peak MUA following the onset of the active state and a total decrease in firing during the 100 ms after the onset. Cooling significantly reduced delta wave amplitudes and the slopes of LFP delta waves, and MUA histograms triggered by the wave onsets (Figures 5B–D). Interestingly, we did not detect significant changes in vertical propagation of wave onsets during



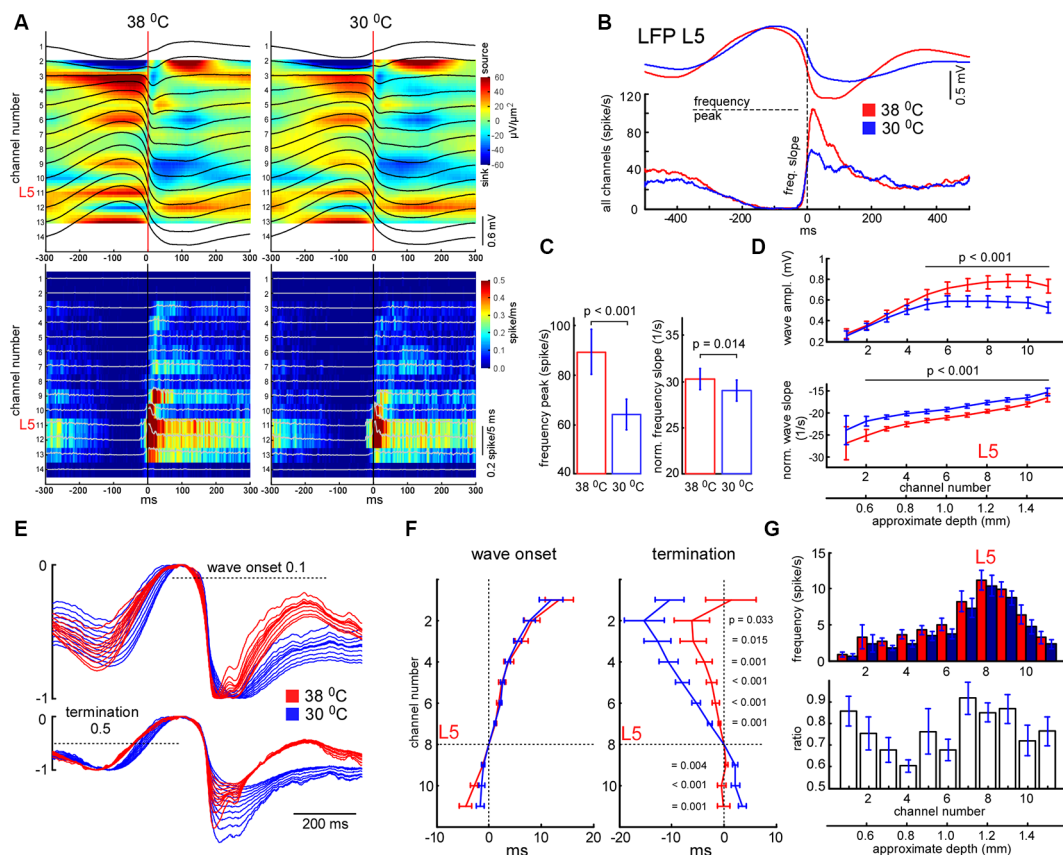
**FIGURE 4 |** Cortical cooling decreased slow-wave occurrence and total time spent in silent states. **(A)** Schematic representation of the analysis in L5. For each interspike interval (ISI) we calculated its duration, maximal LFP value between the spikes, and LFP amplitudes [amp(*i*) and amp(*i* + 1)]. **(B)** Plotting the indicated factors against each other, two clusters corresponding to active and silent states were detected. For automatic cluster discrimination, we used k-means clustering. **(C)** Examples of L5 LFP signals with discriminated ISIs in control and during cortical cooling. **(D)** The mean time spent in silent (calculated as a sum of all “silent” ISIs) and active states in control and during cortical cooling ( $n = 22$ ). **(E)** Slow-wave occurrence in control and during cooling. **(F)** Mean duration of silent states (“silent” ISI) in control and during cooling. **(G)** Ratio (cooling/control) of spike frequency in L5 plotted against silent time ratio and slow-wave occurrence ratio (one dot represents one experiment,  $n = 22$ ). The stronger silent time and slow-wave occurrence were decreased with cooling, the higher L5 spike frequency was increased. Linear fitting is shown with a red line. Spearman and Pearson correlation coefficient is indicated as Sp and Pear, respectively.

cooling (**Figure 5F**); on average, slow waves started in the deep layers and propagated upward (Sanchez-Vives and McCormick, 2000; Sakata and Harris, 2009; Chauvette et al., 2010; Reyes-Puerta et al., 2016). In contrast, cooling significantly changed the propagation profile calculated for the wave termination; the profile displayed high variability in different experiments (not shown), but the slow-wave termination in upper layers occurred earlier with cooling, so that, on average we detected propagation from superficial to deep layers. Taken together, these findings indicate that moderate cooling strongly depresses delta-wave activity through a decrease in delta-wave occurrence and promotes continuity (or desynchronization) of neuronal network activity.

## MUA Firing in Control and During Cortical Cooling

To reveal layer-specific changes in neuronal firing associated with a cooling-dependent decrease in slow-wave activity we analyzed MUA (**Figure 1**). Extracellular spike amplitudes and slopes decline during cooling and using a constant threshold for MUA detection might mislead. To detect MUA in control and hypothermia we corrected spike amplitudes and then used a constant threshold for detection (see “Materials and Methods” section) and calculated time average MUA firing rates for each cortical layer (**Figure 6**). To avoid potential variability of cortical

size between different animals we averaged MUA obtained from each recording channel of the silicon probe around the detected representative channels in L2/3, L4, L5, and at B5B/6A (**Figure 6A**). Absolute values of MUA can be different in different experiments; therefore, to reveal relative changes, we plotted ratios (cooling/control) of MUA rates (**Figure 6B**) and ratios for fast gamma power (same as in **Figure 3D**, shown with black and blue dashed lines, **Figure 6B**). We found that moderate cortical cooling significantly decreased spontaneous firing rate in superficial, L4 and B5B/6A ( $p < 0.001$ ,  $n = 22$ ), but did not change it in L5. MUA rates in L5 showed heterogeneous changes during cooling, with some animals showing increased or decreased firing (**Figure 4G**). The total firing rate (sum of all recording channels) was significantly diminished by cooling ( $89 \pm 3\%$  of the control value,  $p < 0.001$ ,  $n = 22$ ). Qualitatively, the depth profile of the fast gamma power ratio was similar to the depth profile of the MUA ratio, with the maximum in L5 and minimum in L4 and around B5B/6A (**Figure 6B**). To estimate MUA grouping (synchrony) for each cortical layer we calculated the mean number of MUA spikes within a  $\pm 10$  ms window around a spike (MUA grouping index) in control and during cortical cooling, and the ratio of cooling grouping index to control grouping index (**Figures 6C,D**). We found that the MUA grouping index was maximal in the superficial layers and L4 and had a local maximum at B5B/6A. Moderate cortical cooling



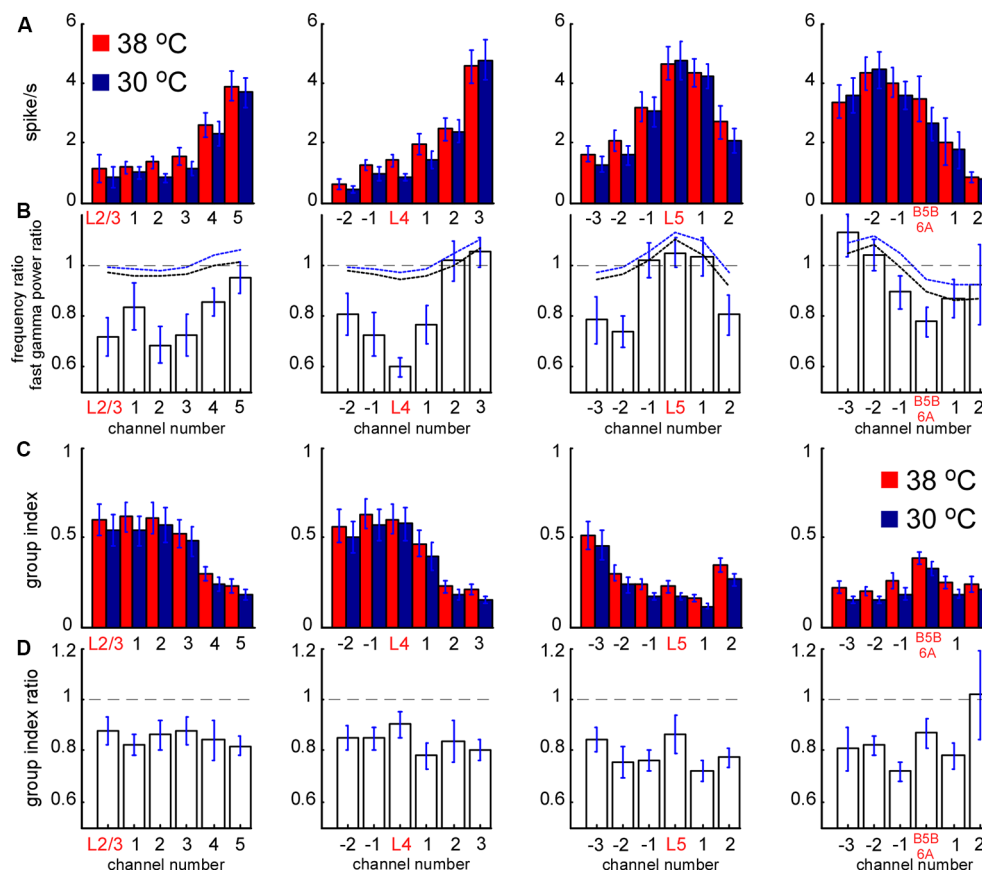
**FIGURE 5 |** Effects of cooling on slow-wave activity and associated spike synchrony. **(A)** Example of averaged LFP slow-wave with current source density plot in control and during cortical cooling (upper plots) and the corresponding MUA histograms and density plots (bottom). Activities averaged around (triggered by) the minimal value of the L5 LFP slope (zero time). **(B)** Example of averaged L5 LFP slow waves and spike histograms in control (shown with red) and during cortical cooling (blue). Here spikes were collected from all the recording channels of the silicon probe. **(C)** Cortical cooling significantly decreased the delta wave associated spike frequency peak (shown in **B**) and its slope during the transition from silent to the active state ( $n = 22$ ). **(D)** Averaged slow wave amplitudes and the wave onset slopes calculated for each recording channel in control and during cooling. Activities were averaged around L5 (channel 8). **(E)** The imposition of normalized slow-wave signals recorded with the silicon probe through the cortex during cooling. **(F)** Vertical propagation of slow waves in control and during cooling. Cooling did not change the depth profile calculated for the wave onset (threshold 0.1); on average, slow waves started in the deep layers and propagated upward. Cooling significantly changed the depth profile calculated for the wave termination (threshold 0.5); despite the pattern being different in each experiment, signals from the superficial layers crossed the threshold earlier with cooling.  $P$  values calculated for cooling-evoked increments  $<$  and  $>0$  for channels 1–7 and 9–11, respectively ( $n = 22$ ). Activities were averaged around L5 (zero time). **(G)** Mean firing rates and their ratios during the first 100 ms after the slow-wave onset ( $n = 22$ ).

diminished the MUA grouping index through all cortical layers. Thus, layer-specific analysis of MUA during cooling-evoked desynchronization revealed that the most significant decrease in firing rate is in the superficial layers, L4, and at B5B/6A.

## Cortical Sensory-Evoked Responses During Active and Silent States in Control and Cooling Conditions

We further compared cortical sensory responses evoked by the principal whisker deflection in control and cooling conditions. Since cortical sensory-evoked potentials (SEPs) depend on the phase of slow oscillation and the network state (see below) we analyzed the responses evoked during the active and silent cortical states separately. To sort responses evoked during different states we applied a threshold to the L5 LFP signal

(Figure 7A). Because L4 SEP is known to be the largest compared to the other layers, we firstly calculated its amplitude and slope. We found that the amplitude of the L4 SEP was significantly higher if the principal whisker was deflected during the silent state as compared to the active state (Figures 7B,C) and did not depend on moderate cooling. The L4 SEP onset slope, normalized by amplitude, was significantly lower if the principal whisker was deflected during the silent state (Figure 7C) and also did not depend on moderate cortical cooling. To compare latencies of the sensory-evoked response we plotted spike histograms (Figure 7F). The latencies were significantly longer if the response was evoked during the silent state independently of moderate changes in temperature (Figure 7H). The latency during active state/hypothermia was significantly longer compared to the latency during active state/normothermia ( $6.5 \pm 0.1$  vs.  $6.17 \pm 0.15$  ms



**FIGURE 6 |** Spontaneous MUA rate in control and during cortical cooling. **(A)** MUA firing rates from different channels of the silicon probe averaged around cortical layers (representative channels) in control (shown with red) and during cortical cooling (blue color). **(B)** Bars represent the MUA frequency ratio (cooling/control) averaged around cortical layers. Dashed black and blue lines represent averaged fast gamma power of the 90–200 Hz and 200–500 Hz ratio, respectively ( $n = 22$ ). **(C)** MUA grouping index calculated for different cortical layers in control and during cooling. The grouping index was calculated as a mean number of spikes in a  $\pm 10$  ms time window around a spike. **(D)** MUA grouping index ratio averaged around cortical layers in control and during cooling.

respectively;  $p = 0.011$ ,  $n = 22$ ). We conclude that the amplitude, slope, and latency of the sensory-evoked response were much more dependent on the network state than on a moderate change in temperature.

It was previously shown that the sensory-evoked high-frequency oscillations (HFO) in the barrel cortex were eliminated with deep cooling that suppresses neuronal activity (Jones and Barth, 1999; Staba et al., 2003). To reveal moderate cooling effects on HFO we averaged LFP signals applied during active or silent states (Figure 7D) and band-passed them within the 300–16,000 Hz range (Figure 7E). We found that HFO was sensitive to moderate cortical cooling. A moderate drop in temperature significantly decreased HFO frequency (Figure 7I) but increased its peak power (Figures 7E,G,I). Also, HFO were generated later if the stimulus was applied during the silent state (Figure 7E).

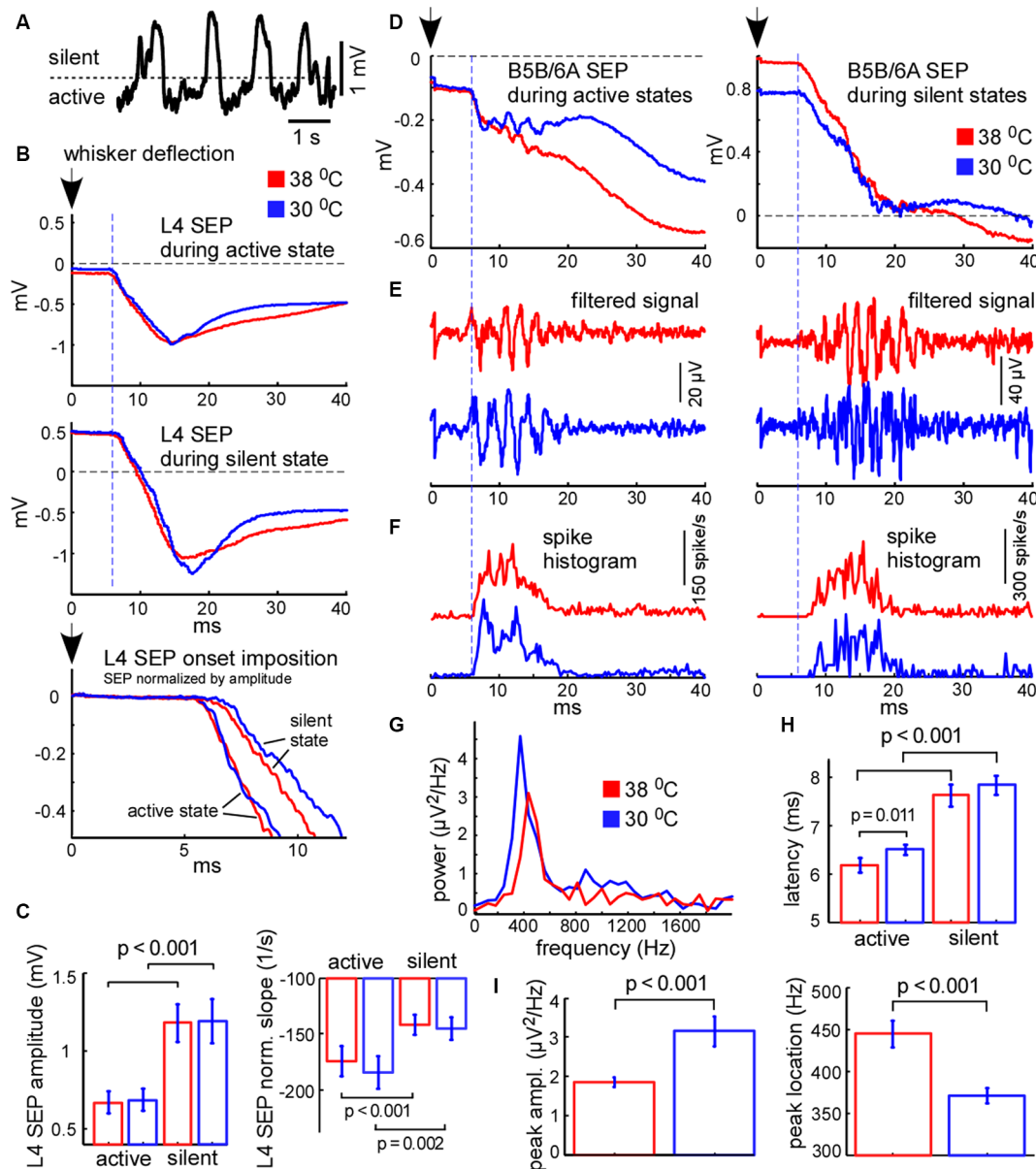
It is known that the amplitude of the cortical response is higher during the silent (hyperpolarized) state as compared to the active (depolarized) state (Sachdev et al., 2004; Crochet et al., 2005), because stimulation during the silent state synchronously depolarizes and activates all the cells in the target area. The

latency of cortical response during the silent state also appears to be longer because membrane potential transition from a hyperpolarized state to the threshold of spike generation takes more time (Shu et al., 2003a; Rosanova and Timofeev, 2005). Therefore, the relative response latencies and amplitudes obtained from extracellular recordings allowed us to make a judgment about the instantaneous membrane potential at the moment of sensory stimulation. Taken together, our results support the idea that cortical cells are depolarized during cooling-evoked active states which we detected with LFP signal and, therefore, active states under moderate hypothermia are mechanistically close to the true active states under normal physiological conditions.

## Sensory-Evoked MUA Firing During Cortical Cooling

We further analyzed the effect of cooling on MUA evoked by principal whisker deflection. MUA response was divided into an early (7–12 ms after the onset) and late component (the following 20 ms). As above, we plotted MUA rates for early (Figure 8B) and late components (Figure 8E) around the detected representative

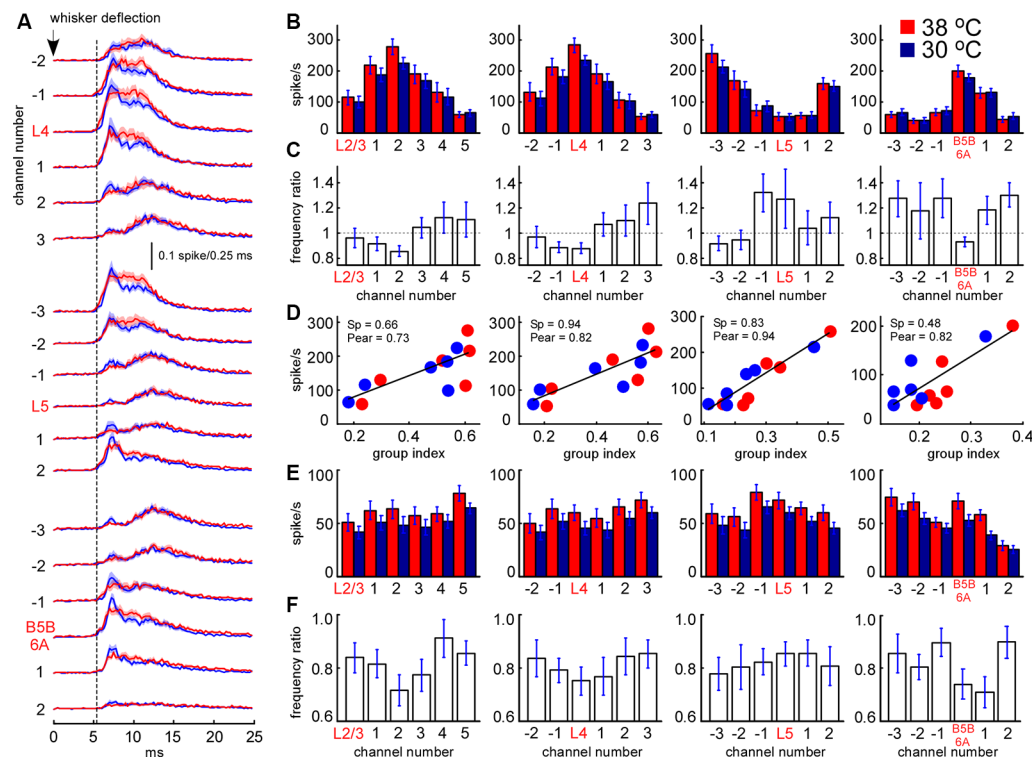




**FIGURE 7 |** Cortical sensory-evoked response in control and during cooling: active state vs. silent state. **(A)** Example of slow-wave activity in L5. The dashed line represents the threshold (0.4 standard deviations of the LFP signal), used for sensory stimuli sorting. **(B)** Examples of averaged L4 sensory-evoked potentials (SEPs) evoked during the active and silent states in control (shown with red) and cooling conditions (blue). The black arrow (zero time) represents the moment of principal whisker deflection. The lower trace represents a superposition of sensory evoked potential (SEPs) normalized by amplitude. **(C)** L4 SEP amplitude was significantly higher if evoked during the silent state compared to the active state independently of temperature ( $n = 22$ ). The L4 SEP slope normalized by amplitude was significantly higher if evoked during active states. **(D)** Examples of averaged B5B/6A SEPs evoked during active states (left panel) and silent states (right panel) in control and cooling conditions. The original wide-band signal was averaged. Note high-frequency oscillation. **(E)** The same activities as in **(D)** band-passed within the 300–16,000 Hz range. Note sensory-evoked high-frequency oscillation. **(F)** Spike histograms (spikes collected from all recording channels) in control and during cooling. Note the higher latency of the sensory-evoked MUA response during the silent state compared to the active state. **(G)** The power spectrum plot calculated for the activities shown in **(E)** for active states. The power spectrum plot revealed a high-frequency oscillation peak. Note, that cortical cooling increased the peak amplitude and decreased the frequency of high-frequency oscillation. **(H)** The latencies of sensory-evoked MUA responses were significantly higher if evoked from silent states compared to active states. **(I)** Cortical cooling significantly increased the power of sensory-evoked high-frequency oscillation (peak amplitude) and decreased its frequency (peak location).

channels in L2/3, L4, L5, and at B5B/6A. We analyzed MUA responses evoked during active states (responses evoked during silent states revealed rather similar cooling-evoked changes, not

shown). Relative changes in sensory-evoked MUA rates were rather similar to the changes in spontaneous MUA. Cooling significantly decreased early sensory-evoked MUA rates in



**FIGURE 8 |** Sensory-evoked MUA rate in control and during cortical cooling. **(A)** Grand average peristimulus spike histograms (evoked during active states) in control (red) and during cortical cooling (blue). Note the increased peak firing with cooling. **(B)** Early sensory-evoked MUA firing rate (evoked during active states) from different recording channels of the silicon probe in control (shown with red) and during cortical cooling (blue). **(C)** Early sensory-evoked MUA frequency ratio (cooling/control) averaged around cortical layers ( $n = 22$ ). **(D)** Early sensory-evoked MUA firing rate plotted against the MUA grouping index shown in **Figure 7C**. Cortical layers with a high grouping index (spontaneous activity) responded to the whisker stimulation with higher firing rates. Linear fitting is shown with black lines. Spearman and Pearson correlation coefficient is indicated as Sp and Pear respectively. **(E)** Late sensory-evoked MUA firing rate (evoked during active states) from different recording channels of the silicon probe averaged around cortical layers (representative channels) in control and during cortical cooling. **(F)** Late sensory-evoked MUA frequency ratio (cooling/control) averaged around the detected cortical layers.

L4 and B5B/6A ( $p < 0.001$ ,  $n = 22$ ), and evoked heterogeneous changes in L5 (**Figure 8C**). We also found that the grouping index calculated for spontaneous MUA positively correlated with the early evoked MUA firing rate (**Figure 8D**): the layers with a high grouping index displayed higher firing rates during the early sensory-evoked response. In contrast to the early MUA response, cooling significantly decreased the late MUA rates through all cortical layers (**Figure 8F**). The depth profiles of relative MUA changes, similar to spontaneous activities, displayed a maximum in L5 and minimum in L4 and around B5B/6A (**Figures 8C,F**). Taken together, despite moderate cortical cooling increasing the duration of cortical active states in anesthetized rats, it also strongly inhibited sensory-evoked MUA response in L4 and at B5B/6A and, in particular, the late component of sensory-evoked MUA response through all cortical layers.

## DISCUSSION

The main finding of the present study is that the effects of cortical cooling on electrical activity are complex and differ across cortical layers and that these changes closely match

a transition from slow-wave activity towards desynchronized network activity. Cortical hypothermia promoted a transition from slow-wave activity to more continuous persistent activity, and this was associated with a decrease in neuronal firing in all cortical layers except L5. We found that the MUA grouping index decreased through all cortical layers with cooling. While cortical cooling increased total time spent in active states, sensory-evoked responses were suppressed during cooling through all layers. With cooling, spontaneous and sensory-evoked MUA rates were more strongly reduced in L4 and B5B/6A, which are the main recipient layers of the thalamic sensory inputs. The cooling-dependent slowdown was detected at the fast time-scale with a decreased frequency of the sensory-evoked HFO.

## Hypothermic State vs. Desynchronized State

We suppose that partial inactivation of potassium channels with moderate/light cooling is the main reason for changes in slow-wave network dynamics, leading to changes in firing rates as a network effect. The majority of open membrane channels at rest, which maintain the resting membrane potential and

determines the neuronal input resistance, are potassium leak channels (Hodgkin and Huxley, 1952; Lesage, 2003). It had been shown that one of the candidates for potassium leak channels, the two-pore-domain potassium [K(2P)] channel, is temperature-sensitive (Talley et al., 2001; Kang et al., 2005; Schneider et al., 2014). Moderate cooling evoked K(2P)-dependent neuronal activation in hippocampal cell cultures (de la Peña et al., 2012). Mild/moderate ( $\sim 5^{\circ}\text{C}$  below normothermia) local cooling in songbirds slows down the song tempo (Long and Fee, 2008; Zhang et al., 2017), but did not prevent the behavior itself in contrast to deep cortical cooling, used for local cortical inactivation (Lomber et al., 1994).

In sharp contrast with intracellular recordings in cats where slow/delta waves is a feature of slow-wave sleep (Steriade et al., 2001; Timofeev et al., 2001), transitions between synchronization and desynchronization have been reported in the rodent barrel cortex also during quiet and alert-whisking wake states. In the absence of whisker movements, large-amplitude membrane potential fluctuations dominate in L2/3 pyramidal cells and interneurons, whereas during active whisking, the large-amplitude membrane potential fluctuations were replaced by smaller amplitude and continuous fluctuations (Crochet and Petersen, 2006; Gentet et al., 2010; Reimer et al., 2014). In the rat barrel cortex, desynchronization suppressed spontaneous spiking activity of L2/3 principal cells but caused enhanced firing in a subset of L5 neurons (de Kock and Sakmann, 2009). Application of cholinergic agonists to the thalamus in urethane-anesthetized mice also caused a desynchronized state in the barrel cortex and suppression of L2/3 principal neuron firing (Hirata and Castro-Alamancos, 2010), but in contrast to other desynchronization models (Gentet et al., 2010; Sakata and Harris, 2012), enhanced fast-spiking neuronal firing. In the rat visual cortex, basal forebrain stimulation decreased MUA firing rates in the superficial layers and lightly increased firing in other layers (Goard and Dan, 2009). In the rat auditory cortex, spontaneous desynchronization and desynchronization evoked by whisking in unanesthetized rats, and desynchronization evoked by electrical stimulation of the pedunculopontine tegmental nucleus in anesthetized rats, suppressed spontaneous spiking activity in the superficial layers but caused heterogeneous changes in deep layer firing. Typically, large tufted L5 pyramidal cells increased the firing rate (Sakata and Harris, 2012). The mean frequency of spontaneous excitatory postsynaptic potentials (EPSPs) is lower in superficial cortical layers in anesthetized rats (compared to deep layers), but the EPSP amplitude is higher, so, a single EPSP can evoke more than one action potential in the superficial layers (Zhu and Connors, 1999). Spontaneous superficial-layer spiking activity during active states is sparse (Brecht et al., 2003; Kerr et al., 2007; Poulet and Petersen, 2008; Sakata and Harris, 2009). In our experiments, cooling-evoked desynchronization led to largely consistent changes in spontaneous MUA firing rates across cortical layers including variable changes in L5 MUA frequency which might reflect a variable response of different L5 neurons and reduced firing in other cortical layers as it occurs during slow-wave—persistent network activity transitions. Analysis of the sensory-evoked response latency, amplitude and slope supported the idea that

cortical cells were depolarized during cooling-evoked active states (Shu et al., 2003a; Crochet et al., 2005; Rosanova and Timofeev, 2005).

## MUA Detection and Cooling-Dependent Slowdown

To compare MUA frequency in control and during cortical cooling we corrected spike amplitudes for their decrease upon cooling. At the first approximation, the extracellular spike amplitude is proportional to the first derivative of the intracellular potential, diminished with cooling, because spikes become wider (Henze et al., 2000; Volgushev et al., 2000; Girardin and Martin, 2009). Using a constant amplitude threshold for MUA detection might lead to artificially decreased/increased firing rates with lower/higher temperatures, respectively. We discarded small-amplitude extracellular events from our analysis, but we cannot exclude that some portion of the detected MUA spikes were not somatic spikes, in particular, in L4 and B5B/6A. It has been reported that up to 20% percent of the early sensory-evoked MUA response in L4 and at the border between layers 5 and 6 are thalamocortical axonal spikes (Vinokurova et al., 2018). The early sensory-evoked MUA response, insensitive to the glutamate receptor antagonists, had been also reported in rat pups (Minlebaev et al., 2007).

We found that cortical cooling decreased the frequency of HFO and delta wave slopes, but did not change SEP slopes. Cooling did not change the depth profile calculated for LFP delta wave onset, but significantly changed the depth profile of LFP delta wave cessation. Interneurons are probably less sensitive to cooling compared to principal cells (Vizi and Sperl gh, 1999; Motamed et al., 2013). Therefore, fast network dynamics, based on neuron-interneuron interactions, might be temperature-sensitive. Cortical feed-forward inhibition plays a central role in the generation of cortical gamma oscillations (Bartos et al., 2007; Whittington et al., 2011; Khazipov et al., 2013) that could explain the significant reduction in HFO frequency with cooling.

## Evoked and Spontaneous MUA During Cooling-Evoked Desynchronization

We found that the polarity of changes in spontaneous L5 MUA rates correlated with changes in the total time spent in the active state (negative correlation with total silent time) along with a decrease in slow-wave occurrence during cooling-evoked desynchronization. Layer 5 neurons, the largest neurons in the cortex (Feldman, 1984), play the central role in triggering and propagation of cortical active states (Sanchez-Vives and McCormick, 2000; Chauvette et al., 2010; Wester and Contreras, 2012; Beltramo et al., 2013). Since L5 neurons were able to generate prolonged active states during decreased activity in other layers, they probably also play a central role in the maintenance of persistent cortical activity (Shu et al., 2003b).

A decrease in slow-wave occurrence was associated with a decrease in the MUA grouping index through all cortical layers. During slow oscillations, neuronal membrane potential rapidly changes between the hyperpolarized and depolarized states with short time delays in neighboring neurons

(Steriade et al., 1993, 2001; Chauvette et al., 2010; Sheroziya and Timofeev, 2014); this strongly increases the probability of firing together. The MUA grouping index (mean number of MUA spikes within the  $\pm 10$  ms window around a spike) mainly provides an estimation of neuronal synchrony near the electrode (and therefore is proportional to local neuron density) with a possible contribution of intrinsically bursting cells and interneurons, whose ISIs can be less than 10 ms. The depth profile of mean MUA rate during the early sensory-evoked response also depends on local neuron density (and thalamic sensory inputs) explaining the similarity with the depth profile of the grouping index (**Figure 8D**).

The reduction in MUA rates with cooling might depend on intrinsic neuronal properties and slow-wave occurrence. L2/3, L4, and L5B in the rodent somatosensory cortex contain intrinsically bursting neurons (Chagnac-Amitai et al., 1990; Schubert et al., 2001; Staiger et al., 2004; de Kock and Sakmann, 2008; Prönnke et al., 2020). First, intrinsically generated high firing rates require fast neuronal membrane re-polarization during the generation of the action potential. With cooling, the voltage-gated potassium current has a higher activation threshold and lower amplitude (Volgushev et al., 2000), which might decrease the frequency of spikes within bursts. Second, cortical neurons intrinsically generate bursts of spikes (or higher firing rates) right after fast depolarization from a hyperpolarized membrane potential with subsequent adaptation (Connors et al., 1982; Chagnac-Amitai et al., 1990; Connors and Gutnick, 1990; Llinás et al., 1991; Nuñez et al., 1993; Gray and McCormick, 1996; Brumberg et al., 2000; Contreras, 2004; Staiger et al., 2004; Hedrick and Waters, 2012). Thus, intrinsically bursting neurons can potentially decrease mean firing rates after the transition to desynchronization. Third, the relative fraction of interneurons across all cortical layers is around 14% (Sirota et al., 2008; Sakata and Harris, 2009; Reyes-Puerta et al., 2015), however, interneuron density was reported to be maximal in L2, L4, and L5 in the rat barrel cortex and the interneuron/principal neuron ratio in these layers is 0.19–0.25 (Meyer et al., 2011; Reyes-Puerta et al., 2015). Similarly to bursting neurons, interneurons can decrease firing after the transition to desynchronization (Gentet et al., 2010; Sakata and Harris, 2012). Nevertheless, the distribution of bursting neurons and interneurons through the barrel cortex does not strictly match L4 and B5B/6A and does not fully explain the differences between spontaneous and evoked MUA ratios.

We observed that MUA rates during the early response decreased in L4 and B5B/6A and then decreased through all cortical layers (late response); in other words, the response became shorter with cooling (**Figures 7B, 8**). On average, peak firing was not diminished with cooling. The L4 vs. L5 and L5 vs. B5B/6A contrast decreased in the late response compared to the early response. The sensory-evoked triphasic response is composed of fast initial excitation, prolonged (intracortical) inhibition, and then delayed excitation (Chapin et al., 1981; Zhu and Connors, 1999). A feature of the barrel cortex is fast VPM-dependent feed-forward inhibition lasting up to 10 ms after cortical activation (Agmon and Connors, 1992; Gil and Amitai, 1996; Gabernet et al., 2005; Sun et al., 2006; Cruikshank

et al., 2007). Interneurons are probably less sensitive to cooling than principal cells (Vizi and Sperlág, 1999; Motamedi et al., 2013). The decreased frequency ( $\sim 445$  Hz in control) and increased HFO peak power with cooling might indicate changes in neuron-interneuron interactions during thalamus-dependent feed-forward inhibition. It was reported that in regular-spiking cells the mean latency from onset of excitation to the onset of the inhibitory outward current was about 2.4 ms ( $\sim 420$  Hz) during disynaptic inhibition in L4 (Cruikshank et al., 2010). Therefore, we propose that hypothermia changed the excitation/inhibition balance (Higley and Contreras, 2006) towards interneurons during the sensory-evoked response. Spiny stellate cells and fast-spiking interneurons form clusters of highly interconnected cells in L4 (Feldmeyer, 2012; Koelbl et al., 2015). In mice, L6A also contains clusters of cells called infrabarrels (Crandall et al., 2017) and has a local inhibitory circuit (Frändolig et al., 2019). From a statistical point of view, the connectivity of local inhibitory cells (and overall inhibitory impact on MUA) might depend on local neuron density that has local maxima in L4 and B5B/6A. Also, it had been reported that tonic inhibition (Semyanov et al., 2004), mediated by extrasynaptic GABA(A) receptors, decreases with cooling (Bright and Smart, 2013), while activation of tonic inhibition in L4 (targeting fast-spiking interneurons) reduces thalamus-dependent feed-forward inhibition (Swadlow, 2003; Krook-Magnuson et al., 2008). Thus, cooling could provide a strong inhibitory effect on MUA in L4 and B5B/6A during the early response. The instantaneous amount of non-spiking cells (decreased MUA rate) was higher during hypothermia that could increase peak firing (**Figure 8A**) in the early response (Reig et al., 2015).

The thalamocortical circuit should be viewed as a single functional and dynamic entity (Crunelli and Hughes, 2010; Sheroziya and Timofeev, 2014) and thalamic contribution to cortical slow oscillations had been already shown (David et al., 2013; Lemieux et al., 2014). Neuronal activity in L4 and B5B/6A in the barrel cortex might strongly depend on the thalamus because cells within these cortical layers receive weak synaptic inputs from other cortical layers (compared to thalamocortical). Spiny stellate cells in L4 do not have excitatory inputs from other cortical layers in the home column while L4 star pyramids have weak and sparse synaptic inputs (Feldmeyer, 2012). Brief high-frequency stimulation of thalamocortical axons in slices triggers widespread recurrent activity in populations of neurons in L4 and then spreads into the adjacent L2/3 and L5 (Beierlein et al., 2002). In our experiments, cooling-evoked cortical desynchronization could decrease the occurrence of thalamic spontaneous low-threshold calcium bursts (Llinás and Yarom, 1981; Timofeev and Steriade, 1996) and respectively decrease spontaneous firing (**Figure 6**) in L4 and B5B/6A (Swadlow and Gusev, 2001). A cortical MUA response to a spontaneous thalamic activation could decrease similarly to the changes in the sensory-evoked MUA response during hypothermia, but the relative thalamic/intracortical impact on spontaneous MUA rates in L4 and B5B/6A remained unclear. Further investigations are required to verify thalamic contribution to vertical propagation of activity at the beginning of delta waves in the barrel cortex.



## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of Kazan State Medical University N9-2013.

## AUTHOR CONTRIBUTIONS

GB, KC, and MS performed the experiments. MS and RK designed the research project and wrote the article. MS analyzed the data.

## REFERENCES

- Abend, N. S., Dlugos, D. J., and Clancy, R. R. (2013). A review of long-term EEG monitoring in critically ill children with hypoxic-ischemic encephalopathy, congenital heart disease, ECMO, and stroke. *J. Clin. Neurophysiol.* 30, 134–142. doi: 10.1097/wnp.0b013e3182872af9
- Abend, N. S., Mani, R., Tschuda, T. N., Chang, T., Topjian, A. A., Donnelly, M., et al. (2011). EEG monitoring during therapeutic hypothermia in neonates, children, and adults. *Am. J. Electroneurodiagnostic. Technol.* 51, 141–164. doi: 10.1080/1086508x.2011.11079816
- Agmon, A., and Connors, B. W. (1992). Correlation between intrinsic firing patterns and thalamocortical synaptic responses of neurons in mouse barrel cortex. *J. Neurosci.* 12, 319–329. doi: 10.1523/JNEUROSCI.12-01-00319.1992
- Andresen, M., Gazmuri, J. T., Marin, A., Regueira, T., and Rovegno, M. (2015). Therapeutic hypothermia for acute brain injuries. *Scand. J. Trauma Resusc. Emerg. Med.* 23:42. doi: 10.1186/s13049-015-0121-3
- Armstrong-James, M., and George, M. J. (1988). Influence of anesthesia on spontaneous activity and receptive field size of single units in rat Sm1 neocortex. *Exp. Neurol.* 99, 369–387. doi: 10.1016/0014-4886(88)90155-0
- Baker, A. L., O'Toole, R. J., and Gullledge, A. T. (2018). Preferential cholinergic excitation of corticopontine neurons. *J. Physiol.* 596, 1659–1679. doi: 10.1113/jp275194
- Bartos, M., Vida, I., and Jonas, P. (2007). Synaptic mechanisms of synchronized  $\gamma$  oscillations in inhibitory interneuron networks. *Nat. Rev. Neurosci.* 8, 45–56. doi: 10.1038/nrn2044
- Bazhenov, M., Timofeev, I., Steriade, M., and Sejnowski, T. J. (2002). Model of thalamocortical slow-wave sleep oscillations and transitions to activated States. *J. Neurosci.* 22, 8691–8704. doi: 10.1523/JNEUROSCI.22-19-08691.2002
- Beierlein, M., Fall, C. P., Rinzel, J., and Yuste, R. (2002). Thalamocortical bursts trigger recurrent activity in neocortical networks: layer 4 as a frequency-dependent gate. *J. Neurosci.* 22, 9885–9894. doi: 10.1523/JNEUROSCI.22-22-09885.2002
- Beltramini, R., D'Urso, G., Dal Maschio, M., Farisello, P., Bovetti, S., Clovis, Y., et al. (2013). Layer-specific excitatory circuits differentially control recurrent network dynamics in the neocortex. *Nat. Neurosci.* 16, 227–234. doi: 10.1038/nn.3306
- Boylan, G. B., Kharoshankaya, L., and Wusthoff, C. J. (2015). Seizures and hypothermia: importance of electroencephalographic monitoring and considerations for treatment. *Semin. Fetal. Neonatal. Med.* 20, 103–108. doi: 10.1016/j.siny.2015.01.001
- Brecht, M., Roth, A., and Sakmann, B. (2003). Dynamic receptive fields of reconstructed pyramidal cells in layers 3 and 2 of rat somatosensory barrel cortex. *J. Physiol.* 553, 243–265. doi: 10.1113/jphysiol.2003.044222

## FUNDING

This work was supported by Russian Science Support Foundation (Ministry of Education and Science of the Russian Federation; 17-15-01271-P; electrophysiological experiments, data analysis) and the subsidy 075-00216-20-05 allocated to Kazan Federal University for the state assignment in the sphere of scientific activities (project 0671-2020-0059), and was performed within the Program of Competitive Growth of Kazan University and collaborative agreement between the French National Institute of Health and Medical Research (INSERM) and Kazan University.

## ACKNOWLEDGMENTS

We thank David Jappy for his helpful comments.

- Bright, D. P., and Smart, T. G. (2013). Methods for recording and measuring tonic GABAA receptor-mediated inhibition. *Front. Neural Circuits* 7:193. doi: 10.3389/fncir.2013.00193
- Brumberg, J. C., Nowak, L. G., and McCormick, D. A. (2000). Ionic mechanisms underlying repetitive high-frequency burst firing in supragranular cortical neurons. *J. Neurosci.* 20, 4829–4843. doi: 10.1523/JNEUROSCI.20-13-04829.2000
- Chagnac-Amitai, Y., Luhmann, H. J., and Prince, D. A. (1990). Burst generating and regular spiking layer 5 pyramidal neurons of rat neocortex have different morphological features. *J. Comp. Neurol.* 296, 598–613. doi: 10.1002/cne.902960407
- Chapin, J. K., Waterhouse, B. D., and Woodward, D. J. (1981). Differences in cutaneous sensory response properties of single somatosensory cortical neurons in awake and halothane anesthetized rats. *Brain Res. Bull.* 6, 63–70. doi: 10.1016/s0361-9230(81)80069-x
- Chauvette, S., Volgushev, M., and Timofeev, I. (2010). Origin of active states in local neocortical networks during slow sleep oscillation. *Cereb. Cortex* 20, 2660–2674. doi: 10.1093/cercor/bhq009
- Compte, A., Sanchez-Vives, M. V., McCormick, D. A., and Wang, X.-J. (2003). Cellular and network mechanisms of slow oscillatory activity (<1 Hz) and wave propagations in a cortical network model. *J. Neurophysiol.* 89, 2707–2725. doi: 10.1152/jn.00845.2002
- Connors, B. W., and Gutnick, M. J. (1990). Intrinsic firing patterns of diverse neocortical neurons. *Trends Neurosci.* 13, 99–104. doi: 10.1016/0166-2236(90)90185-d
- Connors, B. W., Gutnick, M. J., and Prince, D. A. (1982). Electrophysiological properties of neocortical neurons *in vitro*. *J. Neurophysiol.* 48, 1302–1320. doi: 10.1152/jn.1982.48.6.1302
- Constantinople, C. M., and Bruno, R. M. (2013). Deep cortical layers are activated directly by thalamus. *Science* 340, 1591–1594. doi: 10.1126/science.1236425
- Contreras, D. (2004). Electrophysiological classes of neocortical neurons. *Neural Netw.* 17, 633–646. doi: 10.1016/j.neunet.2004.04.003
- Crundall, S. R., Patrick, S. L., Cruikshank, S. J., and Connors, B. W. (2017). Infrabarrels are layer 6 circuit modules in the barrel cortex that link long-range inputs and outputs. *Cell Rep.* 21, 3065–3078. doi: 10.1016/j.celrep.2017.11.049
- Crochet, S., Chauvette, S., Boucetta, S., and Timofeev, I. (2005). Modulation of synaptic transmission in neocortex by network activities. *Eur. J. Neurosci.* 21, 1030–1044. doi: 10.1111/j.1460-9568.2005.03932.x
- Crochet, S., and Petersen, C. C. H. (2006). Correlating whisker behavior with membrane potential in barrel cortex of awake mice. *Nat. Neurosci.* 9, 608–610. doi: 10.1038/nn1690
- Cruikshank, S. J., Lewis, T. J., and Connors, B. W. (2007). Synaptic basis for intense thalamocortical activation of feedforward inhibitory cells in neocortex. *Nat. Neurosci.* 10, 462–468. doi: 10.1038/nn1861

- Cruikshank, S. J., Urabe, H., Nurmikko, A. V., and Connors, B. W. (2010). Pathway-specific feedforward circuits between thalamus and neocortex revealed by selective optical stimulation of axons. *Neuron* 65, 230–245. doi: 10.1016/j.neuron.2009.12.025
- Crunelli, V., and Hughes, S. W. (2010). The slow (>1 Hz) rhythm of non-REM sleep: a dialogue between three cardinal oscillators. *Nat. Neurosci.* 13, 9–17. doi: 10.1038/nn.2445
- David, F., Schmiedt, J. T., Taylor, H. L., Orban, G., Di Giovanni, G., Uebele, V. N., et al. (2013). Essential thalamic contribution to slow waves of natural sleep. *J. Neurosci.* 33, 19599–19610. doi: 10.1523/JNEUROSCI.3169-13.2013
- de Kock, C. P. J., Bruno, R. M., Spors, H., and Sakmann, B. (2007). Layer- and cell-type-specific suprathreshold stimulus representation in rat primary somatosensory cortex. *J. Physiol.* 581, 139–154. doi: 10.1113/jphysiol.2006.124321
- de Kock, C. P. J., and Sakmann, B. (2008). High frequency action potential bursts (>or= 100 Hz) in L2/3 and L5B thick tufted neurons in anaesthetized and awake rat primary somatosensory cortex. *J. Physiol.* 586, 3353–3364. doi: 10.1113/jphysiol.2008.155580
- de Kock, C. P. J., and Sakmann, B. (2009). Spiking in primary somatosensory cortex during natural whisking in awake head-restrained rats is cell-type specific. *Proc. Natl. Acad. Sci. U S A* 106, 16446–16450. doi: 10.1073/pnas.0904143106
- de la Peña, E., Mäklä, A., Vara, H., Caires, R., Ballesta, J. J., Belmonte, C., et al. (2012). The influence of cold temperature on cellular excitability of hippocampal networks. *PLoS One* 7:e52475. doi: 10.1371/journal.pone.0052475
- Eggermann, E., and Feldmeyer, D. (2009). Cholinergic filtering in the recurrent excitatory microcircuit of cortical layer 4. *Proc. Natl. Acad. Sci. U S A* 106, 11753–11758. doi: 10.1073/pnas.0810062106
- Eggermann, E., Kremer, Y., Crochet, S., and Petersen, C. C. H. (2014). Cholinergic signals in mouse barrel cortex during active whisker sensing. *Cell Rep.* 9, 1654–1660. doi: 10.1016/j.celrep.2014.11.005
- Feldman, M. L. (1984). “Morphology of the neocortical pyramidal neuron,” in *Cerebral Cortex Cellular Components of the Cerebral Cortex*, eds A. Peters and E. G. Jones (New York, NY: Plenum), 123–200.
- Feldmeyer, D. (2012). Excitatory neuronal connectivity in the barrel cortex. *Front. Neuroanat.* 6:24. doi: 10.3389/fnana.2012.00024
- Fiáth, R., Kerekes, B. P., Wittner, L., Tãth, K., Beregszászi, P., Horváth, D., et al. (2016). Laminar analysis of the slow wave activity in the somatosensory cortex of anesthetized rats. *Eur. J. Neurosci.* 44, 1935–1951. doi: 10.1111/ejn.13274
- Frändolig, J. E., Matney, C. J., Lee, K., Kim, J., Chevee, M., Kim, S. J., et al. (2019). The synaptic organization of layer 6 circuits reveals inhibition as a major output of a neocortical sublamina. *Cell Rep.* 28, 3131.e5–3143.e5. doi: 10.1016/j.celrep.2019.08.048
- Frauscher, B., Bartolomei, F., Kobayashi, K., Cimbálik, J., van ’t Klooster, M. A., Rapp, S., et al. (2017). High-frequency oscillations: the state of clinical research. *Epilepsia* 58, 1316–1329. doi: 10.1111/epi.13829
- Gabernet, L., Jadhav, S. P., Feldman, D. E., Carandini, M., and Scanziani, M. (2005). Somatosensory integration controlled by dynamic thalamocortical feed-forward inhibition. *Neuron* 48, 315–327. doi: 10.1016/j.neuron.2005.09.022
- Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F., and Petersen, C. C. H. (2010). Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. *Neuron* 65, 422–435. doi: 10.1016/j.neuron.2010.01.006
- Gil, Z., and Amitai, Y. (1996). Properties of convergent thalamocortical and intracortical synaptic potentials in single neurons of neocortex. *J. Neurosci.* 16, 6567–6578. doi: 10.1523/JNEUROSCI.16-20-06567.1996
- Gil, Z., Connors, B. W., and Amitai, Y. (1997). Differential regulation of neocortical synapses by neuromodulators and activity. *Neuron* 19, 679–686. doi: 10.1016/S0896-6273(00)80380-3
- Girardin, C. C., and Martin, K. A. C. (2009). Cooling in cat visual cortex: stability of orientation selectivity despite changes in responsiveness and spike width. *Neuroscience* 164, 777–787. doi: 10.1016/j.neuroscience.2009.07.064
- Gluckman, P. D., Wyatt, J. S., Azzopardi, D., Ballard, R., Edwards, A. D., Ferriero, D. M., et al. (2005). Selective head cooling with mild systemic hypothermia after neonatal encephalopathy: multicentre randomised trial. *Lancet* 365, 663–670. doi: 10.1016/S0140-6736(05)17946-X
- Goard, M., and Dan, Y. (2009). Basal forebrain activation enhances cortical coding of natural scenes. *Nat. Neurosci.* 12, 1444–1449. doi: 10.1038/nn.2402
- Goila, A. K., and Pawar, M. (2009). The diagnosis of brain death. *Indian J. Crit. Care Med.* 13, 7–11. doi: 10.4103/0972-5229.53108
- Gray, C. M., and McCormick, D. A. (1996). Chattering cells: superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. *Science* 274, 109–113. doi: 10.1126/science.274.5284.109
- Gunn, A. J., Laptok, A. R., Robertson, N. J., Barks, J. D., Thoresen, M., Wassink, G., et al. (2017). Therapeutic hypothermia translates from ancient history in to practice. *Pediatr. Res.* 81, 202–209. doi: 10.1038/pr.2016.198
- Hardingham, N. R., and Larkman, A. U. (1998). Rapid report: the reliability of excitatory synaptic transmission in slices of rat visual cortex *in vitro* is temperature dependent. *J. Physiol.* 507, 249–256. doi: 10.1111/j.1469-7793.1998.249bu.x
- Hasselmo, M. E., and McGaughy, J. (2004). High acetylcholine levels set circuit dynamics for attention and encoding and low acetylcholine levels set dynamics for consolidation. *Prog. Brain Res.* 145, 207–231. doi: 10.1016/S0079-6123(03)45015-2
- Hasselmo, M. E., and Sarter, M. (2011). Modes and models of forebrain cholinergic neuromodulation of cognition. *Neuropsychopharmacology* 36, 52–73. doi: 10.1038/npp.2010.104
- Hedrick, T., and Waters, J. (2012). Effect of temperature on spiking patterns of neocortical layer 2/3 and layer 6 pyramidal neurons. *Front. Neural Circuits* 6:28. doi: 10.3389/fncir.2012.00028
- Henze, D. A., Borhegyi, Z., Csicsvari, J., Mamiya, A., Harris, K. D., and Buzsáki, G. (2000). Intracellular features predicted by extracellular recordings in the hippocampus *in vivo*. *J. Neurophysiol.* 84, 390–400. doi: 10.1152/jn.2000.84.1.390
- Higley, M. J., and Contreras, D. (2006). Balanced excitation and inhibition determine spike timing during frequency adaptation. *J. Neurosci.* 26, 448–457. doi: 10.1523/JNEUROSCI.3506-05.2006
- Hill, S., and Tononi, G. (2005). Modeling sleep and wakefulness in the thalamocortical system. *J. Neurophysiol.* 93, 1671–1698. doi: 10.1152/jn.00915.2004
- Hirata, A., and Castro-Alamancos, M. A. (2010). Neocortex network activation and deactivation states controlled by the thalamus. *J. Neurophysiol.* 103, 1147–1157. doi: 10.1152/jn.00955.2009
- Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544. doi: 10.1113/jphysiol.1952.sp004764
- Hypothermia after Cardiac Arrest Study Group. (2002). Mild therapeutic hypothermia to improve the neurologic outcome after cardiac arrest. *N. Engl. J. Med.* 346, 549–556. doi: 10.1056/NEJMoa012689
- Jasper, H. H., Shacter, D. G., and Montplaisir, J. (1970). The effect of local cooling upon spontaneous and evoked electrical activity of cerebral cortex. *Can. J. Physiol. Pharmacol.* 48, 640–652. doi: 10.1139/y70-094
- Jones, M. S., and Barth, D. S. (1999). Spatiotemporal organization of fast (>200 Hz) electrical oscillations in rat Vibrissa/Barrel cortex. *J. Neurophysiol.* 82, 1599–1609. doi: 10.1152/jn.1999.82.3.1599
- Kang, D., Choe, C., and Kim, D. (2005). Thermosensitivity of the two-pore domain K<sup>+</sup> channels TREK-2 and TRAAK. *J. Physiol.* 564, 103–116. doi: 10.1113/jphysiol.2004.081059
- Katz, B., and Miledi, R. (1965). The effect of temperature on the synaptic delay at the neuromuscular junction. *J. Physiol.* 181, 656–670. doi: 10.1113/jphysiol.1965.sp007790
- Kerr, J. N. D., de Kock, C. P. J., Greenberg, D. S., Bruno, R. M., Sakmann, B., and Helmchen, F. (2007). Spatial organization of neuronal population responses in layer 2/3 of rat barrel cortex. *J. Neurosci.* 27, 13316–13328. doi: 10.1523/JNEUROSCI.2210-07.2007
- Khazipov, R., Minlebaev, M., and Valeeva, G. (2013). Early  $\gamma$  oscillations. *Neuroscience* 250, 240–252. doi: 10.1016/j.neuroscience.2013.07.019
- Kiyatkin, E. A. (2007). Brain temperature fluctuations during physiological and pathological conditions. *Eur. J. Appl. Physiol.* 101, 3–17. doi: 10.1007/s00421-007-0450-7
- Koelbl, C., Helmstaedter, M., Lübke, J., and Feldmeyer, D. (2015). A barrel-related interneuron in layer 4 of rat somatosensory cortex with a high intrabarrel connectivity. *Cereb. Cortex* 25, 713–725. doi: 10.1093/cercor/bht263
- Krnjević, K., Pumain, R., and Renaud, L. (1971). The mechanism of excitation by acetylcholine in the cerebral cortex. *J. Physiol.* 215, 247–268. doi: 10.1113/jphysiol.1971.sp009467

- Krook-Magnuson, E. I., Li, P., Paluszkievicz, S. M., and Huntsman, M. M. (2008). Tonic active inhibition selectively controls feedforward circuits in mouse barrel cortex. *J. Neurophysiol.* 100, 932–944. doi: 10.1152/jn.01360.2007
- Lee, S.-H., and Dan, Y. (2012). Neuromodulation of brain states. *Neuron* 76, 209–222. doi: 10.1016/j.neuron.2012.09.012
- Lemieux, M., Chen, J. Y., Lonjers, P., Bazhenov, M., and Timofeev, I. (2014). The impact of cortical deafferentation on the neocortical slow oscillation. *J. Neurosci.* 34, 5689–5703. doi: 10.1523/JNEUROSCI.1156-13.2014
- Lemyre, B., and Chau, V. (2018). Hypothermia for newborns with hypoxic-ischemic encephalopathy. *Paediatr. Child Health* 23, 285–291. doi: 10.1093/pch/pxy028
- Lesage, F. (2003). Pharmacology of neuronal background potassium channels. *Neuropharmacology* 44, 1–7. doi: 10.1016/s0028-3908(02)00339-8
- Llinás, R. R., Grace, A. A., and Yarom, Y. (1991). *In vitro* neurons in mammalian cortical layer 4 exhibit intrinsic oscillatory activity in the 10- to 50-Hz frequency range. *Proc. Natl. Acad. Sci. U S A* 88, 897–901. doi: 10.1073/pnas.88.3.897
- Llinás, R. R., and Yarom, Y. (1981). Properties and distribution of ionic conductances generating electroresponsiveness of mammalian inferior olivary neurones *in vitro*. *J. Physiol.* 315, 569–584. doi: 10.1113/jphysiol.1981.sp.013764
- Lomber, S. G., Cornwell, P., Sun, J. S., MacNeil, M. A., and Payne, B. R. (1994). Reversible inactivation of visual processing operations in middle suprasylvian cortex of the behaving cat. *Proc. Natl. Acad. Sci. U S A* 91, 2999–3003. doi: 10.1073/pnas.91.8.2999
- Long, M. A., and Fee, M. S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456, 189–194. doi: 10.1038/nature07448
- McCormick, D. A. (1992). Neurotransmitter actions in the thalamus and cerebral cortex and their role in neuromodulation of thalamocortical activity. *Prog. Neurobiol.* 39, 337–388. doi: 10.1016/0301-0082(92)90012-4
- Meyer, H. S., Schwarz, D., Wimmer, V. C., Schmitt, A. C., Kerr, J. N. D., Sakmann, B., et al. (2011). Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5A. *Proc. Natl. Acad. Sci. U S A* 108, 16807–16812. doi: 10.1073/pnas.1113648108
- Meyer, H. S., Wimmer, V. C., Hemberger, M., Bruno, R. M., de Kock, C. P. J., Frick, A., et al. (2010a). Cell type-specific thalamic innervation in a column of rat vibrissa cortex. *Cereb. Cortex* 20, 2287–2303. doi: 10.1093/cercor/bhq069
- Meyer, H. S., Wimmer, V. C., Oberlaender, M., de Kock, C. P. J., Sakmann, B., and Helmstaedt, M. (2010b). Number and laminar distribution of neurons in a thalamocortical projection column of rat vibrissa cortex. *Cereb. Cortex* 20, 2277–2286. doi: 10.1093/cercor/bhq067
- Minlebaev, M., Ben-Ari, Y., and Khazipov, R. (2007). Network mechanisms of spindle-burst oscillations in the neonatal rat barrel cortex *in vivo*. *J. Neurophysiol.* 97, 692–700. doi: 10.1152/jn.00759.2006
- Miyazawa, T., Tamura, A., Fukui, S., and Hossmann, K.-A. (2003). Effect of mild hypothermia on focal cerebral ischemia. Review of experimental studies. *Neurol. Res.* 25, 457–464. doi: 10.1179/016164103101201850
- Motamedi, G. K., Lesser, R. P., and Vicini, S. (2013). Therapeutic brain hypothermia, its mechanisms of action and its prospects as a treatment for epilepsy. *Epilepsia* 54, 959–970. doi: 10.1111/epi.12144
- Neske, G. T. (2015). The slow oscillation in cortical and thalamic networks: mechanisms and functions. *Front. Neural Circuits* 9:88. doi: 10.3389/fncir.2015.00088
- Núñez, A., Amzica, F., and Steriade, M. (1993). Electrophysiology of cat association cortical cells *in vivo*: intrinsic properties and synaptic responses. *J. Neurophysiol.* 70, 418–430. doi: 10.1152/jn.1993.70.1.418
- Poulet, J. F. A., and Crochet, S. (2018). The cortical states of wakefulness. *Front. Syst. Neurosci.* 12:64. doi: 10.3389/fnsys.2018.00064
- Poulet, J. F. A., and Petersen, C. C. H. (2008). Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* 454, 881–885. doi: 10.1038/nature07150
- Prönneke, A., Witte, M., Möck, M., and Staiger, J. F. (2020). Neuromodulation leads to a burst-tonic switch in a subset of VIP neurons in mouse primary somatosensory (barrel) cortex. *Cereb. Cortex* 30, 488–504. doi: 10.1093/cercor/bhz102
- Reig, R., Zerlaut, Y., Vergara, R., Destexhe, A., and Sanchez-Vives, M. V. (2015). Gain modulation of synaptic inputs by network state in auditory cortex *in vivo*. *J. Neurosci.* 35, 2689–2702. doi: 10.1523/JNEUROSCI.2004-14.2015
- Reimer, J., Froudarakis, E., Cadwell, C. R., Yatsenko, D., Denfield, G. H., and Tolias, A. S. (2014). Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron* 84, 355–362. doi: 10.1016/j.neuron.2014.09.033
- Reyes-Puerta, V., Sun, J.-J., Kim, S., Kilb, W., and Luhmann, H. J. (2015). Laminar and columnar structure of sensory-evoked multineuronal spike sequences in adult rat barrel cortex *in vivo*. *Cereb. Cortex* 25, 2001–2021. doi: 10.1093/cercor/bhu007
- Reyes-Puerta, V., Yang, J. W., Siwek, M. E., Kilb, W., Sun, J. J., and Luhmann, H. J. (2016). Propagation of spontaneous slow-wave activity across columns and layers of the adult rat barrel cortex *in vivo*. *Brain Struct. Funct.* 221, 4429–4449. doi: 10.1007/s00429-015-1173-x
- Rosanova, M., and Timofeev, I. (2005). Neuronal mechanisms mediating the variability of somatosensory evoked potentials during sleep oscillations in cats. *J. Physiol.* 562, 569–582. doi: 10.1113/jphysiol.2004.071381
- Sachdev, R. N. S., Ebner, F. F., and Wilson, C. J. (2004). Effect of subthreshold up and down states on the whisker-evoked response in somatosensory cortex. *J. Neurophysiol.* 92, 3511–3521. doi: 10.1152/jn.00347.2004
- Sakata, S., and Harris, K. D. (2009). Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. *Neuron* 64, 404–418. doi: 10.1016/j.neuron.2009.09.020
- Sakata, S., and Harris, K. D. (2012). Laminar-dependent effects of cortical state on auditory cortical spontaneous activity. *Front. Neural Circuits* 6:109. doi: 10.3389/fncir.2012.00109
- Sanchez-Vives, M. V., and McCormick, D. A. (2000). Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nat. Neurosci.* 3, 1027–1034. doi: 10.1038/79848
- Schneider, E. R., Anderson, E. O., Gracheva, E. O., and Bagriantsev, S. N. (2014). Temperature sensitivity of two-pore (K2P) potassium channels. *Curr. Top. Membr.* 74, 113–133. doi: 10.1016/b978-0-12-800181-3.00005-1
- Schubert, D., Staiger, J. F., Cho, N., Kötter, R., Zilles, K., and Luhmann, H. J. (2001). Layer-specific intracolumnar and transcolumnar functional connectivity of layer V pyramidal cells in rat barrel cortex. *J. Neurosci.* 21, 3580–3592. doi: 10.1523/JNEUROSCI.21-10-03580.2001
- Schwalm, M., and Easton, C. (2016). Cortical temperature change: a tool for modulating brain states? *eNeuro* 3:ENEURO.0096-16.2016. doi: 10.1523/ENEURO.0096-16.2016
- Sekhon, M. S., Ainslie, P. N., and Griesdale, D. E. (2017). Clinical pathophysiology of hypoxic ischemic brain injury after cardiac arrest: a “two-hit” model. *Crit. Care* 21:90. doi: 10.1186/s13054-017-1670-9
- Semyanov, A., Walker, M. C., Kullmann, D. M., and Silver, R. A. (2004). Tonic active GABA A receptors: modulating gain and maintaining the tone. *Trends Neurosci.* 27, 262–269. doi: 10.1016/j.tins.2004.03.005
- Senzai, Y., Fernandez-Ruiz, A., and Buzsáki, G. (2019). Layer-specific physiological features and interlaminar interactions in the primary visual cortex of the mouse. *Neuron* 101, 500.e5–513.e5. doi: 10.1016/j.neuron.2018.12.009
- Shankaran, S., Laptook, A. R., Ehrenkranz, R. A., Tyson, J. E., McDonald, S. A., Donovan, E. F., et al. (2005). Whole-body hypothermia for neonates with hypoxic-ischemic encephalopathy. *N. Engl. J. Med.* 353, 1574–1584. doi: 10.1056/NEJMcps050929
- Sheroziya, M., and Timofeev, I. (2014). Global intracellular slow-wave dynamics of the thalamocortical system. *J. Neurosci.* 34, 8875–8893. doi: 10.1523/JNEUROSCI.4460-13.2014
- Sheroziya, M., and Timofeev, I. (2015). Moderate cortical cooling eliminates thalamocortical silent states during slow oscillation. *J. Neurosci.* 35, 13006–13019. doi: 10.1523/JNEUROSCI.1359-15.2015
- Shu, Y., Hasenstaub, A., Badoual, M., Bal, T., and McCormick, D. A. (2003a). Barrages of synaptic activity control the gain and sensitivity of cortical neurons. *J. Neurosci.* 23, 10388–10401. doi: 10.1523/JNEUROSCI.23-23-10388.2003
- Shu, Y., Hasenstaub, A., and McCormick, D. A. (2003b). Turning on and off recurrent balanced cortical activity. *Nature* 423, 288–293. doi: 10.1038/nature01616

- Sirota, A., Montgomery, S., Fujisawa, S., Isomura, Y., Zugaro, M., and Buzsaki, G. (2008). Entrainment of neocortical neurons and  $\gamma$  oscillations by the hippocampal theta rhythm. *Neuron* 60, 683–697. doi: 10.1016/j.neuron.2008.09.014
- Staba, R. J., Brett-Green, B., Paulsen, M., and Barth, D. S. (2003). Effects of ventrobasal lesion and cortical cooling on fast oscillations ( $>200$  Hz) in rat somatosensory cortex. *J. Neurophysiol.* 89, 2380–2388. doi: 10.1152/jn.010.98.2002
- Staiger, J. F., Flagmeyer, I., Schubert, D., Zilles, K., Kotter, R., and Luhmann, H. J. (2004). Functional diversity of layer IV spiny neurons in rat somatosensory cortex: quantitative morphology of electrophysiologically characterized and biocytin labeled cells. *Cereb. Cortex* 14, 690–701. doi: 10.1093/cercor/bhh029
- Steriade, M., McCormick, D. A., and Sejnowski, T. J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science* 262, 679–685. doi: 10.1126/science.8235588
- Steriade, M., Timofeev, I., and Grenier, F. (2001). Natural waking and sleep states: a view from inside neocortical neurons. *J. Neurophysiol.* 85, 1969–1985. doi: 10.1152/jn.2001.85.5.1969
- Sun, Q. Q., Huguenard, J. R., and Prince, D. A. (2006). Barrel cortex microcircuits: thalamocortical feedforward inhibition in spiny stellate cells is mediated by a small number of fast-spiking interneurons. *J. Neurosci.* 26, 1219–1230. doi: 10.1523/JNEUROSCI.4727-04.2006
- Swadlow, H. A. (2003). Fast-spike interneurons and feedforward inhibition in awake sensory neocortex. *Cereb. Cortex* 13, 25–32. doi: 10.1093/cercor/13.1.25
- Swadlow, H. A., and Gusev, A. G. (2001). The impact of “bursting” thalamic impulses at a neocortical synapse. *Nat. Neurosci.* 4, 402–408. doi: 10.1038/86054
- Talley, E. M., Solorzano, G., Lei, Q., Kim, D., and Bayliss, D. A. (2001). Cns distribution of members of the two-pore-domain (KCNK) potassium channel family. *J. Neurosci.* 21, 7491–7505. doi: 10.1523/JNEUROSCI.21-19-074.91.2001
- Thoresen, M., and Whitelaw, A. (2005). Therapeutic hypothermia for hypoxic-ischaemic encephalopathy in the newborn infant. *Curr. Opin. Neurol.* 18, 111–116. doi: 10.1097/01.wco.0000162850.44897.c6
- Timofeev, I., Grenier, F., and Steriade, M. (2001). Disfacilitation and active inhibition in the neocortex during the natural sleep-wake cycle: an intracellular study. *Proc. Natl. Acad. Sci. U S A* 98, 1924–1929. doi: 10.1073/pnas.98.4.1924
- Timofeev, I., and Steriade, M. (1996). Low-frequency rhythms in the thalamus of intact-cortex and decorticated cats. *J. Neurophysiol.* 76, 4152–4168. doi: 10.1152/jn.1996.76.6.4152
- Trevelyan, A. J., and Jack, J. (2002). Detailed passive cable models of layer 2/3 pyramidal cells in rat visual cortex at different temperatures. *J. Physiol.* 539, 623–636. doi: 10.1113/jphysiol.2001.013291
- van der Worp, H. B., Macleod, M. R., Kollmar, R., and European Stroke Research Network for, H. (2010). Therapeutic hypothermia for acute ischemic stroke: ready to start large randomized trials? *J. Cereb. Blood Flow Metab.* 30, 1079–1093. doi: 10.1038/jcbfm.2010.44
- Vinokurova, D., Zakharov, A. V., Lebedeva, J., Burkhanova, G. F., Chernova, K. A., Lotfullina, N., et al. (2018). Pharmacodynamics of the glutamate receptor antagonists in the rat barrel cortex. *Front. Pharmacol.* 9:698. doi: 10.3389/fphar.2018.00698
- Vizi, E. S., and Sperl gh, B. (1999). Separation of carrier mediated and vesicular release of GABA from rat brain slices. *Neurochem. Int.* 34, 407–413. doi: 10.1016/s0197-0186(99)00047-9
- Volgushev, M., Kudryashov, I., Chistiakova, M., Mukovski, M., Niesmann, J., and Eysel, U. T. (2004). Probability of transmitter release at neocortical synapses at different temperatures. *J. Neurophysiol.* 92, 212–220. doi: 10.1152/jn.011.66.2003
- Volgushev, M., Vidyasagar, T. R., Chistiakova, M., Yousef, T., and Eysel, U. T. (2000). Membrane properties and spike generation in rat visual cortical cells during reversible cooling. *J. Physiol.* 522, 59–76. doi: 10.1111/j.1469-7793.2000.0059m.x
- Wang, H., Wang, B., Normoyle, K. P., Jackson, K., Spitler, K., Sharrock, M. F., et al. (2014). Brain temperature and its fundamental properties: a review for clinical neuroscientists. *Front. Neurosci.* 8:307. doi: 10.3389/fnins.2014.00307
- Wester, J. C., and Contreras, D. (2012). Columnar interactions determine horizontal propagation of recurrent network activity in neocortex. *J. Neurosci.* 32, 5454–5471. doi: 10.1523/JNEUROSCI.5006-11.2012
- Whittington, M. A., Cunningham, M. O., LeBeau, F. E., Racca, C., and Traub, R. D. (2011). Multiple origins of the cortical  $\gamma$  rhythm. *Dev. Neurobiol.* 71, 92–106. doi: 10.1002/dneu.20814
- Yang, X. F., Ouyang, Y., Kennedy, B. R., and Rothman, S. M. (2005). Cooling blocks rat hippocampal neurotransmission by a presynaptic mechanism: observations using 2-photon microscopy. *J. Physiol.* 567, 215–224. doi: 10.1113/jphysiol.2005.088948
- Zhang, Y. S., Wittenbach, J. D., Jin, D. Z., and Kozhevnikov, A. A. (2017). Temperature manipulation in songbird brain implicates the premotor nucleus HVC in birdsong syntax. *J. Neurosci.* 37, 2600–2611. doi: 10.1523/JNEUROSCI.1827-16.2017
- Zhu, J. J., and Connors, B. W. (1999). Intrinsic firing patterns and whisker-evoked synaptic responses of neurons in the rat barrel cortex. *J. Neurophysiol.* 81, 1171–1183. doi: 10.1152/jn.1999.81.3.1171

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Burkhanova, Chernova, Khazipov and Sheroziya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Accounting for Changing Structure in Functional Network Analysis of TBI Patients

John Dell'Italia<sup>1\*</sup>, Micah A. Johnson<sup>1</sup>, Paul M. Vespa<sup>2</sup> and Martin M. Monti<sup>1,2</sup>

<sup>1</sup> Department of Psychology, University of California, Los Angeles, Los Angeles, CA, United States, <sup>2</sup> Brain Injury Research Center (BIRC), Department of Neurosurgery, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

## OPEN ACCESS

### Edited by:

Maurizio Mattia,  
Istituto Superiore di Sanità (ISS), Italy

### Reviewed by:

Zoltan Somogyvari,  
Wigner Research Centre for  
Physics, Hungary  
Marco Cambiaghi,  
University of Verona, Italy

### \*Correspondence:

John Dell'Italia  
johndellitalia@ucla.edu

**Received:** 30 June 2019

**Accepted:** 05 June 2020

**Published:** 07 August 2020

### Citation:

Dell'Italia J, Johnson MA, Vespa PM  
and Monti MM (2020) Accounting for  
Changing Structure in Functional  
Network Analysis of TBI Patients.  
*Front. Syst. Neurosci.* 14:42.  
doi: 10.3389/fnsys.2020.00042

Over the last 15 years, network analysis approaches based on MR data have allowed a renewed understanding of the relationship between brain function architecture and consciousness. Application of this approach to Disorders of Consciousness (DOC) highlights the relationship between specific aspects of network topology and levels of consciousness. Nonetheless, such applications do not acknowledge that DOC patients present with a dramatic level of heterogeneity in structural connectivity (SC) across groups (e.g., etiology, diagnostic categories) and within individual patients (e.g., over time), which possibly affects the level and quality of functional connectivity (FC) patterns that can be expressed. In addition, it is rarely acknowledged that the most frequently employed outcome metrics in the study of brain connectivity (e.g., degree distribution, inter- or intra-resting state network connectivity, and clustering coefficient) are interrelated and cannot be assumed to be independent of each other. We present empirical data showing that, when the two points above are not taken into consideration with an appropriate analytic model, it can lead to a misinterpretation of the role of each outcome metric in the graph's structure and thus misinterpretation of FC results. We show that failing to account for either SC or the inter-relation between outcome measures can lead to inflated false positives (FP) and/or false negatives (FN) in inter- or intra-resting state network connectivity results (defined, respectively, as a positive or negative result in network connectivity that is present when not accounting for SC and/or outcome measure inter-relation, but becomes not significant when accounting for all variables). Overall, we find that unconscious patients have lower rates of FP and FN for within cortical connectivity, lower rates of FN for cortico-subcortical connectivity, and lower rates of FP for within subcortical connectivity. These lower rates in unconscious patients may reflect differences in their triadic closure and SC metrics, which bias the interpretations of the inter- or intra-resting state network connectivity if the SC metrics and triadic closure are not modeled. We suggest that future studies of functional connectivity in DOC patients (i) incorporate where possible SC metrics and (ii) properly account for the intercorrelated nature of outcome variables.

**Keywords:** network analysis, exponential random graph model, functional magnetic resonance imaging, coma, disorders of consciousness

# 1. INTRODUCTION

Over the last 15 years, the study of the functional organization of the human brain under no overt task-set (i.e., so-called resting state fMRI, rsfMRI; Raichle et al., 2001; Van Dijk et al., 2010), has given rise to an explosion in the study of the relationship between the functional brain network properties and cognitive variables (e.g., Zou et al., 2013; Reineberg et al., 2015), biological variables (e.g., Dosenbach et al., 2010; Wang et al., 2013), and disease (e.g., Sorg et al., 2007; Hacker et al., 2012; Pandit et al., 2013; Werner et al., 2014). Similarly, this approach has also been applied toward the understanding the neural underpinnings of consciousness and its disorders (e.g., Monti et al., 2013; Chennu et al., 2014; Crone et al., 2015, 2018; Demertzi et al., 2019).

Despite the popularity and appeal of using network-based description of brain function to assess task-free neuroimaging data, there are a number of important challenges that must be addressed. We have previously made the case for four main shortcomings in current “standard” approach to applying graph-theoretic methods to task-free neuroimaging data, some of which relate to the implementation of the method itself, while others relate to peculiarities of Disorders of Consciousness (DOC) data (cf. Dell'Italia et al., 2018). In this work, we start by providing context on the four shortcomings and then focus on two specific problems which we empirically show lead to inflated rates of false positives and false negatives by virtue of mis-specifying the model.

The first problem we discussed in our prior work (i.e., Dell'Italia et al., 2018) relates to the fact that network properties are typically estimated on the basis of *sparse* matrices—that is to say, on networks where each node only connects to a subset of other nodes, as opposed to fully connected networks where each node is connected to all other nodes. Functional networks, however, conventionally obtained by correlating across time-courses of a large number of regions of interest (ROI), are fully connected. To overcome this issue, it is conventional to make functional networks sparse by selecting a single arbitrary level of sparseness (so-called proportional thresholding), an arbitrary minimum correlation strength  $r$  below which connections are set to 0 (so-called absolute thresholding), or by computing network summary statistics over several different sparsity levels, typically between the maximum sparsity of  $2 \times \log(N)$ , with  $N$  being the number of nodes in the network, which guarantees that networks are estimable, and the lowest level of sparsity that still yields a network with small-world characteristic  $\sigma$  no lower than 1 (Watts and Strogatz, 1998). Importantly, however, summary measures can only be compared across networks that have the same number of connections (i.e., sparsity), thus leading to the requirement to impose the same sparsity across groups (e.g., healthy volunteers vs. severe brain injury patients, minimally conscious state vs. vegetative state patients) or different time-points along the recovery trajectory (e.g., acute vs. chronic), which are very likely to have different levels of “natural density” (cf., Dell'Italia et al., 2018), thus biasing results toward whichever set of networks happens to have natural density closest to the enforced common density level. In response to this issue, some

have proposed the use of complex networks (Rubinov and Sporns, 2011; Fornito et al., 2013, 2016, i.e., networks that are fully connected and feature both positive and negative edges). To date, however, this approach has only found very limited application ( $\approx 7\%$ ; Nielsen et al., 2013).

Second, it is conventional to treat network outcome statistics (e.g., nodal degree, clustering coefficient, etc) as independent of each other. Yet, these outcome statistics are typically interrelated, which can lead to spurious results. To explain, three generative processes are believed to be key in generating a network's topological structure: sociality, selective mixing, and triadic closure. Sociality, a process that operates at the single-node level, refers to the propensity of some nodes to generate more edges than others (e.g., the propensity of different individuals to make friends; Goodreau et al., 2009). Selective mixing, a dyadic-level process, refers to the propensity of an edge to form between two nodes on the basis of some nodal attributes (e.g., the propensity of an individual to make friends with other individuals of the same political or religious persuasion; McPherson et al., 2001; Goodreau et al., 2009). Assortative mixing indicates a greater propensity for edges to form between nodes of the same category whereas dissortative mixing indexes the opposite pattern. Finally, triadic closure, a triadic-level process, refers to the propensity of an edge to form between two nodes  $p$  and  $q$  that are each already connected to the same third node  $r$  (e.g., the propensity of friends of friends to also be friends; Goodreau et al., 2009). While it is often assumed that each outcome metric maps in a one-to-one fashion to the generative processes, this is typically not the case (cf., Figure 2 in Goodreau et al., 2009). Clustering coefficient, for example, defined as the average fraction of a node's neighbors that are also neighbors of each other (Watts and Strogatz, 1998), is often employed as an outcome measure of triadic closure. Indeed, the greater the triadic closure, the greater the clustering coefficient. Yet, selective mixing can also affect the clustering coefficient: when edges are more likely to form between nodes of the same category (i.e., assortative mixing), closure of the triangle is also more likely to occur (leading to a greater clustering coefficient). To exemplify, friends of friends that share membership in a group (e.g., a synagogue, a book club), are more likely to be—themselves—friends, as compared to friends of friends who do not share such a group membership (i.e., dissortative mixing). In the face of this interdependence between outcome metrics, independent testing of the metrics is not advisable (albeit very frequent) since it can lead to misinterpretation and spurious results.

The third major problem in applying network analysis methods to DOC concerns the fact that there are currently no methods for incorporating structural connectivity in the characterization of functional networks. Structural connectivity plays an important role in shaping the functional connectivity that someone can express (cf., Messé et al., 2015,?; Finger et al., 2016). In the context of DOC patients, the high heterogeneity in underlying brain damage both within and across diagnostic categories, as well as the potential for compensatory neuroplastic mechanisms over time (Voss and Schiff, 2009; Demertzi et al., 2011), makes it all the more important that structural information be incorporated in the analysis of functional

networks, not to make spurious inferences (see Dell'Italia et al., 2018).

Finally, although we will not discuss the topic in the present work, current analytic frameworks do not allow assessing the dynamics of network change over time. Rather, they rely on the comparison of static characterization of brain networks at different time-points along the recovery trajectory (see Dell'Italia et al., 2018).

In what follows, we adopt the powerful and flexible framework of exponential random graph models (ERGM; Hunter et al., 2008) to estimating network statistics to empirically show that, in patients recovering from coma and a patient population and in healthy volunteers (HCP; Van Essen et al., 2013), problem #2 and problem #3, above, lead to false positives (FPs) and false negatives (FNs) in the estimated networks and that, in the patients, FPs and FNs occur at different rates depending on level of consciousness and time since injury (i.e., acute and 6-month follow-up). We define FPs and FNs by comparing four different models with terms for triadic closure and/or SC included to a base model without those terms. A FP is identified as the base model having a significant parameter estimate (PE) for a specific type of mixing (e.g., within default network connectivity) compared to a model that includes an outcome metric for triadic closure (or including SC metrics) produces a non-significant PE for that specific type of mixing. While a FN is identified as a non-significant PE for a specific type of mixing compared to a model that includes an outcome metric for triadic closure (or including SC metrics) produces a non-significant PE for that specific type of mixing. Thus, we are identifying the effects omitting a triadic closure term (i.e., the GWESP effect), SC metrics (i.e., the structural effect), or both triadic closure and SC metrics (i.e., the interaction effect) on FP and FN rates for selective mixing of resting state connectivity. In our current study, we focused on rate differences of FPs and FNs for selective mixing because much of the DOC research is focused on key regions that form the physical substrate of consciousness. Overall, leaving out important generative processes will produce many FPs and FNs, but rate differences will produced biases in the FPs and FNs that may be interpreted as true differences in patient populations. Thus, we wanted to explore the possibility of rate differences with three questions: (1) What are the rate differences in FPs and FNs for HCP participants during resting state imaging days, (2) what are the rate differences in FPs and FNs for unconscious and conscious patients, and (3) what are the rate differences in FPs and FNs for conscious acute and chronic patients?

## 2. METHODS

Our preprocessing remained consistent with our previous study of the single patient (for full details see Dell'Italia et al., 2018). The diffusion data were preprocessed using the following pipeline: DWI preprocessing (brain extraction and bias field correction), registrations (i.e., linear registration to patient's anatomical and non-linear registration to the MNI template using Advanced Normalization Tools; Avants et al., 2008, 2011), probabilistic tractography (i.e., FSL's probtrackx2; Behrens T.

et al., 2003; Behrens T. E. et al., 2003) with tractography thresholding (i.e., MANIA; Shadi et al., 2016). The putative preprocessing steps for FC were performed including: brain extraction, slice timing correction, motion correction, band-pass filtering ( $0.08 \leq \text{Hz} \leq 0.1$ ), removal of linear and quadratic trends, nuisance regression for signals of non-interest (e.g., motion parameters, white matter, cerebral spinal fluid, and full-brain mean signal), and affine registration of functional data to MNI template (Advanced Normalization Tools; Avants et al., 2008, 2011). The brain network construction remained unchanged using 154 ROIs spanning the cortex, sub-cortical nuclei, cerebellum and brainstem. This parcellation scheme, which was defined independently of our data, is made freely available by Craddock et al. (2012). Additionally, we used the Oxford thalamic connectivity atlas (Behrens T. et al., 2003) to further refine the parcellation of the thalamus from 6 to 14 and we parcellated the basal ganglia into 6 ROIs (caudate, putamen, and globus pallidus each in the left and right hemispheres) for a total of 154 ROI (i.e., 134 Craddock ROIs, 6 basal ganglia ROIs, and 14 Thalamic ROIs). All the HCP data were downloaded with the preprocessing completed using the minimally processed pipeline (Glasser et al., 2013; Van Essen et al., 2013). Functional connectivity was assessed with a partial correlation method using the Markov Network Toolbox (MoNeT; Narayan et al., 2015) in MATLAB. MoNeT is a tool which combines a penalized maximum likelihood estimation with a resampling-based model selection procedure in order to find the most stable level of sparse brain graph given a set of time-dependent measurements. Each fMRI time series is bootstrapped and resampled in order to estimate the within-subject variability and a random penalization is applied iteratively to find the most stable solution. This method attempts to reduce the spurious connections that occur from indirect sources, which plague any method using Pearson's R method. Thus, each patient or participant has their own sparse connectivity from direct sources. This estimation procedure was used on both the HCP participants and the patients with DOC. While there are differences in signal quality and exact preprocessing steps between the HCP participants and our patient cohort, our comparisons involved only within HCP participants (across imaging sessions) and within patients (across imaging sessions). Any biases from increased signal quality or differences in preprocessing steps should be controlled for by comparing the data to other data with equal parameters (i.e., within HCP participants and within patients with DOC).

### 2.1. Human Connectome Project Participants

The data for this analysis was taken from the HCP, which is a public repository of high quality structural and functional MR data in a large set of healthy volunteers. For the purposes of this study, we selected a subset of the data ( $N = 12$ ) so to match the characteristics (i.e., age and gender) of the the final set of patients. These HCP participants were randomly sampled from the S1200 ( $n = 9$ ) release, the Q3 release ( $n = 1$ ), S500 ( $n = 1$ ), and S900 ( $n = 1$ ) releases to best match the age and gender of the patients at the time of their injury (see **Table 1** for the patients that were

matched and see **Table 2** for their demographics). These HCP participants only had a single DWI imaging session, which we will use for both days of resting state data.

### 2.1.1. HCP Experimental Design

From the HCP dataset, we made use of anatomical (T1-weighted), diffusion (Diffusion Tensor Imaging; DTI), and functional (T2\*-weighted) data. T1-weighted images were acquired with a 3D MPRAGE sequence (repetition time [TR] = 2,400 ms, echo time [TE] = 2.14 ms, flip angle [FA] = 8 deg). DTI images were acquired with a spin-echo echo planar sequence (TR = 5,520 ms, TE = 89.5 ms, FA = 78 deg/160 deg, 96 directions). Finally, blood oxygenation level dependent (BOLD) functional image were acquired with a gradient-echo echo planar image (TR = 720 ms; TE = 33.1 ms; FA = 52 deg).

### 2.1.2. Functional Connectivity Patient Cohort

Of the original 16 patients, 4 patients were excluded due to BOLD artifacts (patient P083), preprocessing errors (patient P086 and P097), or registration errors (patient P100). After these final exclusions there were 12 patients (see **Table 1**), which seven of these patients were male and five were female. All of these patients were presented with a post-resuscitation GCS during the acute stage of TBI which was transformed into an inferred GOS-E (Crone et al., 2018). Additionally, the GOS-E was assessed at the chronic stage of TBI. Together, the inferred GOS-E and chronic GOS-E were used to split 7 patients in the unconscious group and 17 patients in conscious group (i.e., 5 acute patients and 12 chronic patients).

### 2.1.3. Patients' Experimental Design

The 16 patients underwent two imaging sessions over the span of at least 158 days to at most 222 days. The first session occurred at most 37 days post injury (see **Table 2**), and the follow-up session took place 238 days post-injury. At each session the patient underwent (among other clinical and research sequences) anatomical (T1-weighted) and functional (T2\*-weighted) data protocols. T1-weighted images were acquired with a 3D MPRAGE sequence (repetition time [TR] = 1900 ms, echo time [TE] = 3.43,  $1 \times 1 \times 1$  mm). BOLD functional data were acquired with a gradient-echo echo planar image (see **Table 1** for the slice thickness, TR count, TE, and TR). Diffusion weighted data were acquired with an echo planar sequence (for number of gradient directions, TR, TE, and slice thickness see **Table 3**) using a *b*-value of 1,000 and acquiring an additional B0 image. Acute data were acquired on the in-patient 3 Tesla Siemens TimTrio system at the Ronald Reagan University Medical Center for patients P003, P005, P007, P014, P018, P021, P023, P024, P026, P027, P029, P039, and P066, and rest of the patients' acute data were acquired on a 3 Tesla Siemens Prisma system. All the chronic data were acquired on the out-patient 3 Tesla Siemens Prisma system also at the Ronald Reagan Medical Center at the University of California Los Angeles. The study was approved by the UCLA institutional review board (IRB). Informed consent was obtained from the legal surrogate, as per state regulations.

## 2.2. ERGM and Graph Statistics

The core idea underlying ERGM is the estimation of possible network statistics that generate a family of graphs. Edges are treated as a random variable generated by a stochastic process that could have been sampled for a number of possible graphs, which are produced from similar generative processes. A logistic regression with multiple predictor variables is used to estimate the unique contribution of each network statistic defined by:

$$P_{\theta}(Y = y) = \frac{\exp(\theta^T g(y))}{c(\theta)} \quad (1)$$

where  $\theta$  is a parameter vector that is modeled by  $g(y)$  (i.e., vector of graph statistics used in each model). The parameter  $c(\theta)$  is a normalizing constant representing the parameter estimate for all possible graphs (Hunter et al., 2008). This normalizing constant is not able to be analytically solved in most models, instead a Markov Chain Monte Carlo (MCMC) methods is used to sample and estimate the population mean. These methods assume Markovian principles of independent draws and the ability to reach equilibrium. Equilibrium is the state in which any edge that is toggled on or off results in an equally probable graph.

For both the HCP participants' and patients' datasets, we ran 4 ERGM: complete model (i.e., edges term, nodemix for resting state connectivity, the SC metrics, and triadic closure terms), structural model (i.e., edges term, nodemix for resting state connectivity, and the SC metrics), geometrically weighted edged shared partners (GWESP) model (i.e., all terms except SC metrics), and base model (i.e., only the edges term and nodemix for resting state connectivity). The base model was specified as follows, where  $P_{\theta}(Y=y)$ :

$$= \frac{\exp(\theta_1 \text{edges} + \theta_2 \text{nodemix}(\text{rest}))}{c(\theta)} \quad (2)$$

The edges term is used to control for the overall density of each graph allowing for each graph to have a different density. The nodemix(rest) term (i.e., the selective mixing for resting state networks and subcortical regions) creates multiple terms for each possible inter- (e.g., inter-frontoparietal and default network connectivity), intra-resting state network connectivity (e.g., within default network connectivity), inter-subcortical (e.g., thalamo-basal ganglia connectivity), intra-subcortical (e.g., within basal ganglia connectivity), and between subcortical and resting state connectivity (e.g., thalamo-frontoparietal network connectivity).

The structural model includes structural connectivity covariates estimated from the structural connectivity adjacency matrix, where  $P_{\theta}(Y=y)$ :

$$= \frac{\exp(\theta_1 \text{edges} + \theta_2 \text{nodecov}(\text{degree}) + \theta_3 \text{nodecov}(\text{efficiency}) + \theta_4 \text{nodematch}(\text{latent}) + \theta_5 \text{nodecov}(\text{cluster}) + \theta_6 \text{nodemix}(\text{rest}))}{c(\theta)} \quad (3)$$

The structural model has the same terms (i.e., edges and nodemix) as the base model with three additional nodal covariates (i.e., degree, local efficiency, and nodal



**TABLE 1** | Patients' functional MRI parameters.

	Matched MRI	Acute TR count	Chronic TR count	Acute slice thickness	Chronic slice thickness	Acute TR	Chronic TR	Acute TE	Chronic TE
P054	Yes	200	200	3.4375 × 3.4375 × 3 mm, 40 Slices	3.4375 × 3.4375 × 3 mm, 40 Slices	3,000 ms	3,000 ms	25 ms	25 ms
P055	Yes	200	200	3.5 mm isotropic, 52 Slices	3.4375 × 3.4375 × 3 mm, 64 Slices	3,000 ms	3,000 ms	25 ms	25 ms
P066	No	200	200	3.4375 × 3.4375 × 3 mm, 64 Slices	3.4375 × 3.4375 × 3 mm, 50 Slices	3,000 ms	3,000 ms	25 ms	25 ms
P069	Yes	200	200	3.4375 × 3.4375 × 3 mm, 50 Slices	3.4375 × 3.4375 × 3 mm, 52 Slices	3,000 ms	3,140 ms	25 ms	25 ms
P074	Yes	200	200	3.4375 × 3.4375 × 3 mm, 50 Slices	3.4375 × 3.4375 × 3 mm, 50 Slices	3,000 ms	3,000 ms	25 ms	25 ms
P079	Yes	200	200	3.4375 × 3.4375 × 3 mm, 50 Slices	3.4375 × 3.4375 × 3 mm, 50 Slices	3,000 ms	3,000 ms	25 ms	25 ms
P084	Yes	200	200	3.4375 × 3.4375 × 3 mm, 50 Slices	3.4375 × 3.4375 × 4.25 mm, 37 Slices	3,000 ms	2,006 ms	25 ms	25 ms
P085	Yes	200	200	3.4375 × 3.4375 × 3 mm, 40 Slices	3.4375 × 3.4375 × 4.25 mm, 37 Slices	3,000 ms	2,006 ms	25 ms	25 ms
P089	Yes	200	200	3.00 × 3.00 × 3.99, 33 Slices	3.4375 × 3.4375 × 4.25 mm, 37 Slices	2,000 ms	2,006 ms	30 ms	25 ms
P092	Yes	300	300	3.4375 × 3.4375 × 4.25 mm, 37 Slices	3.4375 × 3.4375 × 4.25 mm, 37 Slices	2,000 ms	2,006 ms	25 ms	25 ms
P096	Yes	300	300	3.4375 × 3.4375 × 4.25 mm, 37 Slices	3.4375 × 3.4375 × 4.25 mm, 37 Slices	2,000 ms	2,006 ms	25 ms	25 ms
P099	Yes	300	300	3.4375 × 3.4375 × 4.25 mm, 37 Slices	3.4375 × 3.4375 × 4.25 mm, 37 Slices	2,000 ms	2,006 ms	25 ms	25 ms

The functional MRI parameters are tabulated for each patient. These parameters' descriptions are the same as the DWI parameters's descriptions, except they are for the functional MRI imaging session. An additional parameter for the number to TRs are tabulated under TR count.

**TABLE 2** | Patients' demographics.

	Gender	Cause of injury	Age at injury	Acute TSI	Chronic TSI	Delta TSI	Acute GCS	Acute GOS-E	Chronic GOS-E	Group
P054	Female	Fall	36+	25	186	161	7 (E:2, V:1, M:4)	2	8	Unconscious
P055	Male	Fall	36+	1	195	194	6 (E:1, V:1, M:4)	2	4	Unconscious
P066	Male	Bicycle vs. Automobile	36+	0	222	222	7 (E:2, V:1, M:4)	2	5	Unconscious
P069	Male	Fall	36+	1	186	185	7 (E:2, V:1, M:4)	2	8	Unconscious
P074	Female	Automobile accident	18–25	1	180	179	8 (E:1, V:2, M:5)	3	8	Conscious
P079	Female	Pedestrian vs. Automobile	18–25	2	181	179	7 (E:3, V:3, M:1)	2	7	Unconscious
P083	Male	Pedestrian vs. Automobile	18–25	1	180	179	3 (E:1, V:1, M:1)	2	8	Unconscious
P084	Female	Automobile accident	18–25	5	183	178	3 (E:1, V:1, M:1)	2	3	Unconscious
P085	Female	Fall	36+	2	177	175	15 (E:4, V:5, M:6)	3	7	Conscious
P086	Male	Fall	36+	26	221	195	6 (E:1, V:1, M:4)	3	5	Conscious
P089	Male	Fall	36+	17	181	164	10 (E:2, V:1, M:5)	3	3	Conscious
P092	Male	Automobile accident	18–25	1	158	157	10 (E:4, V:1, M:5)	2	4	Unconscious
P096	Male	Automobile accident	26–30	37	176	139	3 (E:1, V:1, M:1)	3	4	Conscious
P097	Male	Fall	18–25	17	170	153	8 (E:2, V:1, M:5)	3	6	Conscious
P099	Male	Bicycle vs. Automobile	18–25	18	184	166	8 (E:2, V:1, M:5)	3	8	Conscious
P100	Male	Fall	36+	4	173	169	8 (E:2, V:1, M:5)	2	5	Unconscious

For the 16 patients, the following demographics for each patient is tabulated: gender, cause of injury, age at injury, time since injury for the acute imaging session (Acute TSI), time since injury for the chronic imaging session (Chronic TSI), and the difference in time between the two imaging sessions (Delta TSI). Additionally for each patient, the level of consciousness at the acute imaging session (Acute GCS) and chronic imaging session (Chronic GOS-E) are tabulated. The acute GOS-E is inferred from the GCS scores after the acute imaging session. Finally, these GOS-E scores are used to group patients into two different recovery groups: unconscious and conscious groups.

clustering coefficient). These nodal covariates incorporate structural connectivity differences into the functional connectivity modeling.

The GWESP model has the same terms (i.e., edges and nodemix) as the base model with a term for triadic closure, where  $P_{\theta}(Y=y)$ :

$$= \frac{\exp(\theta_1 \text{edges} + \theta_2 \text{nodemix}(\text{rest}) + \theta_3 \text{gwesp}(\alpha = \lambda))}{c(\theta)} \quad (4)$$

The GWESP term was added to the edges and nodemix(rest) terms to account for the triadic closure. This term models the number of edge shared partners in each graph, but it applies a geometrically weighted distribution to penalize higher counts of edge shared partners (see Hunter, 2007, for a complete description)

The complete model contains both the structural terms and the GWESP term, where  $P_{\theta}(Y = y)$ :

**TABLE 3 |** Patients' DWI parameters.

	Matched MRI	Acute Bvec	Chronic Bvec	Acute slice thickness	Chronic slice thickness	Acute TR	Chronic TR	Acute TE	Chronic TE
P054	Yes	64	63	2 mm isotropic, 69 Slices	2.125 × 2.125 × 2 mm, 69 Slices	9,000 ms	9,000 ms	90 ms	90 ms
P055	Yes	56	56	2 × 2 × 3 mm, 50 Slices	2.125 × 2.125 × 2 mm, 81 Slices	8,000 ms	9,000 ms	96 ms	90 ms
P066	No	41	38	2 mm isotropic, 81 Slices	2 mm isotropic, 69 Slices	9,300 ms	9,000 ms	90 ms	90 ms
P069	Yes	59	61	2 mm isotropic, 69 Slices	2.125 × 2.125 × 2 mm, 77 Slices	9,000 ms	9,900 ms	90 ms	90 ms
P074	Yes	58	62	2 mm isotropic, 69 Slices	2.125 × 2.125 × 2 mm, 69 Slices	9,000 ms	9,000 ms	90 ms	90 ms
P079	Yes	64	62	2 mm isotropic, 69 Slices	2.125 × 2.125 × 2 mm, 69 Slices	9,000 ms	9,000 ms	93 ms	90 ms
P084	Yes	64	62	2 mm isotropic, 69 Slices	2 × 2 × 3 mm, 52 Slices	9,000 ms	9,000 ms	90 ms	90 ms
P085	Yes	54	62	2 mm isotropic, 69 Slices	2 × 2 × 3 mm, 52 Slices	10,100 ms	9,000 ms	90 ms	90 ms
P089	Yes	60	62	2 × 2 × 3 mm, 50 Slices	2 mm isotropic, 72 Slices	9,500 ms	9,500 ms	90 ms	90 ms
P092	Yes	61	63	2 mm isotropic, 78 Slices	2 mm isotropic, 75 Slices	10,100 ms	9,500 ms	90 ms	90 ms
P096	Yes	45	60	2 mm isotropic, 78 Slices	2 mm isotropic, 78 Slices	10,100 ms	10,100 ms	90 ms	90 ms
P099	Yes	62	64	2 mm isotropic, 78 Slices	2 mm isotropic, 78 Slices	10,100 ms	10,100 ms	90 ms	90 ms

The following parameters for each DWI imaging session (i.e., acute and chronic) varied from patient to patient due to clinical requirements: the number of gradient directions (Bvec), the slice thickness, the repetition time (TR), and the echo times (TE). Additionally, the matched MRI indicates which patients had the same MRI system in both the acute and chronic imaging sessions.

$$= \frac{\exp(\theta_1 \text{edges} + \theta_2 \text{nodecov}(\text{degree}) + \theta_3 \text{nodecov}(\text{efficiency}) + \theta_4 \text{nodematch}(\text{latent}) + \theta_5 \text{nodecov}(\text{cluster}) + \theta_6 \text{nodemix}(\text{rest}) + \theta_7 \text{gwesp}(\alpha = \lambda))}{c(\theta)} \quad (5)$$

These four models will be used to estimate rate differences using a multinomial regression (see section 2.3). Each model was assessed for its overall fit to the observed data. Due to the large number of total ERGM conducted (192 in total across patients and HCP participants), we will only compare for two patients (i.e., P092 in acute stage TBI and P085 in chronic stage TBI) and HCP patients (i.e., HCP002 from the first and HCP008 resting state imaging sessions) for the complete and GWESP models assessed by using goodness of fit (GOF) plots (Hunter et al., 2008). After the model was estimated, a thousand simulations were run from the model statistics that generated 1,000 separate graphs with the generative processes captured from each of the four models. These simulations provided a distribution of mixing terms (see **Table 4** for patients, and for HCP participants, see **Table 5**) and edge shared partners (see **Table 6** for patients) that were compared to the original graph's edge probabilities. The mean and variance of these distribution were used to test for differences between the simulated distributions and the original data's distribution. This is to ensure that the model represents a graph similar to the original data that it was modeled from. We will assess the overall model statistics from equation 5 and edge shared partner distributions.

## 2.3. Multinomial Regressions

We compared these 4 ERGMs in three combinations: base model to structural model (i.e., structural effect), base model to GWESP model (i.e., GWESP effect), and base model to complete model (i.e., interaction effect). The first comparison was to isolate the effects of leaving out structural terms discussed in problem #3. The second comparison was to isolate the effects of leaving out a term that accounts for triadic closure (i.e., the GWESP

term) discussed in problem #2. Finally, the third comparison demonstrates the effects of leaving out the structural terms while still accounting for triadic closure (i.e., a combination of problem #2 and #3). We labeled one model as the full model and one as the partial model in each comparison.

To compare the affects of not accounting for specific terms, we tallied the change in PEs when the terms were omitted. If a PE was significant in the full model (i.e., the model with more terms for that specific comparison), but not the partial model (i.e., the base model), we label this as a FN. FP was a PE that was significant in partial model, but not the full model. The true positives (TPs) and true negatives (TNs) are the terms that are significant or non-significant—respectively, in both models. We group the PEs based on whether they belonged to the cortical regions or subcortical to see if within cortical connectivity was affected, within subcortical connectivity, or between cortical to subcortical connectivity.

These tallies of FP and FN were compared for differences between patients grouped based on their level of consciousness at each imaging session using a behavior assessment. During the acute session, patients were evaluated with a post-resuscitation Glasgow Coma Scale (GCS; Teasdale and Jennett, 1974). The GCS has three subscales: eyes opening (E), verbal response (V), and motor response (M). Crone et al. (2018) used the GCS subscales of V and M to transform the GCS scores into the Glasgow Outcome Scale-Extended (GOS-E; Wilson et al., 1998). A patient with a GCS V subscale of less than or equal 3 and a GCS M subscale of less than or equal to 4 were assigned an inferred GOS-E of 2, while a patient with higher scores for GCS V and M were assigned an inferred GOS-E of 3. While DoC diagnoses are typically not made at such an acute stage, a patient with a

**TABLE 4 |** Goodness of fit differences for graph statistics.

	Complete model					GWESP model			
	Observed	min	M	max	p-value	min	M	max	p-value
<i>HCP001 Rest1</i>									
Inter-frontoparietal subcortical	11	5	10	19	1	3	8.600	14	0.200
Inter-default visual	82	65	78.700	90	0.800	65	77.400	88	0.200
Inter-default limbic	69	57	64.800	70	0.200	56	68	80	1
Inter-subcortical thalamus	14	13	16.500	20	0.200	3	13	21	1
Inter-default ventral attention	79	75	86.100	100	0.200	62	78.300	86	1
<i>HCP002 Rest1</i>									
Edges	2,044	2,023	2,059.600	2,113	1	1,901	1,983.800	2,064	0.200
Inter-limbic thalamus	20	15	20.500	26	1	9	14.900	22	0.200
Within subcortical	159	144	161.800	173	0.600	112	143.400	169	0.200
Gwesp Fixed, $\lambda=0.6$	3,634.600	3,593.262	3,660.370	3,756.905	1	3,367.806	3,521.756	3,670.710	0.200
<i>HCP003 Rest1</i>									
Inter-dorsal attention dorsa attention	7	6	8.500	12	0.200	3	6.700	11	1
Inter-subcortical visual	10	6	8.300	14	0.200	4	10.300	18	0.800
<i>HCP004 Rest1</i>									
Within thalamus	23	16	25.900	33	0.400	23	32.200	39	0.200
Within subcortical	138	118	144.300	160	0.400	140	160.600	169	0
<i>HCP005 Rest2</i>									
Inter-frontoparietal somatomotor	45	29	38.800	45	0.200	30	43.300	50	1
Within subcortical	174	156	184.100	200	0.200	148	167.400	176	0.600
<i>HCP006 Rest2</i>									
Inter-frontoparietal somatomotor	45	29	38.800	45	0.200	30	43.300	50	1
Within subcortical	174	156	184.100	200	0.200	148	167.400	176	0.600
<i>HCP008 Rest2</i>									
Inter-limbic thalamus	26	19	23	30	0.400	16	20.800	29	0.200
Inter-somatomotor thalamus	75	64	73.800	87	0.800	60	67.900	76	0.200
Inter-limbic visual	28	15	27.100	37	0.800	18	21.900	27	0
<i>HCP009 Rest1</i>									
Inter-limbic subcortical	16	14	17.400	21	1	17	19.200	24	0
Inter-frontoparietal ventral attention	46	35	46.700	52	0.800	45	49.900	57	0.200
Inter-dorsal attention visual	28	21	28.200	38	1	20	24.200	34	0.200
<i>HCP010 Rest1</i>									
Inter-basal ganglia default	33	24	31.700	44	0.400	25	29.200	32	0
Inter-default dorsal attention	41	29	40.100	53	0.800	36	44.400	47	0.200
<i>HCP010 Rest2</i>									
Inter-basal ganglia limbic	13	7	11.100	18	0.200	8	12.900	21	1
<i>HCP012 Rest2</i>									
Inter-basal ganglia ventral attention	18	17	20.800	30	0.200	12	16.200	24	0.600

The observed column is the original data's values for each graph statistic, while the minimum, mean, maximum, and p-value for the simulated graphs based on each of the ERGM models are displayed. They are the biggest difference between the complete model and the GWESP model. Overall, all the patients' ERGM for the complete model and GWESP effect fit the data well based on the graph statistics modeled but the GWESP model had 4  $p < 0.05$ .

GOS-E of 2 is consistent with a VS, and patient with a GOS-E of 3 has recovered from VS. This allows organizing patients into two groups (see **Table 2**): unconscious patients vs. conscious patients. To mirror this in the HCP datasets, we compared the first resting state imaging to the second resting state imaging session.

Using the nnet package Venables and Ripley (2013) in R, we used the mlogit function to multinomial regressions to predict

the differences between unconscious patients' FP and FN rates for the cortical groupings (i.e., within cortical connectivity, within subcortical connectivity, and cortico-subcortical connectivity). These cortical groupings are for the mixing of the resting state networks and subcortical structures. The no error for all grouping was the reference group for the outcome variable and the acute conscious patients were the reference group for the predictor variable. For the cortical group there were

**TABLE 5 |** Goodness of fit differences for graph statistics.

	Complete model					GWESP model			
	Observed	min	<i>M</i>	max	<i>p</i> -value	min	<i>M</i>	max	<i>p</i> -value
<i>P054 Acute</i>									
Edges	1,796	1,702	1,780.300	1,839	0.800	1,793	1,839.300	1,886	0.200
Intra-limbic	19	15	21.300	28	0.400	18	22.100	28	0.200
Inter-somatomotor-ventral attention	56	50	57.700	69	1	54	63.300	74	0.200
GWESP (fixed, $\lambda = 0.45$ )	2,766.980	2,614.125	2,741.347	2,834.899	0.800	2,761.997	2,835.273	2,908.383	0.200
<i>P084 Acute</i>									
Inter-frontoparietal-somatomotor	40	28	39	52	1	6	10.400	16	0.200
Inter-frontoparietal-thalamus	16	13	18.250	28	0.700	13	18.200	22	0.200
<i>P084 Chronic</i>									
Inter-dorsal attention-limbic	9	9	11.300	14	0.200	2	8	12	1
Within subcortical	153	151	159	172	0.200	127	144.300	161	0.600
<i>P085 Acute</i>									
Inter-dorsal attention-ventral attention	12	6	9.700	15	0.200	10	15	21	0.800

The observed column is the original data's values for each graph statistic, while the minimum, mean, maximum, and *p*-value for the simulated graphs based on each of the ERGM models are displayed. None of these are bad fits, but they are the biggest difference between the complete model and the GWESP model. Overall, all the patients' ERGM for the complete model and GWESP effect fit the data well based on the graph statistics modeled because neither model produced any  $p < 0.05$  for any graph statistic.

**TABLE 6 |** Goodness of fit differences for edge shared partners.

	Complete model P054 acute					GWESP model P054 acute					Complete model P084 chronic					GWESP model P084 chronic				
	Observed	min	<i>M</i>	max	<i>p</i> -value	min	<i>M</i>	max	<i>p</i> -value	Observed	min	<i>M</i>	max	<i>p</i> -value	min	<i>M</i>	max	<i>p</i> -value		
esp0	4	0	0.100	1	0	0	0	0	0	16	0	0.700	2	0	0	1.600	4	0		
esp1	20	1	3.200	6	0	1	3.200	6	0	63	42	56.800	73	0.800	50	58.500	67	0.400		
esp2	58	50	65	83	0.600	54	74	89	0.400	140	198	231.700	254	0	182	226.200	258	0		
esp3	140	226	245.900	269	0	224	252.800	292	0	243	311	341.300	371	0	326	347.500	391	0		
esp4	213	357	385	410	0	326	359.100	383	0	275	311	335.200	365	0	316	332.900	353	0		
esp5	255	362	384	406	0	360	380.300	401	0	260	200	237.600	277	0.400	217	240.600	267	0.200		
esp6	275	276	311.200	351	0	261	292.500	326	0.600	201	113	142.900	167	0	124	149.600	174	0		
esp7	245	176	205	231	0	170	198.300	231	0	143	51	76.700	95	0	67	80.100	94	0		
esp8	197	89	116.100	147	0	81	110.600	135	0	72	26	35.600	47	0	20	35.700	46	0		
esp9	158	54	64.200	78	0	40	60.900	79	0	40	7	14.600	26	0	11	14.700	25	0		
esp10	93	26	34.800	45	0	17	28.500	41	0	17	3	7	15	0	3	5.400	9	0		
esp11	68	6	14.600	22	0	9	12.900	22	0	12	0	2.500	6	0	0	1.800	3	0		
esp12	28	4	6	9	0	1	4.300	7	0	6	0	1	3	0	0	1.100	4	0		
esp13	24	1	2.700	5	0	0	2.200	4	0	1	0	0.500	1	1	0	0.100	1	0.200		
esp14	11	0	0.900	3	0	0	0.600	3	0	0	0	0	0	1	0	0	0	1		

The observed column is the original data's values for each edge shared partner type, while the minimum, mean, maximum, and *p*-value for the simulated graphs based on each of the ERGM models are displayed. The edge shared partner types are based on the number of triangles sharing a common edge (e.g., the esp10 term has 10 triangles all sharing common edge). Overall, all the patients' ERGM for the complete model and GWESP effect did not fit the data well based on the graph statistics modeled because both models had at least than 11 of the 14  $p < 0.05$  for types of edge shared partner type.

6 possible categories predicted, which were FN and FP for each grouping. The same comparisons were conducted for the HCP participants, but there was only one possible comparison between the first and second resting state scan. Finally, we transformed all the logits into odds ratios for reporting and interpretations.

### 3. RESULTS

As shown in **Figure 1**, the brain network construction using MoNeT resulted in different estimated densities. Overall, the density varied between resting state session 1 and session 2 in all participants (except HCP011) within the range between



**TABLE 7 |** The effect of daily variability in resting state connectivity on FPs and FNs.

Comparison:	Multinomial Regression: Cortical Nodal Labeling					
	Interaction effect		GWESP effect		Structural effect	
	Constant	Rest2	Constant	Rest2	Constant	Rest2
False negatives for Cortical to Subcortical	0.00597*** (0.355)	1.49 (0.458)			0.00614*** (0.355)	1.97 (0.435)
False positives for Cortical to Subcortical	0.0269*** (0.169)	0.716 (0.260)	0.039*** (0.141)	0.681 (0.220)	0.0338*** (0.153)	0.628 (0.245)
False negatives for Within Cortical	0.0157*** (0.220)	0.614 (0.355)	0.001*** (0.708)	0.984 (1.001)	0.937*** (0.198)	0.797 (0.296)
False positives for Within Cortical	0.0299*** (0.160)	1.09 (0.222)	0.031*** (0.157)	0.937 (0.224)	0.0515*** (0.125)	0.869 (0.183)
False negatives for Within Subcortical	0.00448*** (0.409)	0.826 (0.607)	0.002*** (0.578)	0.656 (0.914)	0.00307*** (0.501)	1.73 (0.628)
False positives for Within Subcortical	0.00149*** (0.708)	0.496 (1.23)	0.002*** (0.578)	<0.0001 (<0.0001)	0.000768*** (1.00)	1.97 (1.23)
Observations	2904		2904		2904	
Log Likelihood	-1072.262		-1702.825		-1316.831	
Akaike Inf. Crit.	2168.524		1722.825		2657.662	

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ . For the cortical nodal labeling, the FPs and FNs for each type of connectivity pattern were predicted for the first resting state imaging days. The change in odds ratios and their standard errors in parentheses are listed for the interaction, GWESP and structural effect comparisons. Overall, there are no significant change in odds ratios for any of the three model comparisons.

**TABLE 8 |** The effect of level of consciousness and stage of TBI on FPs and FNs.

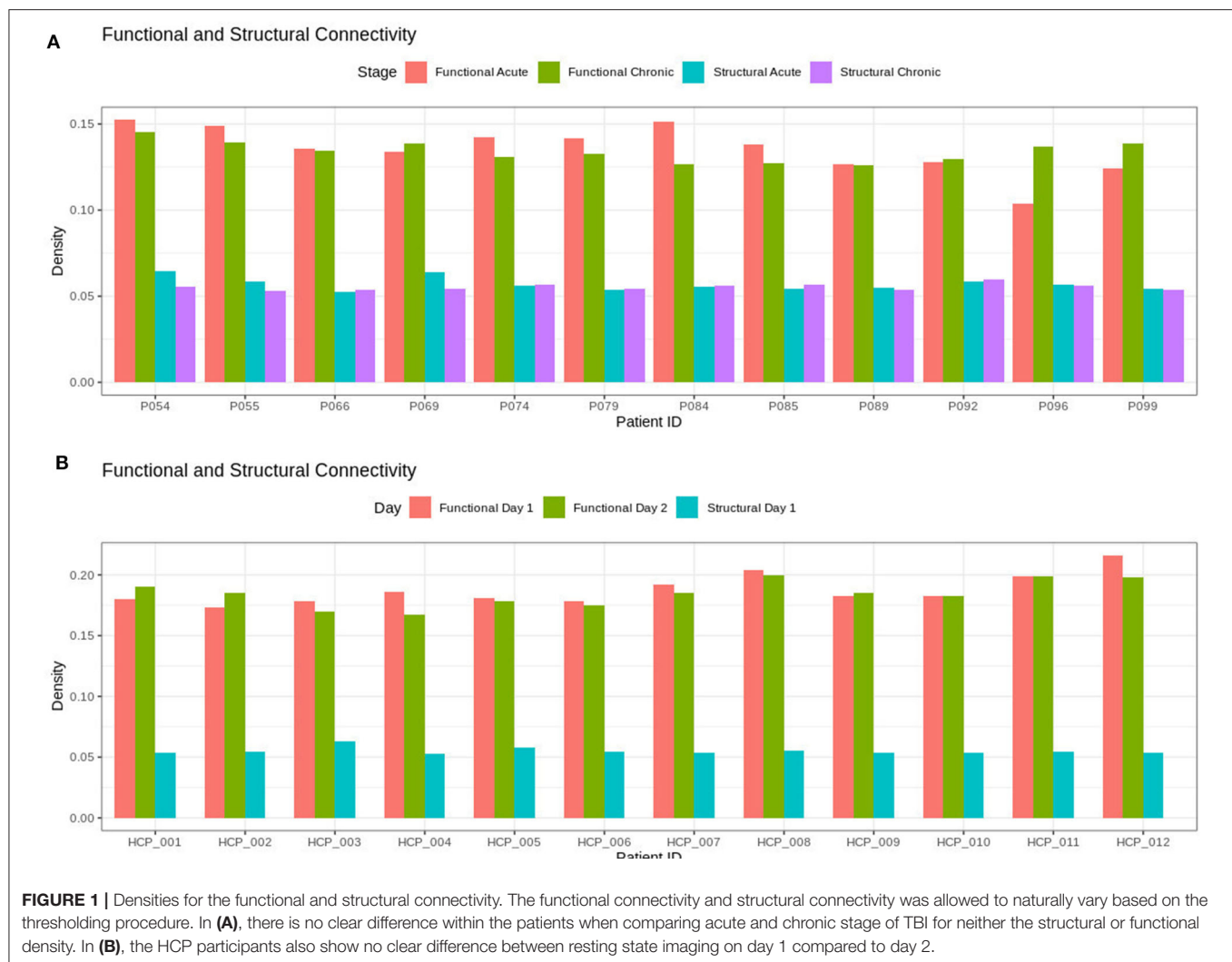
Comparison:	Multinomial Regression: Cortical Nodal Labeling								
	Interaction effect			GWESP effect			Structural effect		
	Constant	Unconscious	Chronic	Constant	Unconscious	Chronic	Constant	Unconscious	Chronic
False negatives for Cortical to Subcortical	0.0272*** (0.271)	0.277*** (0.491)	1.16 (0.317)	NA (NA)	NA (NA)	NA (NA)	0.0183*** (0.319)	0.274* (0.594)	0.401* (0.450)
False positives for Cortical to Subcortical	0.0350*** (0.240)	0.575 (0.348)	1.14 (0.281)	0.05628*** (0.188)	0.83284 (0.254)	0.94166 (0.225)	0.00733*** (0.502)	0.685 (0.709)	1.00 (0.594)
False negatives for Within Cortical	0.0447*** (0.213)	0.366*** (0.352)	1.31 (0.246)	0.01313*** (0.380)	0.69403 (0.537)	0.23741* (0.629)	0.0201*** (0.305)	1.31 (0.376)	1.09 (0.356)
False positives for Within Cortical	0.0485*** (0.205)	0.259*** (0.378)	1.18 (0.239)	0.05816*** (0.185)	0.60447 (0.269)	1.05868 (0.218)	0.0366*** (0.228)	0.274** (0.422)	0.321*** (0.339)
False negatives for Within Subcortical	0.0136*** (0.381)	0.370 (0.629)	0.551 (0.507)	NA (NA)	NA (NA)	NA (NA)	0.0256*** (0.271)	0.636 (0.389)	0.746 (0.335)
False positives for Within Subcortical	0.00583*** (0.579)	0.431* (0.915)	2.43 (0.628)	0.00750*** (0.502)	0.86763 (0.673)	0.10388* (1.12)			
Observations	2904			2904			2904		
Log Likelihood	-1749.64			-1381.539			-1024.384		
Akaike Inf. Crit.	3535.28			2799.077			2078.769		

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

For the cortical nodal labeling, the FPs and FNs for each type of connectivity pattern were predicted for unconscious patients compared to all conscious patients. The change in odds ratios and their standard errors in parentheses are listed for the interaction, GWESP and structural effect comparisons. Overall, the interaction effect has a significant decrease in odds ratios for FNs in cortical to subcortical connectivity and within cortical connectivity, and FPs in within cortical and subcortical connectivity.

0.003 to 0.0175. Across HCP participants, the densities ranged from 0.1676 to 0.2159 and the structural connectivity had less variability in the densities of the graphs across subjects ranging from 0.0531 to 0.0632. For the patients, the density

of the functional connectivity differed between resting acute session and chronic session in all patients within the range between 0.0019 to 0.0331. Across patients, the densities ranged from 0.1039 to 0.1524 and the structural connectivity had less

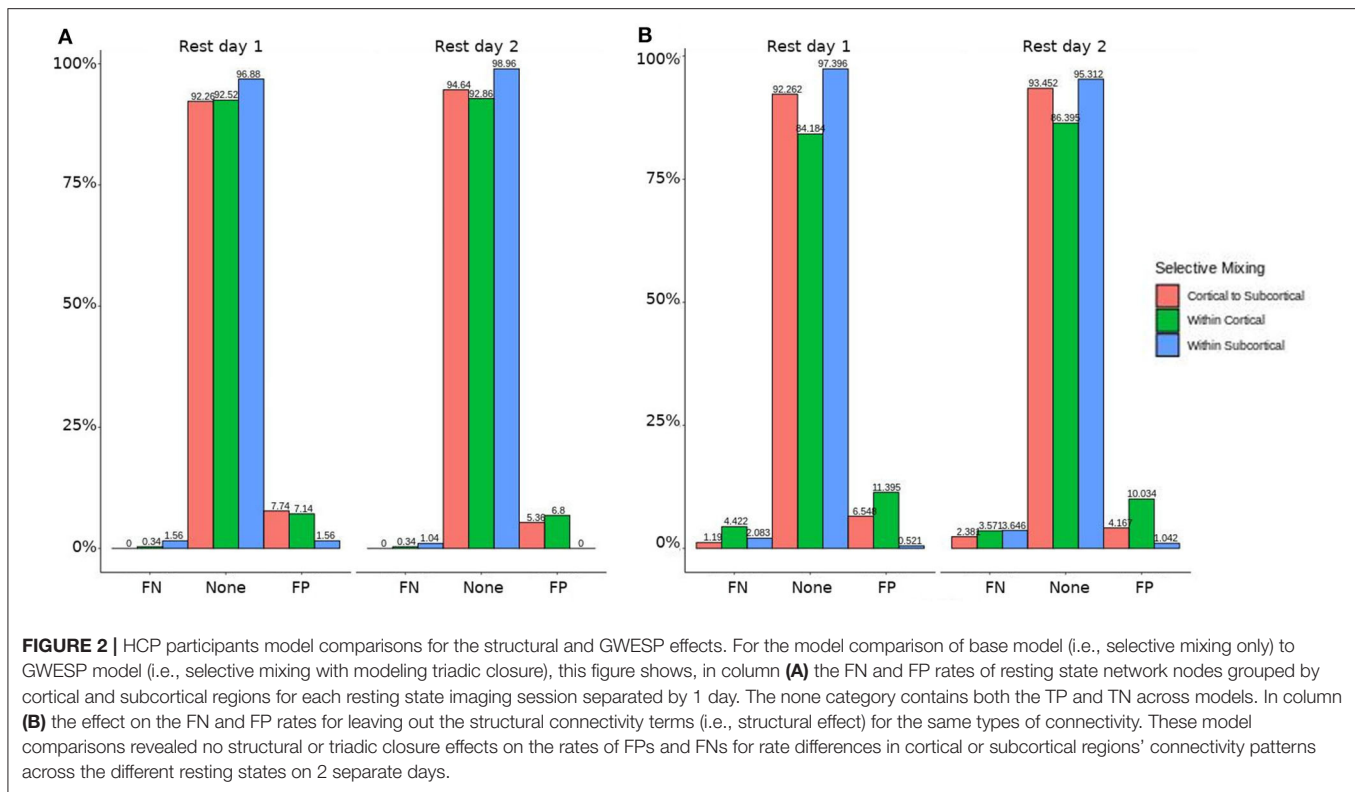


variability across acute and chronic imaging sessions in the densities (i.e., a difference between 0.005 to 0.0095).

In both HCP participants and patients, we assessed the functional connectivity types with ERGM using four models (i.e., base, structural, GWESP, and complete) to reveal the FPs and FNs that result from the omissions of the structural terms (i.e., the structural effect), triadic closure term (i.e., the GWESP effect), and both structural and triadic closure (i.e., the interaction effect). Leaving out the structural terms, the HCP participants all had either FP or FNs for cortical/subcortical connectivity (see column B in **Figure 2**). Using multinomial regression we tested for differences in FP and FN rates between their two resting state imaging, the structural effect, GWESP effect and interaction effect comparisons all revealed no significant change in odds ratios for the resting state imaging session performed on day 2 compared to day 1 (see **Table 7**). Despite these lack of differences in resting state imaging days, all patients had FP or FN rates for leaving out the triadic closure term (see column A in **Figure 2**, and for the interaction effect

of leaving out both the structural and GWESP terms (see **Figures 3, 4**).

The omission of structural terms generate FPs or FNs for all patients (see row B in **Figure 5**) did produce a 0.401 times decrease in FN for cortico-subcortical connectivity and a 0.321 times decrease for FP in within cortical connectivity for conscious chronic compared to conscious acute stage of TBI patients (see column 3, in **Table 8**). A similar pattern was found for unconscious patients compared to conscious acute patients, where there was a significant 0.274 times decrease in FN for cortico-subcortical connectivity and a 0.274 times significant decrease for FP in within cortical connectivity. For the GWESP (see row A in **Figure 5**), only the chronic patients had a significant decrease in FP for within subcortical connectivity and significant decrease in FP for within cortical connectivity compared to acute patients (see column 2, in **Table 8**). Finally, the interaction effect of leaving out GWESP and structural terms (see **Figures 6, 7**) produced FPs and FNs significant decreases in odds ratios of 0.277 times, 0.366 times, 0.259 times, and 0.431



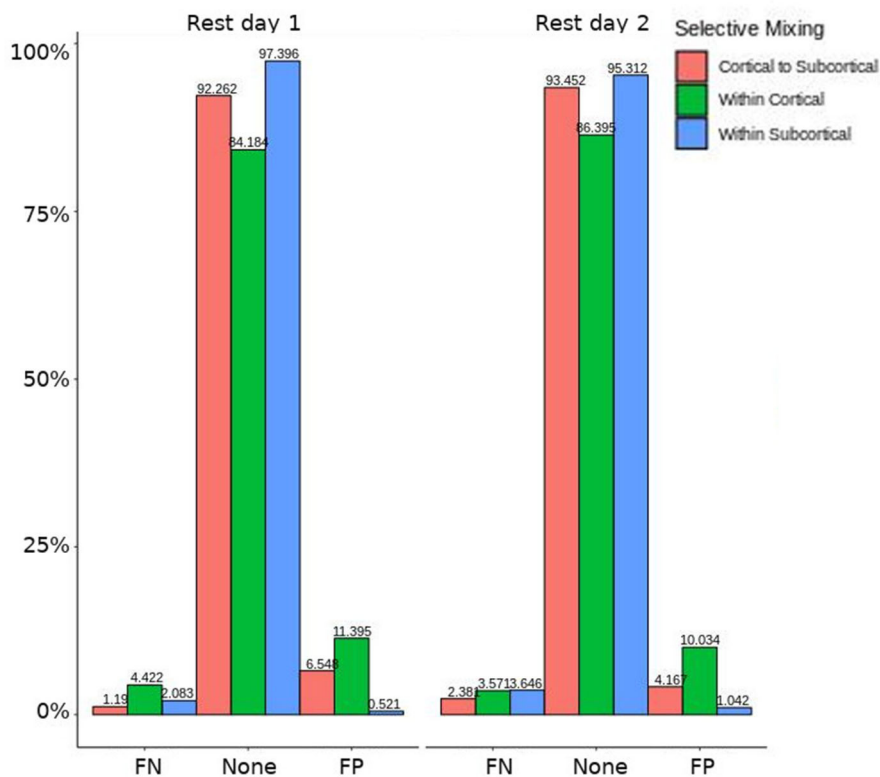
times for FNs in cortico-subcortical connectivity, FNs and FPs for within cortical, and FP for within subcortical, respectively. Overall, the patients with DOC did have differing rates FP and FN and their rates were not driven by a few conscious patients with DOC, rather there were clear high rates of FP and FNs for seven of the seventeen conscious patients (i.e., P074 chronic, P085 chronic, P089 acute, P089 chronic, P092 acute, P096 chronic, and P099 chronic; see **Figure 7**) compared to none of the seven unconscious patients having rates close to the seven conscious patients.

## 4. DISCUSSION

Overall, our results highlight two important and overlooked issues in graph analysis. First, brain graphs are susceptible to having different natural levels of density (see Milham et al., 2012; Nielsen et al., 2013) at which they are the most stable and thus, likely, better representations of the structure of the sampled graph. In our data, FC densities ranged from 16.76 to 21.5% for HCP volunteers (see **Table 2**), and from 10.39 to 15.24% in TBI patients. The use of MoNeT (Narayan et al., 2015), together with ERGM, acknowledges such inter-individual and inter-group differences and views density as a potentially interesting feature of a graph as opposed to a nuisance to be addressed by either imposing an arbitrary level (or range) of density across all graphs (Rubinov and Sporns, 2010; van Wijk et al., 2010) or by assuming fully connected, complex, networks (Rubinov and Sporns, 2011; Fornito et al., 2013, 2016). The issue is all the more important in the context of brain injury and DOC, where full-brain functional

and structural connectivity are known to vary across groups (e.g., acute vs. chronic, healthy volunteers vs. patients, coma vs. MCS, etc; Vanhaudenhuyse et al., 2010; Fernández-Espejo et al., 2012; Monti et al., 2015; Zheng et al., 2017; Crone et al., 2018).

Second, our data also show that even when graph density is allowed to vary, failure to account analytically for the interdependence of network measures (i.e., problem #2) and for the structural substrate of functional graphs (i.e., problem #3) significantly affects model estimation. To clarify this issue, it is worth noting that while the conventional approach to graph theory attempts to “summarize” observed networks by measuring their topological properties (e.g., characteristic path length, clustering, and small-worldness Rubinov and Sporns, 2010; Bullmore and Sporns, 2012), ERGM attempts to reconstruct which “social” processes, or combination thereof, are the most likely to have generated the observed networks (hence the term generative processes; Robins et al., 2007; Goodreau et al., 2009). Somewhat analogously to a multiple regression framework (Goodreau et al., 2009), ERGM can thus be viewed as a procedure to determine which combination of generative processes are most likely to explain the observed network (e.g., triadic closure, which is measured by the GWESP term, and captures the degree to which, if node A and node B both connect to node C, A and B more likely to be connected compared to any other two nodes; selective mixing, which is measured by the mixing term, and captures the degree to which nodes that are part of the same group [e.g., hemisphere] more likely to connect with each other). Thus, much like in



**FIGURE 3 |** HCP participants model comparison for the interaction effect. For the model comparison of base model (i.e., selective mixing only) to complete model (i.e., selective mixing with modeling both triadic closure and SC), this figure shows the FN and FP rates of resting state network nodes grouped by cortical and subcortical regions for each resting state imaging session separated by 1 day. The none category contains both the TP and TN across models. This comparison revealed no rate differences in cortical or subcortical regions' connectivity patterns across the different resting states on 2 separate days.

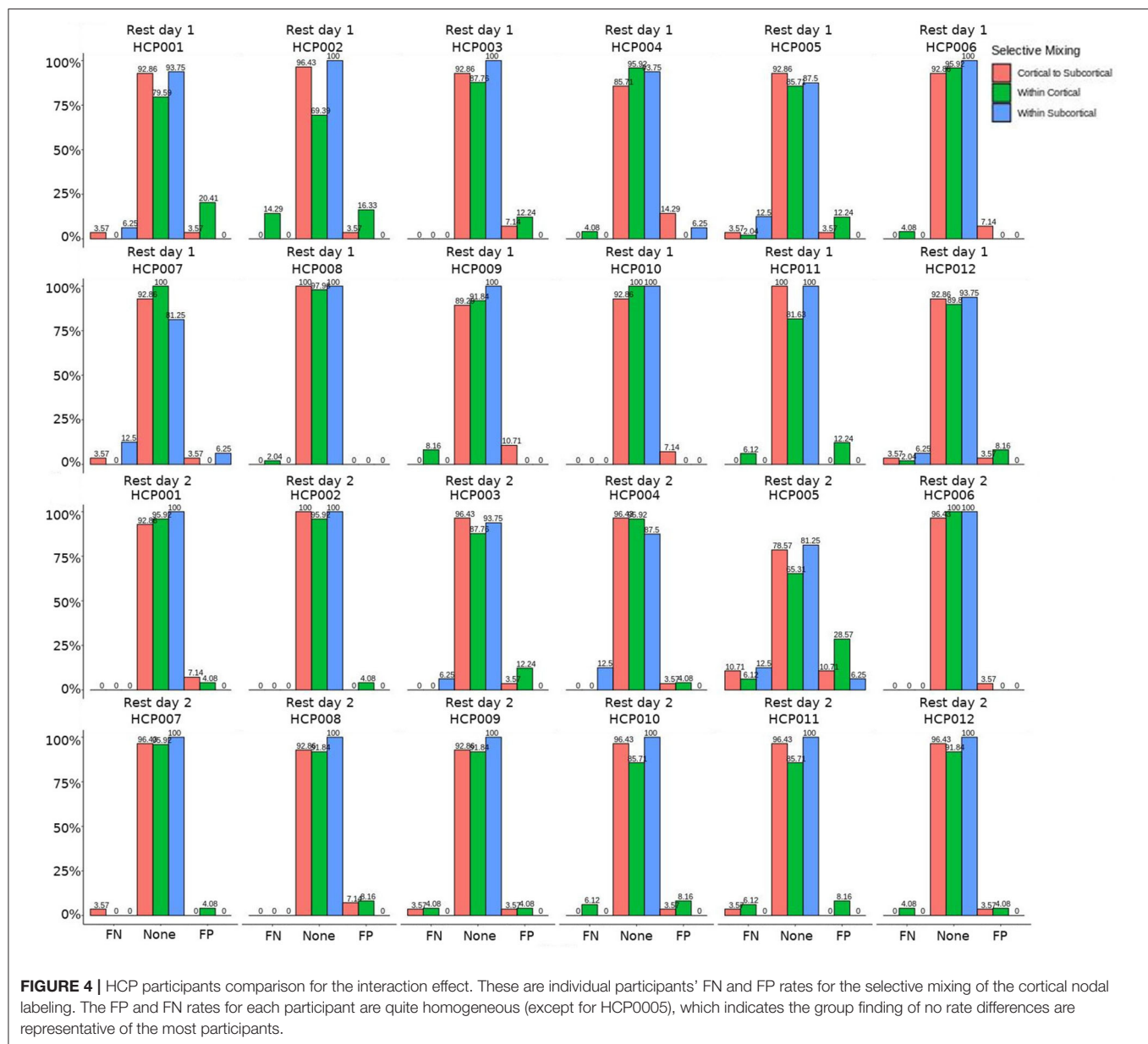
a conventional multiple regression, omission of informative explanatory variables from the model (i.e., terms) results in incorrect estimation of the association between a dependent variable (which, in ERGM, is the observed network) and each explanatory variable (in ERGM, the outcome metrics associated with the generative processes; for further discussion see Dell'Italia et al., 2018).

The importance of this issue becomes clearer when considering that many of the questions of interest in the context of DOC and TBI are questions about specific generative processes. For example, asking whether cortico-cortical or cortico-subcortical connectivity play different roles in the maintenance or recovery of consciousness (e.g., Laureys et al., 2000a,b; Boly et al., 2009, 2011; Vanhaudenhuyse et al., 2010; Crone et al., 2014, 2018; Amico et al., 2017) is, in the ERGM framework, a question of selective mixing: is level of consciousness associated with greater/smaller proclivity for nodes within cortex to associate preferentially with nodes in the same group (i.e., cortex)? In order to properly answer this question, the model needs to account (i.e., parcel out, keeping the analogy of the multiple regression) for the shared variance of selective mixing and other generative processes (e.g., triadic closure measured by number of triangles in a graph).

Indeed, the present results show that both patients and HCP participants had non-zero rates for FP and FN, if either triadic closure or the structural term were omitted. Crucial for this literature, however, conscious and unconscious patients exhibited very different rates, potentially because, as we discuss more below, of the different effect of mis-modeling on the two populations (due to their different underlying structural/functional properties).

Leaving out a known interrelated measures (e.g., GWESP term for triadic closure term; see column 2 in Table 8) increases risks for differences in rates in FN for within cortical connectivity and FP within subcortical connectivity for comparisons between acute and chronic stage TBI patients. As we discussed in the introduction, triadic closure is one the key generative processes (Goodreau et al., 2009) present most graphs, but more importantly these differences in FP and FN rates could be due to the importance of triadic closure to consciousness. Clustering coefficients have been used to differentiate patients with DOC (using structural connectivity; Tan et al., 2019) and for differentiating levels of consciousness while undergoing anesthesia (Monti et al., 2013). Due to the empirical findings of different levels of triadic closure within patient groups, the



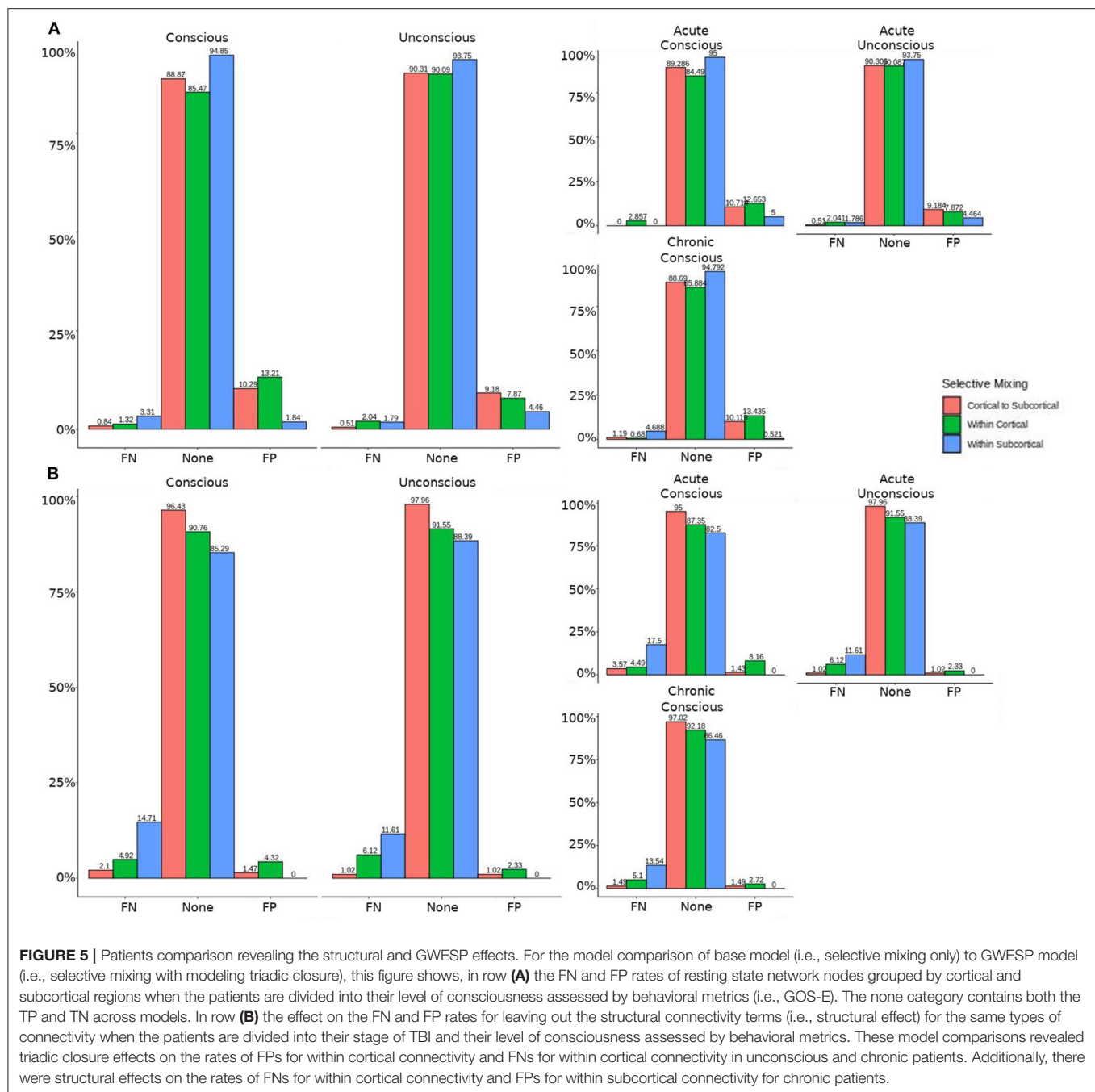


best way to avoid these FP and FN rate differences is to model the selective mixing and triad closure processes in a single model.

Similarly, omitting the structural terms (see column 3 in **Table 8**) affects the FP rates differently for the unconscious patients in their acute stage of TBI and the conscious patients in the chronic stage of their TBI for both within cortical connectivity and cortico-subcortical connectivity. These effects are quite problematic for DOC research due to the interest in comparing cortico-cortical and thalamo-cortical connectivity (e.g., Laureys et al., 2000a,b; Boly et al., 2009, 2011; Vanhaudenhuyse et al., 2010; Crone et al., 2014, 2018; Amico et al., 2017) and its importance to consciousness. Due to the inherent nature

of structural damage affecting the structural connectivity, the inclusion of structural terms into a model assessing functional connectivity will help to avoid the FP and FN rate differences between patient in acute and chronic stage of TBI, and the difference in rates between unconscious and conscious patients.

Finally, leaving out both the structural terms and the GWESP term (see column 1 in **Table 8**) has differing rates in FNs for cortico-subcortical connectivity, FN and FPs for within cortical connectivity, and FPs for within subcortical connectivity. Since all these effects are for unconscious patients, it affects all of DOC research for all three possible types of comparisons between cortical and subcortical connectivity. As can be seen in **Table 8**, the combination of leaving both

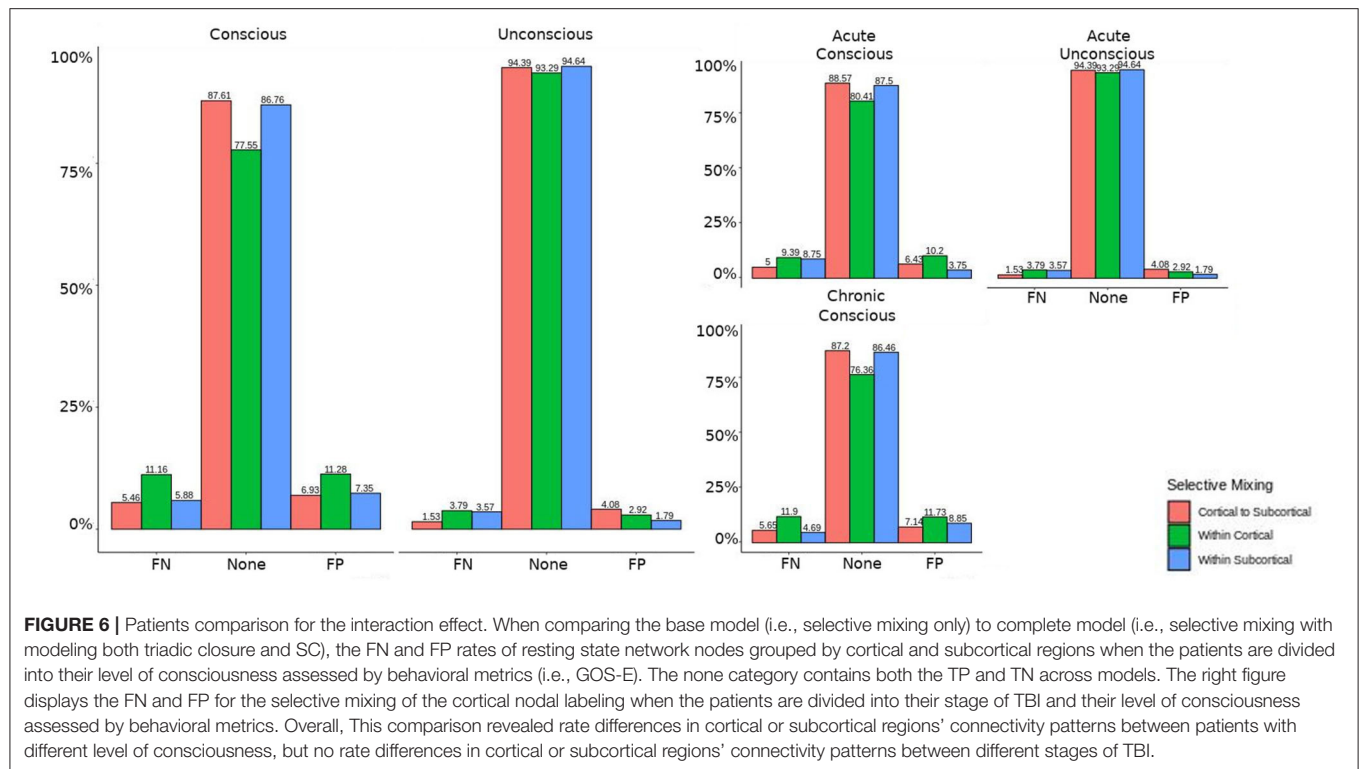


terms is not just a simple combination of leaving out each term, but has a specific impact on unconscious patients. A possible explanation is that, for unconscious patients, structural connectivity alterations related to the traumatic injury affects the patients' triadic closure. These differences, however, are might by not accounting for patients' structural connectivity metrics. These types of interactions are key reasons for including all the generative processes in a graph and other key contributing factors such structural connectivity metrics in patient work, particularly when structural pathology is such a

prominent phenotype of this cohort (Lutkenhoff et al., 2015, 2019).

#### 4.1. Limitations

While we have demonstrated that all three problems would have affected this analysis, we did not fit the data well for the specific effects of triadic closure. The GWESP term matched the overall observed values (see Table 5 for patients, and for HCP participants, see Table 4), but our fits were sub-optimal



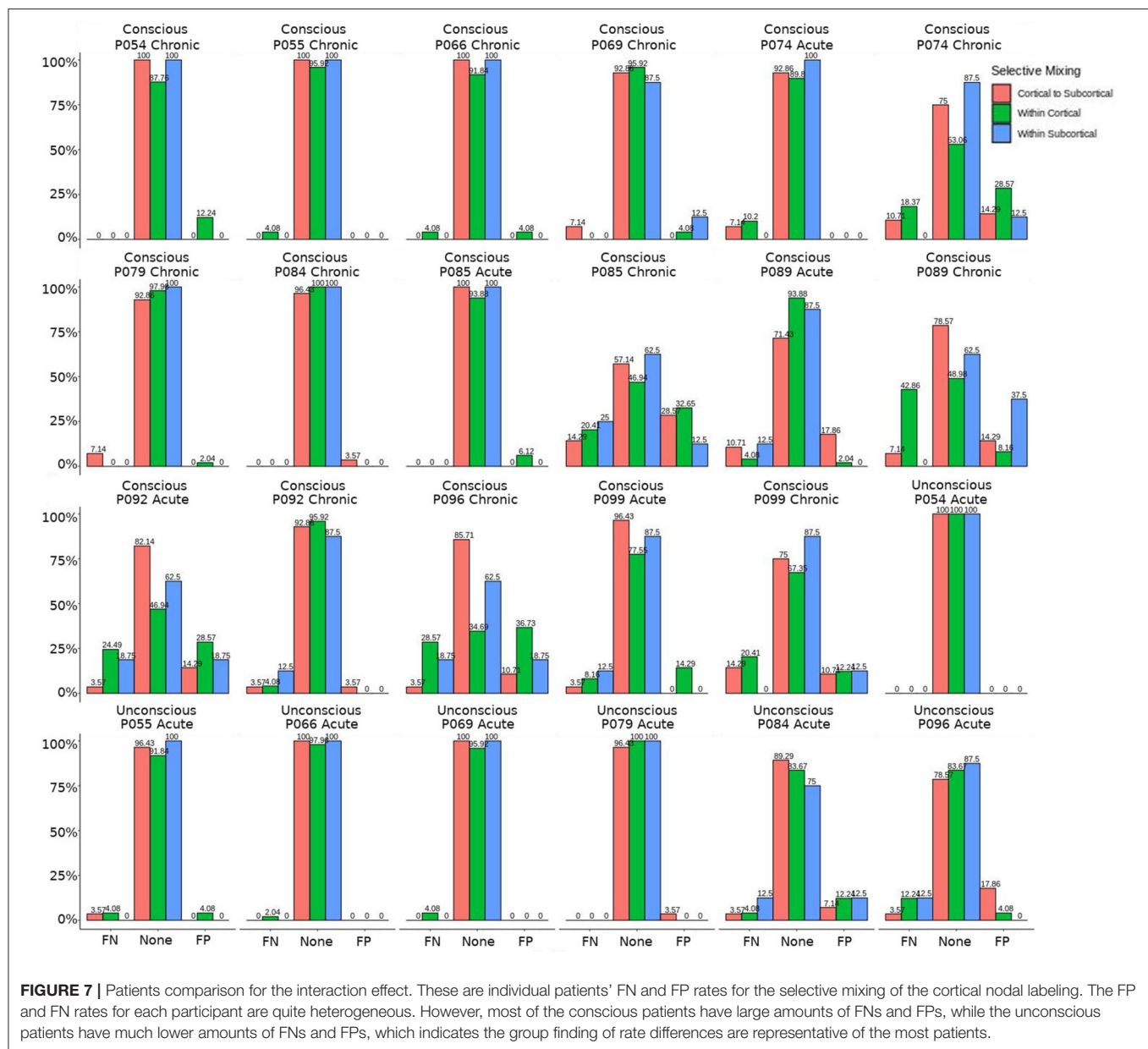
for most specific edge shared partners (see Table 6). The edge shared partners are a type of triadic closure that measures the number of triangles that share an edge. The effect of these poor fits are not well-understood in our field. There is not any work that we are aware of, which explains specific amounts of triadic closure and their neural mechanisms. There is some work describing larger scale interpretations of outcome metrics (e.g., local and global clustering coefficient; Rubinov and Sporns, 2010, 2011) or to characteristic network blocks (i.e., motifs; Sporns and Kötter, 2004), but these accounts are for general interpretations without linking them to key neural mechanism. More theoretical work is needed to understand how these generative process (i.e., sociality, selective mixing, and triadic closure) arise from neural mechanisms.

Finally, we only accounted for the structural terms by using measures of centrality and higher order clustering terms (e.g., clustering coefficient and latent space clusters). This is the equivalent of parceling out the effects of structural connectivity on functional connectivity, similar to using covariates in a regression analysis to control for confounding effects. We have not truly estimated the effects of structural connectivity on functional connectivity. There are structural effects which could account for generative processes that we are attributing to the functional connectivity. For example, if a three nodes have two structural edges, does the functional close this triad? A process like this would represent a possible structural and functional relationship from triad closure from the combination of structural and functional connectivity.

## 5. CONCLUSIONS AND FUTURE WORK

For unconscious patients, the lower rate of false positives and/or false negatives for within cortical connectivity, within subcortical connectivity, and cortico-subcortical connectivity reveal the problems of interrelated outcome statistics and leaving out structural connectivity. These effects can result in misinterpretations of selective mixing terms, if the other generative processes are not included. The proper interpretation selective mixing in functional connectivity for unconscious patients are key to understanding how disruptions of functional connectivity can disrupt consciousness. Additionally, the biases of not including all the generative processes will affect any group comparisons between unconscious patients and other patients with a DOC. We suggest future studies in DOC patients model all the generative processes and include structural connectivity metrics when possible.

We did not compare the patients to the HCP data due to the differences in quality of BOLD imaging and imaging parameters. However, we suspect that there would be differences similar to comparing the groups of patients because the underlying cause of these false positives and false negatives is the importance in generating structure in the brain. Patient populations differ in this structure, which would be reflected by difference in sociality, selective mixing, and triadic closure. Leaving out one of these generative processes will affect the rest (Goodreau et al., 2009). Additionally, the structural terms are specifically important to DOC due to the TBI resulting in structural damage that is part of the recovery process. While there may be alternatives to solving



the three of the four problems we outlined (i.e., problem#1, #2, and #3), we chose to use exponential random graph models due to the flexibility it provides to model all the generative processes and other covariates of interest (e.g., size of regions of interest or amount of atrophy within an region of interest). This flexible framework would allow researchers to generate models that fit their specific needs and research questions.

In addition, a method is needed that accounts for the social processes that generate the complex interactions between structural and functional connectivity. Structural and functional connectivity are part of a multi-level problem (i.e., there is a structural layer, functional layer within the brain and finally an interaction between these layers). By this, we mean that each have their own generative processes that govern their structure

and there are interactions between the levels that drive the brain dynamics. Multi-level exponential random graph models (Wang et al., 2013, 2016; Lazega and Snijders, 2016) have been developing to capture the the nested structure of networks. A concrete example is collaborative research (Lazega et al., 2008), in which, researchers have advice networks for their research problems and the laboratories have collaboration networks. The researchers advice network would be the micro-level network because they are nested within the laboratories (i.e., macro-level). Each of these levels have their own generative processes associated with the structure, but there is a third layer (i.e., the meso-level) that affects both levels. In the collaborative research example, the researchers' affiliation with laboratories are the meso-level. This example could be extended to the



structural and functional layers of the brain. The functional layer is the micro-level because it is nested within the macro-level structural layer. The meso-level could be the locations of the functional layer within the structural layer or an estimation of joint functional structural connectivity could be the meso-level (e.g., hybrid connICA; Amico et al., 2017). These multi-level exponential random graph models would allow for a more complete solution to the four problems posed in this thesis because all the levels of functional and structural connectivity could be jointly estimated and properly modeled. This would be a large step in the direction of unraveling the complex interactions between structural and functional connectivity that generate the ability for our complex behavior.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## REFERENCES

- Amico, E., Marinazzo, D., Di Perri, C., Heine, L., Annen, J., Martial, C., et al. (2017). Mapping the functional connectome traits of levels of consciousness. *Neuroimage* 148, 201–211. doi: 10.1016/j.neuroimage.2017.01.020
- Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. doi: 10.1016/j.media.2007.06.004
- Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54, 2033–2044. doi: 10.1016/j.neuroimage.2010.09.025
- Behrens, T., Johansen-Berg, H., Woolrich, M., Smith, S., Wheeler-Kingshott, C., Boulby, P., et al. (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat. Neurosci.* 6:750. doi: 10.1038/nn1075
- Behrens, T. E., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., et al. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50, 1077–1088. doi: 10.1002/mrm.10609
- Boly, M., Garrido, M. I., Gosseries, O., Bruno, M.-A., Boveroux, P., Schnakers, C., et al. (2011). Preserved feedforward but impaired top-down processes in the vegetative state. *Science* 332, 858–862. doi: 10.1126/science.1202043
- Boly, M., Tshibanda, L., Vanhaudenhuyse, A., Noirhomme, Q., Schnakers, C., Ledoux, D., et al. (2009). Functional connectivity in the default network during resting state is preserved in a vegetative but not in a brain dead patient. *Hum. Brain Mapp.* 30, 2393–2400. doi: 10.1002/hbm.20672
- Bullmore, E., and Sporns, O. (2012). The economy of brain network organization. *Nat. Rev. Neurosci.* 13, 336–349. doi: 10.1038/nrn3214
- Chennu, S., Finoia, P., Kamau, E., Allanson, J., Williams, G. B., Monti, M. M., et al. (2014). Spectral signatures of reorganised brain networks in disorders of consciousness. *PLoS Comput. Biol.* 10:e1003887. doi: 10.1371/journal.pcbi.1003887
- Craddock, R. C., James, G. A., Holtzheimer, P. E., Hu, X. P., and Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* 33, 1914–1928. doi: 10.1002/hbm.21333
- Crone, J. S., Bio, B. J., Vespa, P. M., Lutkenhoff, E. S., and Monti, M. M. (2018). Restoration of thalamo-cortical connectivity after brain injury: recovery of consciousness, complex behavior, or passage of time? *J. Neurosci. Res.* 96, 671–687. doi: 10.1002/jnr.24115
- Crone, J. S., Schurz, M., Höller, Y., Bergmann, J., Monti, M., Schmid, E., et al. (2015). Impaired consciousness is linked to changes in effective connectivity of the posterior cingulate cortex within the default mode network. *Neuroimage* 110, 101–109. doi: 10.1016/j.neuroimage.2015.01.037
- Crone, J. S., Soddu, A., Höller, Y., Vanhaudenhuyse, A., Schurz, M., Bergmann, J., et al. (2014). Altered network properties of the fronto-parietal network and the thalamus in impaired consciousness. *Neuroimage Clin.* 4, 240–248. doi: 10.1016/j.nicl.2013.12.005
- Dell'Italia, J., Johnson, M. A., Vespa, P. M., and Monti, M. M. (2018). Network analysis in disorders of consciousness: four problems and one proposed solution (exponential random graph models). *Front. Neurol.* 9:439. doi: 10.3389/fneur.2018.00439
- Demertzi, A., Schnakers, C., Soddu, A., Bruno, M.-A., Gosseries, O., Vanhaudenhuyse, A., et al. (2011). Neural plasticity lessons from disorders of consciousness. *Front. Psychol.* 1:245. doi: 10.3389/fpsyg.2010.00245
- Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Barttfeld, P., Raimondo, F., et al. (2019). Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Science Adv.* 5:eaat7603. doi: 10.1126/sciadv.aat7603
- Dosenbach, N. U., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., et al. (2010). Prediction of individual brain maturity using fMRI. *Science* 329, 1358–1361. doi: 10.1126/science.1194144
- Fernández-Espejo, D., Soddu, A., Cruse, D., Palacios, E. M., Junque, C., Vanhaudenhuyse, A., et al. (2012). A role for the default mode network in the bases of disorders of consciousness. *Ann. Neurol.* 72, 335–343. doi: 10.1002/ana.23635
- Finger, H., Bönstrup, M., Cheng, B., Messé, A., Hilgetag, C., Thomalla, G., et al. (2016). Modeling of large-scale functional brain networks based on structural connectivity from DTI: comparison with EEG derived phase coupling networks and evaluation of alternative methods along the modeling path. *PLoS Comput. Biol.* 12:e1005025. doi: 10.1371/journal.pcbi.1005025
- Fornito, A., Zalesky, A., and Breakspear, M. (2013). Graph analysis of the human connectome: promise, progress, and pitfalls. *Neuroimage* 80, 426–444. doi: 10.1016/j.neuroimage.2013.04.087
- Fornito, A., Zalesky, A., and Bullmore, E. (2016). *Fundamentals of Brain Network Analysis*. Cambridge, MA: Academic Press.
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage* 80, 105–124. doi: 10.1016/j.neuroimage.2013.04.127
- Goodreau, S. M., Kitts, J. A., and Morris, M. (2009). Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks. *Demography* 46, 103–125. doi: 10.1353/dem.0.0045
- Hacker, C. D., Perlmutter, J. S., Criswell, S. R., Ances, B. M., and Snyder, A. Z. (2012). Resting state functional connectivity of the striatum in Parkinson's disease. *Brain* 135, 3699–3711. doi: 10.1093/brain/awr281

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by UCLA Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JD, MM, and PV designed the experiment. JD and MJ analyzed the data. JD and MM interpreted the results. JD drafted the manuscript. All authors provided the feedback.

## FUNDING

This work was supported by the Tiny Blue Dot foundation (to MM) and the Brain Injury Research Center at UCLA.

- Hunter, D. R. (2007). Curved exponential family models for social networks. *Soc. Netw.* 29, 216–230. doi: 10.1016/j.socnet.2006.08.005
- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). ERGM: a package to fit, simulate and diagnose exponential-family models for networks. *J. Stat. Softw.* 24:nihp54860. doi: 10.18637/jss.v024.i03
- Laureys, S., Faymonville, M. E., Degueldre, C., Fiore, G. D., Damas, P., Lambermont, B., et al. (2000a). Auditory processing in the vegetative state. *Brain* 123(Pt 8), 1589–1601. doi: 10.1093/brain/123.8.1589
- Laureys, S., Faymonville, M. E., Luxen, A., Lamy, M., Franck, G., and Maquet, P. (2000b). Restoration of thalamocortical connectivity after recovery from persistent vegetative state. *Lancet* 355, 1790–1791. doi: 10.1016/S0140-6736(00)02271-6
- Lazega, E., Jourda, M.-T., Mounier, L., and Stofer, R. (2008). Catching up with big fish in the big pond? multi-level network analysis through linked design. *Soc. Netw.* 30, 159–176. doi: 10.1016/j.socnet.2008.02.001
- Lazega, E., and Snijders, T. (2016). *Multilevel Network Analysis for the Social Sciences: Theory, Methods and Applications*. Vol. 12. New York, NY: Springer. doi: 10.1007/978-3-319-24520-1
- Lutkenhoff, E. S., Chiang, J., Tshibanda, L., Kamau, E., Kirsch, M., Pickard, J. D., et al. (2015). Thalamic and extrathalamic mechanisms of consciousness after severe brain injury. *Ann. Neurol.* 78, 68–76. doi: 10.1002/ana.24423
- Lutkenhoff, E. S., Wright, M. J., Shrestha, V., Real, C., McArthur, D. L., Buitrago-Blanco, M., et al. (2019). The thalamic basis of outcome and cognitive impairment in traumatic brain injury. *bioRxiv* doi: 10.1101/669390
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annu. Rev. Sociol.* 27, 415–444. doi: 10.1146/annurev.soc.27.1.415
- Messé, A., Rudrauf, D., Giron, A., and Marrelec, G. (2015). Predicting functional connectivity from structural connectivity via computational models using MRI: an extensive comparison study. *Neuroimage* 111, 65–75. doi: 10.1016/j.neuroimage.2015.02.001
- Milham, M. P., Fair, D., Mennes, M., and Mostofsky, S. H. (2012). The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Front. Syst. Neurosci.* 6:62. doi: 10.3389/fnsys.2012.00062
- Monti, M. M., Lutkenhoff, E. S., Rubinov, M., Boveroux, P., Vanhaudenhuyse, A., Gosseries, O., et al. (2013). Dynamic change of global and local information processing in propofol-induced loss and recovery of consciousness. *PLoS Comput. Biol.* 9:e1003271. doi: 10.1371/journal.pcbi.1003271
- Monti, M. M., Rosenberg, M., Finoia, P., Kamau, E., Pickard, J. D., and Owen, A. M. (2015). Thalamo frontal connectivity mediates top-down cognitive functions in disorders of consciousness. *Neurology* 84, 167–173. doi: 10.1212/WNL.0000000000001123
- Narayan, M., Allen, G. I., and Tomson, S. (2015). Two sample inference for populations of graphical models with applications to functional connectivity. *arXiv [Preprint] arXiv:1502.03853*.
- Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., et al. (2013). Multisite functional connectivity MRI classification of autism: abide results. *Front. Hum. Neurosci.* 7:599. doi: 10.3389/fnhum.2013.00599
- Pandit, A. S., Expert, P., Lambiotte, R., Bonnelle, V., Leech, R., Turkheimer, F. E., et al. (2013). Traumatic brain injury impairs small-world topology. *Neurology* 80, 1826–1833. doi: 10.1212/WNL.0b013e3182929f38
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682. doi: 10.1073/pnas.98.2.676
- Reineberg, A. E., Andrews-Hanna, J. R., Depue, B. E., Friedman, N. P., and Banich, M. T. (2015). Resting-state networks predict individual differences in common and specific aspects of executive function. *Neuroimage* 104, 69–78. doi: 10.1016/j.neuroimage.2014.09.045
- Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph ( $p^*$ ) models for social networks. *Soc. Netw.* 29, 173–191. doi: 10.1016/j.socnet.2006.08.002
- Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069. doi: 10.1016/j.neuroimage.2009.10.003
- Rubinov, M., and Sporns, O. (2011). Weight-conserving characterization of complex functional brain networks. *Neuroimage* 56, 2068–2079. doi: 10.1016/j.neuroimage.2011.03.069
- Shadi, K., Bakhshi, S., Gutman, D. A., Mayberg, H. S., and Dovrolis, C. (2016). A symmetry-based method to infer structural brain networks from probabilistic tractography data. *Front. Neuroinform.* 10:46. doi: 10.3389/fninf.2016.00046
- Sorg, C., Riedl, V., Mühlaus, M., Calhoun, V. D., Eichele, T., Läer, L., et al. (2007). Selective changes of resting-state networks in individuals at risk for Alzheimer's disease. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18760–18765. doi: 10.1073/pnas.0708803104
- Sporns, O., and Kötter, R. (2004). Motifs in brain networks. *PLoS Biol.* 2:e369. doi: 10.1371/journal.pbio.0020369
- Tan, X., Zhou, Z., Gao, J., Meng, F., Yu, Y., Zhang, J., et al. (2019). Structural connectome alterations in patients with disorders of consciousness revealed by 7-tesla magnetic resonance imaging. *Neuroimage Clin.* 22:101702. doi: 10.1016/j.nicl.2019.101702
- Teasdale, G., and Jennett, B. (1974). Assessment of coma and impaired consciousness: a practical scale. *Lancet* 304, 81–84. doi: 10.1016/S0140-6736(74)91639-0
- Van Dijk, K. R., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., and Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J. Neurophysiol.* 103, 297–321. doi: 10.1152/jn.00783.2009
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., et al. (2013). The Wu-Minn human connectome project: an overview. *Neuroimage* 80, 62–79. doi: 10.1016/j.neuroimage.2013.05.041
- van Wijk, B. C. M., Stam, C. J., and Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph theory. *PLoS ONE* 5:e13701. doi: 10.1371/journal.pone.0013701
- Vanhaudenhuyse, A., Noirhomme, Q., Tshibanda, L. J.-F., Bruno, M.-A., Boveroux, P., Schnakers, C., et al. (2010). Default network connectivity reflects the level of consciousness in non-communicative brain-damaged patients. *Brain* 133, 161–171. doi: 10.1093/brain/awp313
- Venables, W. N., and Ripley, B. D. (2013). *Modern Applied Statistics with S-PLUS*. New York, NY: Springer Science & Business Media.
- Voss, H. U., and Schiff, N. D. (2009). MRI of neuronal network structure, function, and plasticity. *Prog. Brain Res.* 175, 483–496. doi: 10.1016/S0079-6123(09)17532-5
- Wang, P., Robins, G., Pattison, P., and Lazega, E. (2013). Exponential random graph models for multilevel networks. *Soc. Netw.* 35, 96–115. doi: 10.1016/j.socnet.2013.01.004
- Wang, P., Robins, G., Pattison, P., and Lazega, E. (2016). Social selection models for multilevel networks. *Soc. Netw.* 44, 346–362. doi: 10.1016/j.socnet.2014.12.003
- Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of “small-world?? networks. *Nature* 393:440. doi: 10.1038/30918
- Werner, C. J., Dogan, I., Saß, C., Mirzazade, S., Schiefer, J., Shah, N. J., et al. (2014). Altered resting-state connectivity in Huntington's disease. *Hum. Brain Mapp.* 35, 2582–2593. doi: 10.1002/hbm.22351
- Wilson, J. L., Pettigrew, L. E., and Teasdale, G. M. (1998). Structured interviews for the Glasgow Outcome Scale and the extended Glasgow Outcome Scale: guidelines for their use. *J. Neurotrauma* 15, 573–585. doi: 10.1089/neu.1998.15.573
- Zheng, Z. S., Reggente, N., Lutkenhoff, E., Owen, A. M., and Monti, M. M. (2017). Disentangling disorders of consciousness: insights from diffusion tensor imaging and machine learning. *Hum. Brain Mapp.* 38, 431–443. doi: 10.1002/hbm.23370
- Zou, Q., Ross, T. J., Gu, H., Geng, X., Zuo, X.-N., Hong, L. E., et al. (2013). Intrinsic resting-state activity predicts working memory brain activation and behavioral performance. *Hum. Brain Mapp.* 34, 3204–3215. doi: 10.1002/hbm.22136

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Dell'Italia, Johnson, Vespa and Monti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Decreased Evoked Slow-Activity After tDCS in Disorders of Consciousness

Armand Mensen<sup>1\*</sup>, Olivier Bodart<sup>1,2</sup>, Aureore Thibaut<sup>1,3,4</sup>, Sarah Wannez<sup>1</sup>, Jitka Annen<sup>1</sup>, Steven Laureys<sup>1,3</sup> and Olivia Gosseries<sup>1,3</sup>

<sup>1</sup> Coma Science Group, GIGA Consciousness, University of Liège, Liège, Belgium, <sup>2</sup> Department of Neurology, University Hospital of Liège, Liège, Belgium, <sup>3</sup> Centre du Cerveau<sup>2</sup>, University Hospital of Liège, Liège, Belgium, <sup>4</sup> Neuromodulation Center, Spaulding Rehabilitation Hospital, Harvard Medical School, Boston, MA, United States

## OPEN ACCESS

### Edited by:

Umberto Olcese,  
Faculty of Science, University  
of Amsterdam, Netherlands

### Reviewed by:

Giulio Bernardi,  
IMT School for Advanced Studies  
Lucca, Italy  
Michael Scharner,  
Université de Genève, Switzerland

### \*Correspondence:

Armand Mensen  
research.mensen@gmail.com

**Received:** 21 May 2020

**Accepted:** 28 July 2020

**Published:** 25 September 2020

### Citation:

Mensen A, Bodart O, Thibaut A,  
Wannez S, Annen J, Laureys S and  
Gosseries O (2020) Decreased  
Evoked Slow-Activity After tDCS  
in Disorders of Consciousness.  
*Front. Syst. Neurosci.* 14:62.  
doi: 10.3389/fnsys.2020.00062

Due to life-saving medical advances, the diagnosis and treatment of disorders of consciousness (DOC) has become a more commonly occurring clinical issue. One recently developed intervention option has been non-invasive transcranial direct current stimulation. This dichotomy of patient responders may be better understood by investigating the mechanism behind the transcranial direct current stimulation (tDCS) intervention. The combination of transcranial magnetic stimulation and electroencephalography (TMS-EEG) has been an important diagnostic tool in DOC patients. We therefore examined the neural response using TMS-EEG both before and after tDCS in seven DOC patients (four diagnosed as in a minimally conscious state and three with unresponsive wakefulness syndrome). tDCS was applied over the dorsolateral prefrontal cortex, while TMS pulses were applied to the premotor cortex. None of the seven patients showed relevant behavioral change after tDCS. We did, however, find that the overall evoked slow activity was reduced following tDCS intervention. We also found a positive correlation between the strength of the slow activity and the amount of high-frequency suppression. However, there was no significant pre-post tDCS difference in high frequencies. In the resting-state EEG, we observed that both the incidence of slow waves and the positive slope of the wave were affected by tDCS. Taken together, these results suggest that the tDCS intervention can reduce the slow-wave activity component of bistability, but this may not directly affect high-frequency activity. We hypothesize that while reduced slow activity may be necessary for the recovery of neural function, especially consciousness, this alone is insufficient.

**Keywords:** disorders of consciousness, transcranial direct current stimulation, transcranial magnetic stimulation, electroencephalography, bistability, slow activity, diagnosis, treatment

## INTRODUCTION

Disorders of consciousness (DOC) have become a growing clinical concern with the advance of medical technologies, which have limited the fatal consequences of severe brain injury events. Distinctions have been made between the levels of recovery from coma, unresponsive wakefulness syndrome/vegetative state (UWS), minimally conscious state (MCS), to emergence from MCS (EMCS) (Giacino et al., 2002; Laureys et al., 2010). Each step is associated with increasing signs

of awareness of one's self or the patient's environment. While in UWS the patient is awake, no such behavioral signs can be reliably found in the patients within this group. Reliable evidence of awareness then characterizes the patient as MCS, which has been subsequently divided into MCS– and MCS+ depending on the absence or presence of language processing (Bruno et al., 2011). Further signs of functional communication or object use transition the patient to EMCS (Giacino et al., 2002). Patients who are unresponsive at the bedside (i.e., UWS) may, however, present similar brain activity to patients in MCS, and they are referred as non-behavioral MCS or MCS\* (Gosseries et al., 2014b; Stender et al., 2014). The development of accurate diagnostic tools to distinguish these states has been key in the understanding of the basic mechanisms of neural activity and the clinical treatment options and also has important ethical and legal implications (Gosseries et al., 2014a; Fins, 2016).

Transcranial direct current stimulation (tDCS) has recently shown promise as a non-invasive, non-pharmaceutical intervention in DOC patients (Thibaut et al., 2019). This technique of passing a small electric current through two electrodes attached to the scalp has been well established to modulate the neural activity of the underlying networks (Nitsche and Paulus, 2000). One stimulation site in particular, the left dorsolateral prefrontal cortex (DLPFC, with the cathode over the right supraorbital area), has been shown to be effective in a number of studies (Angelakis et al., 2014; Thibaut et al., 2014, 2017, 2018a; Naro et al., 2015; Estraneo et al., 2017; Martens et al., 2018; Cavinato et al., 2019; Wu et al., 2019; Hermann et al., 2020). For instance, of 30 MCS patients undergoing just a single session of tDCS, 13 showed new signs of conscious behavior following the intervention (Thibaut et al., 2014). More recently, it was shown that repeated sessions of the same protocol over 5 days significantly improved scores on the Coma Recovery Scale – Revised (CRS-R) in 9 of 16 chronic MCS patients, even 1 week after the end of the stimulation protocol (Thibaut et al., 2017).

While the evidence for clinical improvement is mounting, we nonetheless lack a plausible mechanism for these changes. For over a decade now, the level of consciousness can be probed using single pulses of transcranial magnetic stimulation (TMS), with the neural effects of this stimulation captured using electroencephalography (EEG). This TMS-EEG combination has been shown to be an effective discriminator of conscious levels between participants in sleep and wake conditions (Massimini et al., 2005), including periods of dreaming within rapid and non-rapid eye movement sleep (Massimini et al., 2010; Niemenen et al., 2016; Lee et al., 2019). The approach has also differentiated the effects on (un)conscious level in distinct anesthetic agents (Sarasso et al., 2015; Darracq et al., 2018). In the clinical diagnostic setting, the neural response from the TMS pulse has been quantified into the perturbational complexity index (PCI), where values above an empirically defined cutoff of 0.31 can accurately stratify patients into the same clinical categories made with behavioral assessments (Casali et al., 2013; Casarotto et al., 2016; Bodart et al., 2017). Changes to the complexity of a neural signal can theoretically be achieved in a number of ways, yet two principal mechanisms have been identified under the umbrella term “bistability” (Pigorini et al., 2015;

D'Andola et al., 2018). Firstly, low complexity states found in deep sleep, anesthesia, and DOC show slow activity indicative of a neural “down-state,” associated with hyperpolarization and the cessation of neural firing. Accompanying this slow wave is a relative suppression of high-frequency neural activity. Following the slow wave, spontaneous neural activity might resume, yet this is no longer causally locked to the initial stimulus. This break in causal influence is thought to underlie the loss of consciousness (Pigorini et al., 2015; Tononi et al., 2016). With this *a priori* mechanistic approach, we aimed to directly examine the features of bistability using the well-established TMS-EEG methodology to perturbing neural activity, both before and after tDCS, in a set of DOC patients.

## MATERIALS AND METHODS

### Participants and Diagnostic Assessments

Twenty-one patients with chronic DOC were enrolled in this study at the University Hospital of Liege. Thirteen patients did not complete the entire study due to having no brain responses to TMS-EEG prior to tDCS, significant TMS artifacts or muscle artifacts, patients moving too much, and/or technical issues. Eight chronic DOC patients thus successfully completed the whole study, but one participant was excluded from the analyses due to noisy data. Inclusion criteria were that the patient be over 18; diagnosed as UWS or MCS based on repeated assessments with the CRS-R (Giacino et al., 2004) performed by trained and experienced clinicians; had no contraindications for MRI or TMS (i.e., no history of epilepsy, metallic implants, or external shunts); and no history of other neurological or psychiatric problems. This study was approved by the Ethics Committee of the Faculty of Medicine of the University of Liège. Legal guardians of the patients were informed about the study and signed a written informed consent sheet.

### TDCS and TMS Acquisitions

Patients underwent an established TMS-EEG protocol, as previously published (Rosanova et al., 2012), before and after anodal tDCS on the left DLPFC (see Thibaut et al., 2014). tDCS was performed and lasted 20 min at 2 mA with the cathode on the right supra-orbitofrontal area using the DC Stimulator PLUS (NeuroConn, Germany). The TMS target was the medial premotor cortex (just outside the primary effect location of the tDCS), for 400 trials separated by random intervals between 2 and 3 s using a figure-of-eight coil (Focal Bipulse, Nexstim Plc, Finland). Navigation on patients' 3D T1 MRI image allowed avoiding stimulating over brain lesions (Gosseries et al., 2015). Ultimately, five patients were stimulated over the left premotor cortex while two over the right premotor cortex due to brain lesion location. Resting-state EEG was also acquired before the first TMS-EEG and after the last TMS-EEG session in each patient. Experimenters encouraged the patients to remain awake with their eyes open during the entire time period of recordings. If the patients were unable to do this, breaks were taken and/or a standard arousal protocol was performed (Giacino et al., 2004).



This arousal protocol is part of the CRS-R and uses deep pressure stimulation, starting from the facial musculature to the toes on both sides of the body. The muscles are firmly grasped between the thumb and forefinger and are “rolled” back and forth several times. This helps in prolonging the length of time the patient maintains arousal (i.e., eye opening).

## EEG Recording and Preprocessing

EEG was recorded using a TMS-compatible amplifier, with a 60-channel, low-profile cap at 1,450-Hz sampling rate (eXimia, Nexstim Plc, Finland). The raw EEG data was imported into Matlab (version 2017a, MathWorks, Natick Inc., Boston) and processed using EEGLAB (Delorme and Makeig, 2004) and custom, open-source, scripts<sup>1</sup>. An example of the TMS-evoked activity after each preprocessing step is shown in **Supplementary Figure 1**. A first-order high-pass filter at 0.1 Hz was applied to remove any DC drift. Then, a notch filter (Butterworth 48–52 Hz, sixth order) was applied to remove any power mains noise. This was followed by a bandpass filter (Butterworth 0.5–45 Hz, eighth order). Trials were then created 800 ms before and after each TMS pulse. Bad channels were manually detected and removed ( $Mean = 5.86$ ,  $SD = 2.93$ ). Trials with excessive muscle artifacts were also manually selected and removed, as well as trials with eye movements around the time of the pulse ( $Mean = 113.0$  of 400,  $SD = 49.82$ ).

Independent component analysis was then performed with a reduction of data to the first 25 principal components (as implemented in EEGLAB’s “*binica*” function). Twenty-five were selected so that each patient had the same number of possible components, regardless of the number of bad channels that were previously removed. Moreover, a reduction in the number of components made the assumption of sufficient data points for adequate component convergence more tenable. Independent components were visualized according to the spatial topography of their weights, trial-by-trial activity, mean evoked activity, spectral transformation, and the component time series. Furthermore, we could explore, online, the effect of a single component removal (or re-addition) on the channel time series. Clear eye blink/movement components were principally identified by their topography and removed. Muscular artifacts were identified by their higher-frequency spectral power, generally lateral topography, and no clearly evoked activity (see **Supplementary Figure 2** for examples). Residual activity directly related to the TMS pulse was identified through their mean evoked activity confined within 20 ms of the event marker. After component removal, activity of the missing channels was recalculated using the spline interpolation approach, and the data was then referenced to the average activity of all channels at each sample.

## TMS-EEG Analysis

Event-related potentials were analyzed at the individual level by comparing pre-tDCS trials with post-tDCS trials using *t*-tests for each channel and time point. This mass univariate approach was corrected for multiple comparisons using a threshold-free

cluster enhancement (TFCE) technique followed by a maximum permutation approach (Mensen and Khatami, 2013; Pernet et al., 2015). The same approach was used at the group level using paired *t*-tests for each participant between pre- and post-tDCS on the mean event-related potential (ERP).

Parameters of bistability were assessed in line with those previously used for intracranial EEG (Pigorini et al., 2015). Slow activity was defined as the evoked activity found after filtering the original data between 0.5 and 6 Hz (fourth-order Butterworth filter). Event-related spectral perturbation (ERSP) was calculated using EEGLAB’s *newtimef* function, which uses Morlet wavelets to decompose each trial into the power at increasing frequency bands and baseline corrects the evoked activity for each trial using the baseline period from –400 to –200 ms (low frequencies will leak temporally into the baseline period, so –200 ms is a safe period). This was then converted to decibels (using  $\log_{10}$ ) and then averaged across all trials. From this, the high frequencies were considered in the range of 20–45 Hz. Both high-frequency activation and suppression can be calculated by examining the fifth lowest percentile of activity below zero for suppression and the 95th highest percentile above zero for activation across the range of high frequencies and time points within defined windows described below.

The individual nature of the TMS response made point-to-point analysis unreliable. However, specific aspects of the response could be summarized across the topography and time series for each participant and effectively compared at the group level. Given the strong prior hypotheses on the importance of slow activity, we summarized this effect by taking the minimum (i.e., largest negative amplitude) activity for the slow-wave, bistability measure for each participant in three time windows. The baseline period ran from –400 to –100 ms, the maximum early slow response was taken between 0 and 200 ms, while the late response was taken between 200 and 500 ms. Main effects of the individual terms were assessed using a likelihood ratio test on the complete model vs. the reduced model without that particular term. We explored the consistency of any effect found by examining the minimum 50–95% percentiles of activity; not only the absolute minimum response from each participant. This summary data was assessed for statistically significant differences by a linear mixed model with time window, tDCS condition, and behavioral diagnosis as fixed factors and participant as a random effect.

## Resting-State EEG Analysis

Each participant had two resting-state recordings, one before and one after tDCS, each lasting between 5 and 8 min. Preprocessing proceeded along identical lines as with the TMS-EEG recordings, with a few exceptions: the data was kept as a single continuous recording, and artifacts were manually marked and excluded from the independent component analysis (ICA) and all further analyses. Individual slow waves were detected using an open-source MATLAB-based toolbox<sup>2</sup> (Mensen et al., 2016). In a first stage, the negative envelope of all channels was calculated and then bandpass filtered between 0 and 4 Hz (Chebyshev type

<sup>1</sup><https://github.com/CSC-UW/csc-eeeg-tools>

<sup>2</sup>[www.github.com/mensen/swa-matlab](https://www.github.com/mensen/swa-matlab)

2). The negative envelope is given by taking the mean of the most negative three channels at any given time point (default toolbox setting). This new canonical time series was examined for slow waves by checking for the amplitude and duration of waves between the downward and subsequent upward zero crossings. The duration criterion was kept to its default between 250 and 1,250 ms, while the amplitude criterion was set to an absolute value of 15  $\mu$ V. The amplitude criterion is lower than that in previous analyses for three reasons: average reference tends to reduce amplitudes, especially in the case of more global waves over more channels; the high-pass filter of the negative envelope will baseline shift the entire time series positive; and we can include the individual wave amplitude in the analysis to see whether this had a significant interaction with the other parameters. Several parameters were taken from each detection and examined for pre- vs. post-tDCS effects: wave amplitudes; globality (as percent of channels involved in the slow wave); its duration (essentially the inverse of its base frequency); both the negative and positive peaks; and, finally, the time since the previous wave (an indicator of wave incidence).

## RESULTS

### Clinical Results

The data of seven patients with DOC (four females; mean age of  $34.7 \pm 10.5$  years, mean time since the event  $70.9 \pm 72$  weeks, range 13–200) were analyzed for this study. Four of the patients suffered a traumatic brain injury, two had a hemorrhagic stroke, and one suffered damage from anoxia. Three patients were diagnosed MCS+ (i.e., relative preservation of language function), one MCS– (i.e., signs of consciousness not related to language such as fixation), and three UWS (i.e., only reflex behaviors). However, two UWS patients show brain activity compatible with consciousness using TMS-EEG, positron

emission tomography, and/or functional magnetic resonance imaging and were thus considered as MCS\*. No significant behavioral changes were observed in any of the patients after tDCS. None of the patients showed a change of diagnosis or new signs of consciousness after the tDCS session. **Table 1** reports the individual demographical and clinical data.

### Individual-Evoked Potentials

All participants showed some significant changes between the pre- and post-tDCS trials at the individual level (see **Figure 1** for an example). The pattern of these changes, however, was highly variable between patients as none shared a peak significant channel or time point. Nor was any single channel or time point significantly different for even six of the seven patients (only 15 channels shared significant differences for five of seven participants). This non-overlapping of significant individual differences was confirmed in the group analysis comparing the average evoked potentials, which found no significant changes to the ERPs at the group level [peak channel E9 at 33 ms, corresponding approximately to channel F2 in the 10/20 system;  $T(6) = 12.512$ ,  $p = 0.365$ ].

### Bistability Measures

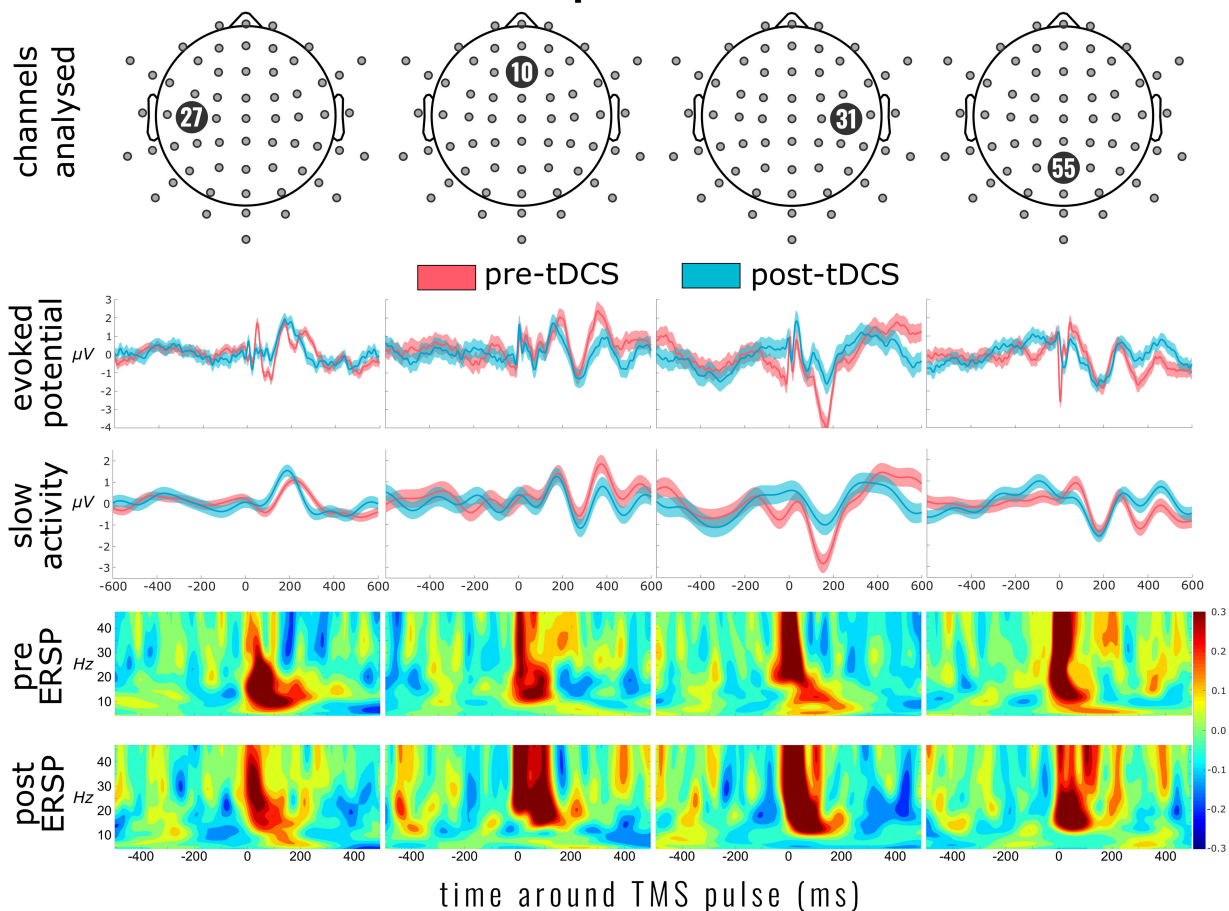
Two distinct aspects of the EEG response were derived to capture proxies of cortical bistability: the slow response and high-frequency activity. The results for the slow responses showed that, at the bottom fifth percentile of each individual patient's response, there was a significant attenuation of the slow activity after tDCS at both the early and later time windows [mean ERP amplitude change =  $0.45 \pm 0.56 \mu$ V,  $\chi^2(2) = 12.141$ ,  $p = 0.002$ ] (see **Figure 2**). There was also a main effect of the time window with a smaller (less negative) slow response at the later time [ $\chi^2(2) = 8.308$ ,  $p = 0.0157$ ], yet no interaction with the tDCS condition [ $\chi^2(1) = 2.021$ ,  $p = 0.155$ ]. This pattern of results was significant from the fifth percentile to the 45th lowest percentile

**TABLE 1 |** Demographical and clinical data of patients with disorders of consciousness (DOC).

Patients	Diagnosis pre-tDCS	Diagnosis post-tDCS	Gender	Age	Etiology	Time since injury (weeks)	Structural imaging	Treatment
1	UWS	UWS	F	25	TBI	33	Right basal ganglia hemorrhagic lesion, DAI maximal in both frontal and mesiotemporal areas	Lamictal
2	MCS+	MCS+	F	26	TBI	145	Right frontal CSF shunt, hydrocephaly, signs of ancient right subdural hematoma	Amantadine
3	MCS+	MCS+	F	32	TBI	200	DAI, global atrophy, no focal lesions	/
4	MCS–	MCS–	M	54	Anoxic	23	Global atrophy	Keppra, amantadine
5	UWS (MCS*)	UWS	M	40	TBI	45	Right frontal CSF shunt, DAI in both frontal lobes, lesions on the parieto-occipital junction on both sides, pre-rolandique lesions more on the left side, left frontal superior gyrus, pons, moderate global atrophy	Amantadine, diazepam, baclofen intrathecal, gabapentine
6	MCS+	MCS+	M	39	Stroke	37	Right MCA hemorrhagic lesion with incidental hygroma, basal ganglia involvement on the left side	Keppra, baclofen, sirdalud
7	UWS (MCS*)	UWS	F	27	Stroke	13	Right temporo-parieto-occipital, left orbitofrontal, thalamic and mesencephalic lesions. Right shunt	Amantadine, zolpidem (to sleep), Tramadol

Patients 5 and 7 were diagnosed MCS\* based on partial metabolic preservation of the frontoparietal network (Stender et al., 2014) and high values of the perturbational complexity index (PCI) (Casali et al., 2013). Patient 7 also showed robust and specific brain activity during active paradigms (Monti et al., 2010) with functional magnetic resonance imaging. tDCS, transcranial direct current stimulation; UWS, unresponsive wakefulness syndrome; MCS, minimally conscious state; M, male; F, female; TBI, traumatic brain injury; DAI, diffuse axonal injury; CSF, cerebrospinal fluid; MCA, middle cerebral artery.

# Individual bistability profile (MCS patient)



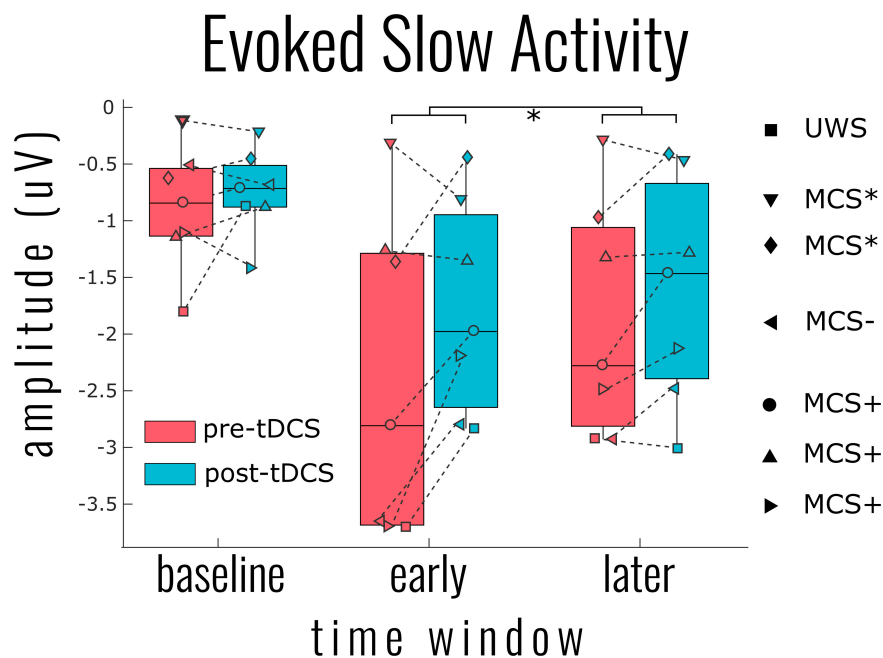
**FIGURE 1 |** Individual bistability profile of a patient in minimally conscious state (MCS; patient 6). *Top panel* indicates the exemplar channels 27, 10 [channel in the center of the supposed effect area of the transcranial direct current stimulation (tDCS)], 31 [nearest channel to the transcranial magnetic stimulation (TMS) pulse], and 55. The evoked potential panel shows the average electroencephalography (EEG) activity 600 ms before and after the single TMS pulses (mean of 267 pulses pre-tDCS and 220 post-tDCS). Below this is the mean slow activity, a low-pass-filtered signal of the TMS-evoked potential. The slow activity profile of channel 31 (*third column*) shows the reduction of slow activity after tDCS. The *two panels below* show the event-related spectral perturbation before and after tDCS intervention. Yellow to red indicates an increase in the power of the corresponding frequency (compared to the mean baseline activity), while shades of blue indicate a suppression of those frequencies.

response, yet with steadily decreasing variance being accounted for (model  $R^2$  from 0.87 down to 0.66). At the individual channel level, there was no significant decrease in slow activity after tDCS, however, the general pattern of results indicated that, for any given patient, the decrease was most likely to be found in posterior channels. Information about the DOC classification of the patients (UWS or MCS) was not a significant predictor of slow wave activity [ $\chi^2(4) = 3.847$ ,  $p = 0.427$ ].

High-frequency activity was examined in a similar approach to slow-wave activity. As **Figure 1** (ERSP) exemplifies, the time period around the TMS pulse was saturated in power across a large range of frequencies; we therefore adjusted the early time period between 100 and 200 ms. As **Figure 3** illustrates, there was a significant activation [ $\chi^2(4) = 17.090$ ,  $p = 0.002$ ] and suppression [ $\chi^2(4) = 18.053$ ,  $p = 0.001$ ] of high frequencies

evoked by the TMS pulse compared to baseline. However, unlike the slow activity, we found no significant differences before and after tDCS in either activation [mean activation change =  $-0.04 \pm 0.13$ ,  $\chi^2(2) = 1.999$ ,  $p = 0.368$ ] or suppression [mean suppression change =  $0.003 \pm 0.03$ ,  $\chi^2(3) = 0.231$ ,  $p = 0.973$ ]. Information on the DOC diagnosis of individual patients significantly improved the overall model for high-frequency suppression [ $\chi^2(6) = 15.397$ ,  $p = 0.017$ ], but not activation [ $\chi^2(6) = 4.987$ ,  $p = 0.546$ ].

We also examined the relationship between the slow response and high-frequency suppression directly. We found that these two measures significantly correlated with one another [ $r = 0.657$ ,  $\chi^2(1) = 13.664$ ,  $p < 0.001$ ]. Furthermore, once the slow activity was included in the mixed model with the tDCS condition, there was a significant interaction between these two factors on the



**FIGURE 2 |** Group results for the effect of transcranial magnetic stimulation (tDCS) on the evoked slow activity. The single point per individual indicates the most negative 5% of activity across all channels within the indicated time windows: baseline (–400 to –100 ms), early (0–200 ms), and later (200–400 ms). This percentile approach within time windows was necessary to generalize the patterns of results over all patients given the highly individualized responses. Linear mixed model analysis indicated a significant reduction of slow activity (i.e., smaller negative amplitude) at both the early and later time windows [ $\chi^2(2) = 12.141$ ,  $p = 0.002$ ]. A significant tDCS effect was found for all negative percentiles across channels from 5% (shown) to 50%.

early time window [ $\chi^2(1) = 7.092$ ,  $p = 0.008$ ]. As **Supplementary Figure 5** illustrates, prior to tDCS, the relationship between these two factors was positive, but weak and non-significant [ $\chi^2(1) = 1.052$ ,  $p = 0.305$ ]. After tDCS, the more the slow activity was reduced, the less the high frequency was suppressed [ $\chi^2(1) = 7.715$ ,  $p = 0.006$ ]. This interaction between slow activity and condition was not the case for the later time window [ $\chi^2(1) = 0.215$ ,  $p = 0.643$ ], nor for the high-frequency activations [ $\chi^2(1) = 0.301$ ,  $p = 0.583$ ]. Here, again, we found a global effect of DOC diagnosis [ $\chi^2(6) = 14.507$ ,  $p = 0.025$ ] on the prediction of high-frequency suppression even when slow activity was also included.

## Resting State

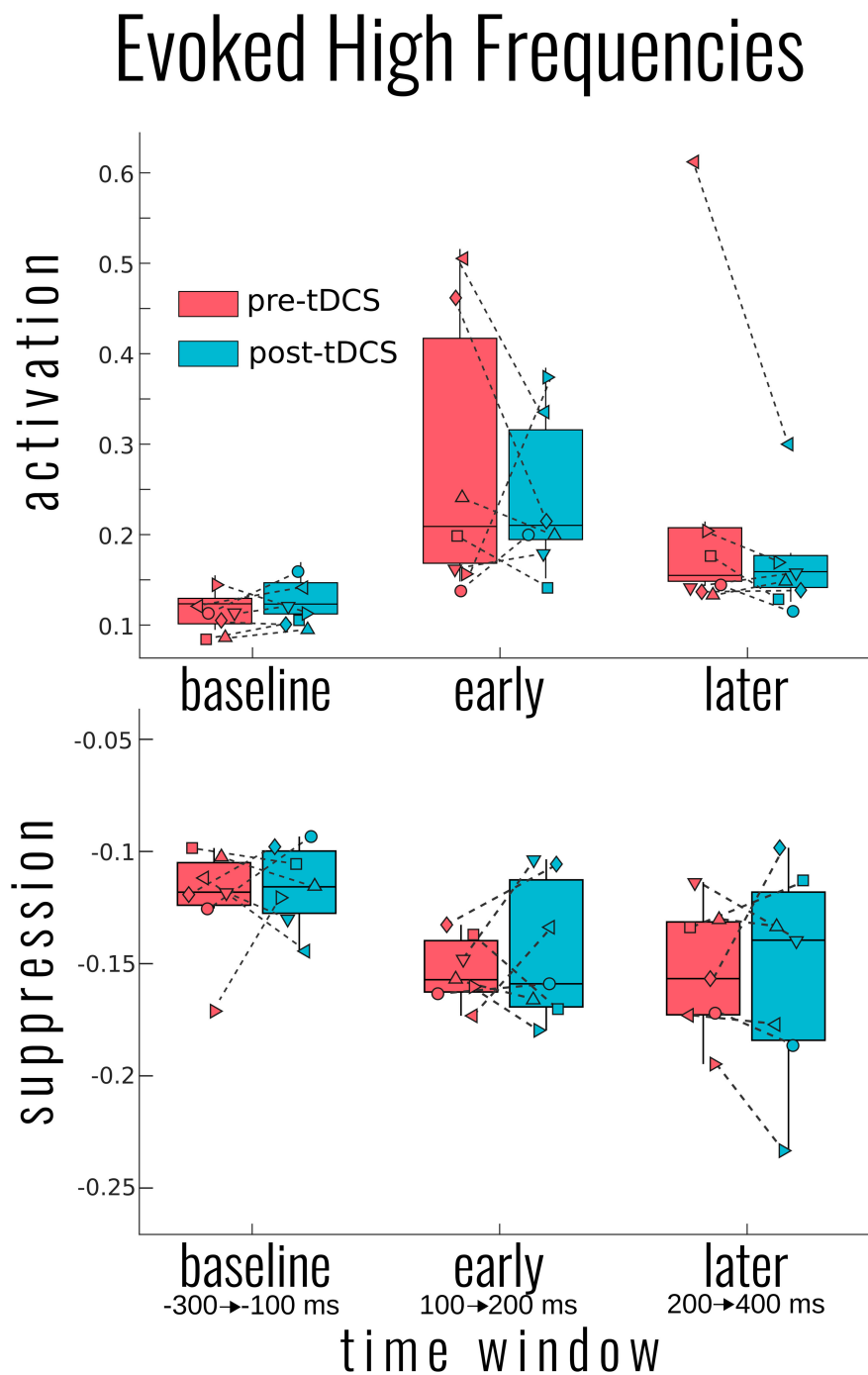
Given the effects of tDCS on slow activity, we further examined patients' resting-state recordings. For two MCS patients, no slow waves were detected using the current criteria, and so the analysis focused on the five remaining participants. A total of 962 waves were detected in all recordings, 608 in resting state prior to tDCS and 354 after tDCS. As with the TMS-EEG response, at the individual participant level, several distinct parameters could be found to be significantly different between the two conditions (see **Supplementary Figure 4** for an example). At the group level, we found main effects on two primary slow-wave parameters. After tDCS, the positive slope of the wave was significantly reduced [ $\chi^2 = 10.116$ ,  $p = 0.002$ ]. This effect remained when the wave amplitude was controlled for [ $\chi^2 = 10.298$ ,  $p = 0.001$ ]. Yet while

we found the expected correlation with the negative slope ( $r^2 = 0.764$ ,  $\chi^2 = 996.830$ ,  $p < 0.001$ ), tDCS did not seem to systematically affect this parameter at the group level ( $\chi^2 = 1.490$ ,  $p = 0.222$ ). Time since last wave, effectively the inter-wave temporal spacing, was also found to be significantly longer after tDCS (estimated difference of 1.287 s;  $\chi^2 = 18.524$ ,  $p < 0.001$ ). Stepwise addition of the other parameters to this model revealed no further interactions or unique effects beyond the main effect of time.

## DISCUSSION

This study sought to examine the effect of prefrontal tDCS on cortical bistability in a small sample of DOC patients. The combination of TMS-EEG is regularly used to aid in the classification of DOC patients, specifically in the differentiation of UWS and MCS, by perturbing the state of the brain and capturing the causal influence (Napolitani et al., 2014; Sarasso et al., 2014). Firstly, none of the seven patients demonstrated any novel signs of consciousness following tDCS. Given that previous studies have shown that around half of MCS patients showed some improvement, this was somewhat unexpected (Thibaut et al., 2014; Estraneo et al., 2017). However, with our focus on the TMS-EEG measures, behavioral assessments were delayed compared to the previous studies and so could have missed acute changes. Another reason for the lack of behavioral changes was likely the high intensity of the entire experimental protocol, lasting between





**FIGURE 3 |** Group results for the effect of transcranial magnetic stimulation (tDCS) on the evoked high-frequency power. Activation (**top**) and suppression (**bottom**) of high frequencies (20–50 Hz) for each participant and group average in box plots as measured by the bottom and top 5% of the event-related spectral perturbation over each of the indicated time windows. No significant differences in pre- to post-tDCS were found for either high-frequency measure. Each patient's disorders of consciousness (DOC) classification can be found in **Figure 2** above.

4 and 6 h, undoubtedly causing fatigue for the patients. Regarding the response to TMS, we found that each patient had relatively unique evoked potentials in response to the TMS pulse, with significant effects of tDCS on the direct spatiotemporal activity at the individual level, but with no consistent, specific effects over

all patients. We then examined the two hallmark signatures of cortical bistability: the slow wave and the following suppression of high-frequency activity. We found a significant group-level reduction in slow activity after tDCS, yet with a non-specific spatial profile consistent with individual variability. Furthermore,

patients' slow activity was related to the amount of high-frequency suppression, consistent with the concept of bistability. However, there was no group-level reduction in the amount of suppression or in the activation of high frequencies, suggesting that, while correlated, slow activity and high frequencies can be, and were, independently modulated. The effect of tDCS on slow activity was further supported by demonstrating that the properties of individual slow waves were also significantly altered after tDCS. While their amplitudes remained unchanged, there was a significant reduction in the incidence of slow waves.

Previously, Bai et al. (2017) demonstrated an increase in cortical excitability in both MCS and UWS patients following the same tDCS protocol as used here. Their measure of cortical excitability was the global mean field power (GFP) of the mean EEG response following TMS. Given that most of the power in the EEG signals will indefinitely reside in the lower spectral frequencies (i.e., slow activity), this GFP increase initially seems to counter the decrease of slow activity presented in our study. Therefore, one key difference in these studies is the analytical approaches used. While Bai and colleagues examined more general and widely used metrics, here, we focused on measures of bistability that have a closer theoretical underpinning to the study of consciousness (Pigorini et al., 2015). However, when we reanalyzed our data using GFP (see **Supplementary Figure 3**), we found an early decrease, but which was not significant. We believe that the key distinction between these projects is the site of the TMS. Bai and colleagues stimulated directly over the left DLPFC, presumably over the region of the cortex directly altered by the tDCS intervention (Bai et al., 2017). On the other hand, we perturbed the premotor cortex, which lies outside the main electric field of the tDCS and, therefore, over areas of the cortex that would have retained more of their physiological responses. Here, we were primarily concerned with electrophysiological measures of conscious awareness and not in the cortical effects of tDCS *per se*. We are therefore confident that perturbation outside the direct influence of the intervention is a more meaningful way to examine the global neural effects. In this regard, it is not unreasonable to expect that, had our TMS site been over the DLPFC, or that of Bai and colleagues been outside this area, we would have found similar results.

It is critical to note that none of the patients we recorded showed a significant behavioral improvement despite various degrees of slow activity modulation. On its surface, this suggests that the reduction of this neural slow activity is insufficient for the recovery of important networks underlying potential behavioral improvement. There are three potential elements at work here, none mutually exclusive. The first is that, despite some reduction in the slow activity, further reduction, i.e., larger clusters of neurons out of the bistable state, is necessary for normal functioning. Secondly, while slow activity can be reduced, this is not enough for the resumption of high-frequency activity associated with the normal functioning of critical networks, and despite their correlation, there is an intermediate step between these elements that is often not directly affected through tDCS. Thirdly, there is an interaction of slow activity and high frequencies within certain brain areas, cortical layers, or across

distinct networks that are damaged in certain patients and not others. The latter is certainly the case, to some extent, and likely the most difficult situation to disentangle.

Regarding the resting-state EEG results, we mainly found a significant reduction in the incidence of slow waves after tDCS. In the same line, a recent EEG study reported decreases in the delta band when assessing coherence in patients with DOC after tDCS over the precuneus (Guo et al., 2019). Other EEG studies showed increased frontoparietal coherence in the alpha, beta, and/or theta frequency bands (Bai et al., 2018; Cavinato et al., 2019), along with an increase in P300 amplitude (Zhang et al., 2017) after tDCS over the DLPFC in MCS patients. Comparing tDCS responders to non-responders, increased EEG power, network centrality, and functional connectivity were found in the theta and theta-alpha bands, with larger P300 responses after tDCS (Thibaut et al., 2018b; Hermann et al., 2020).

Thibaut et al. (2018a) present a case study of a patient showing limited signs of consciousness, but whose passive neuroimaging examination showed preserved activity and metabolism. tDCS was able to unveil signs of conscious behavior in this patient. With respect to the results here, tDCS can modulate neural networks, but patients likely already need to be close to the necessary conditions for recovery of function. That is, that key networks need to be structurally preserved, albeit possibly in a bistable state. Modulation of patients' slow activity might then physiologically be accompanied by a reactivation of the high frequencies necessary for behavior improvement. Future work should aim to preselect patients based on their prior demonstrated responsiveness to tDCS (or lack of), with extensive qualification of the lesions involved (both structurally and functionally). While here we included patients with UWS diagnosis, the previous study that successfully modulated behavior with tDCS was directed toward MCS patients. Here, we present half of the potential results of such a study; that tDCS can indeed modulate bistability, as hypothesized, but this modulation alone is insufficient to improve the behavioral state of patients. We have provided some evidence that the correlated high-frequency suppression may be another key missing element, yet not directly altered by the tDCS. We can only speculate that, in patients who respond behaviorally to the tDCS intervention, the reduction of slow activity is combined with the reemergence of high-frequency activity.

## Limitations and Future Perspectives

Our limited sample size of seven patients is a clear limiting factor in generalizing the results of such a study. However, the demands on the patient and the experimenter of performing multiple sessions of TMS-EEG assessments, in combination with the tDCS intervention, are extraordinarily high. Given the large parameter space of settings and localization of tDCS and the variability in the etiology of patients, we believe that the in-depth characterization of effects at the individual patient level is nevertheless a fruitful path to navigate through this space. A further limitation was the long experimental duration and the non-random nature of the pre-post measures. Patients here may have become fatigued throughout the day, which could have contributed to the pattern of results found here.

Further studies could include sham tDCS stimulation, a shorter TMS-EEG protocol, or multiple measures post-tDCS where, presumably, the effects of the intervention may decrease but fatigue would increase to delineate these effects. Future studies may also wish to focus on patients who had previously shown improvement with tDCS to increase the chance of correlating the behavioral improvements to the changes in the bistability measures demonstrated here.

## CONCLUSION

The present results suggest that a single session of prefrontal tDCS can reduce the slow-wave activity component of bistability, but this may not directly affect the high-frequency activity. We hypothesize that while reduced slow activity may be necessary for the recovery of neural function, especially consciousness, this alone is insufficient as we did not observe significant clinical improvement. Repeated sessions of tDCS may be necessary to induce a behavioral response paralleled with a reduction of slow activity combined with the reemergence of high-frequency activity.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Committee of the University Hospital of Liege, Liège, Belgium. Written informed consent was provided by the patients' legal guardians.

## REFERENCES

- Angelakis, E., Liouta, E., Andreadis, N., Korfiatis, S., Ktonas, P., Stranjalis, G., et al. (2014). Transcranial direct current stimulation effects in disorders of consciousness. *Arch. Phys. Med. Rehabil.* 95, 283–289.
- Bai, Y., Xia, X., Kang, J., Yang, Y., He, J., Li, X., et al. (2017). TDCS modulates cortical excitability in patients with disorders of consciousness. *Neuroimage Clin.* 15, 702–709. doi: 10.1016/j.nicl.2017.01.025
- Bai, Y., Xia, X., Wang, Y., Guo, Y., Yang, Y., He, J., et al. (2018). Frontoparietal coherence response to tDCS modulation in patients with disorders of consciousness. *Int. J. Neurosci.* 128, 587–594. doi: 10.1080/00207454.2017.1403440
- Bodart, O., Gosseries, O., Wannez, S., Thibaut, A., Annen, J., Boly, M., et al. (2017). Measures of metabolism and complexity in the brain of patients with disorders of consciousness. *Neuroimage Clin.* 14, 354–362. doi: 10.1016/j.nicl.2017.02.002
- Bruno, M. A., Vanhaudenhuyse, A., Thibaut, A., Moonen, G., and Laureys, S. (2011). From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: recent advances in our understanding of disorders of consciousness. *J. Neurol.* 258, 1373–1384. doi: 10.1007/s00415-011-6114-x

## AUTHOR CONTRIBUTIONS

AM, OB, AT, SL, and OG planned and designed the study. OB, AT, SW, and JA acquired the data. AM and OG analyzed and interpreted the data. AM drafted the manuscript. All the authors revised and approved the manuscript.

## FUNDING

The study was supported by the Belgian National Funds for Scientific Research (FRS-FNRS), the University and University Hospital of Liège, the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement Nos. 785907 and 945539 (Human Brain Project SGA2 and SGA3), the Luminous project (EU-H2020-fetopen-ga686764), the Fund Generet, the King Baudouin Foundation, the AstraZeneca Foundation, the DOCMA project (EU-H2020-MSCA-RISE-778234), the BIAL Foundation, the James McDonnell Foundation, Mind Science Foundation, IAP research network P7/06 of the Belgian Government (Belgian Science Policy), the European Commission, the Public Utility Foundation “Université Européenne du Travail,” and “Fondazione Europea di Ricerca Biomedica”. AT is a post-doctoral fellow, OG is research associate, and SL is research director at FRS-FNRS.

## ACKNOWLEDGMENTS

We thank the patients and their families.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnsys.2020.00062/full#supplementary-material>

- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* 5:198ra105. doi: 10.1126/scitranslmed.3006294
- Casarotto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fecchio, M., Napolitani, M., et al. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Ann. Neurol.* 80, 718–729. doi: 10.1002/ana.24779
- Cavinato, M., Genna, C., Formaggio, E., Gregorio, C., Storti, S. F., Manganotti, P., et al. (2019). Behavioural and electrophysiological effects of tDCS to prefrontal cortex in patients with disorders of consciousness. *Clin. Neurophysiol.* 130, 231–238. doi: 10.1016/j.clinph.2018.10.018
- D'Andola, M., Rebollo, B., Casali, A. G., Weinert, J. F., Pigorini, A., Villa, R., et al. (2018). Bistability, causality, and complexity in cortical networks: an in vitro perturbational study. *Cereb. Cortex* 28, 2233–2242. doi: 10.1093/cercor/bhx122
- Darracq, M., Funk, C. M., Polyakov, D., Riedner, B., Gosseries, O., Niemenen, J. O., et al. (2018). Evoked alpha power is reduced in disconnected consciousness during sleep and anesthesia. *Sci. Rep.* 8:16664.
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

- Estraneo, A., Pascarella, A., Moretta, P., Masotta, O., Fiorenza, S., Chirico, G., et al. (2017). Repeated transcranial direct current stimulation in prolonged disorders of consciousness: a double-blind cross-over study. *J. Neurol. Sci.* 375, 464–470. doi: 10.1016/j.jns.2017.02.036
- Fins, J. J. (2016). Neuroethics and disorders of consciousness: discerning brain states in clinical practice and research. *AMA J. Ethics* 18, 1182–1191. doi: 10.1001/journalofethics.2016.18.12.ecas2-1612
- Giacino, J. T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., Katz, D. I., et al. (2002). The minimally conscious state: definition and diagnostic criteria. *Neurology* 58, 349–353. doi: 10.1212/wnl.58.3.349
- Giacino, J. T., Kalmar, K., and Whyte, J. (2004). The JFK coma recovery scale-revised: measurement characteristics and diagnostic utility. *Arch. Phys. Med. Rehabil.* 85, 2020–2029. doi: 10.1016/j.apmr.2004.02.033
- Gosseries, O., Di, H., Laureys, S., and Boly, M. (2014a). Measuring consciousness in severely damaged brains. *Annu. Rev. Neurosci.* 37, 457–478. doi: 10.1146/annurev-neuro-062012-170339
- Gosseries, O., Zasler, N. D., and Laureys, S. (2014b). Recent advances in disorders of consciousness: focus on the diagnosis. *Brain Inj.* 28, 1141–1150. doi: 10.3109/02699052.2014.920522
- Gosseries, O., Sarasso, S., Casarotto, S., Boly, M., Schnakers, C., Napolitani, M., et al. (2015). On the cerebral origin of EEG responses to TMS: insights from severe cortical lesions. *Brain Stimul.* 8, 142–149. doi: 10.1016/j.brs.2014.10.008
- Guo, Y., Bai, Y., Xia, X., Li, J., Wang, X., Dai, Y., et al. (2019). Effects of long-lasting high-definition transcranial direct current stimulation in chronic disorders of consciousness: a pilot study. *Front. Neurosci.* 13:412. doi: 10.3389/fnins.2019.00412
- Hermann, B., Raimondo, F., Hirsch, L., Huang, Y., Denis-Valente, M., Perez, P., et al. (2020). Combined behavioral and electrophysiological evidence for a direct cortical effect of prefrontal tDCS on disorders of consciousness. *Sci. Rep.* 10:4323.
- Laureys, S., Celesia, G. G., Cohadon, F., Lavrijsen, J., Leon-Carrion, J., Sannita, W. G., et al. (2010). Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. *BMC Med.* 8:68. doi: 10.1186/1741-7015-8-68
- Lee, M., Baird, B., Gosseries, O., Nieminen, J. O., Boly, M., Postle, B. R., et al. (2019). Connectivity differences between consciousness and unconsciousness in non-rapid eye movement sleep: a TMS-EEG study. *Sci. Rep.* 9:5175.
- Martens, G., Lejeune, N., O'Brien, A. T., Fregni, F., Martial, C., Wannez, S., et al. (2018). Randomized controlled trial of home-based 4-week tDCS in chronic minimally conscious state. *Brain Stimul.* 11, 982–990. doi: 10.1016/j.brs.2018.04.021
- Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., Tononi, G., et al. (2005). Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. doi: 10.1126/science.1117256
- Massimini, M., Ferrarelli, F., Murphy, M., Huber, R., Riedner, B., Casarotto, S., et al. (2010). Cortical reactivity and effective connectivity during REM sleep in humans. *Cogn. Neurosci.* 1, 176–183. doi: 10.1080/17588921003731578
- Mensen, A., and Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *Neuroimage* 67, 111–118. doi: 10.1016/j.neuroimage.2012.10.027
- Mensen, A., Riedner, B., and Tononi, G. (2016). Optimizing detection and analysis of slow waves in sleep EEG. *J. Neurosci. Methods* 274, 1–12. doi: 10.1016/j.jneumeth.2016.09.006
- Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., et al. (2010). Willful modulation of brain activity in disorders of consciousness. *N. Engl. J. Med.* 362, 579–589. doi: 10.1056/nejmoa0905370
- Napolitani, M., Bodart, O., Canali, P., Seregni, F., Casali, A., Laureys, S., et al. (2014). Transcranial magnetic stimulation combined with high-density EEG in altered states of consciousness. *Brain Inj.* 28, 1180–1189. doi: 10.3109/02699052.2014.920524
- Naro, A., Russo, M., Leo, A., Bramanti, P., Quartarone, A., Calabro, R. S., et al. (2015). A single session of repetitive transcranial magnetic stimulation over the dorsolateral prefrontal cortex in patients with unresponsive wakefulness syndrome: preliminary results. *Neurorehabil. Neural. Repair* 29, 603–613. doi: 10.1177/1545968314562114
- Nieminen, J. O., Gosseries, O., Massimini, M., Saad, E., Sheldon, A. D., Boly, M., et al. (2016). Consciousness and cortical responsiveness: a within-state study during non-rapid eye movement sleep. *Sci. Rep.* 6:30932.
- Nitsche, M. A., and Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J. Physiol.* 527(Pt 3), 633–639. doi: 10.1111/j.1469-7793.2000.t01-1-00633.x
- Pernet, C. R., Latinus, M., Nichols, T. E., and Rousselet, G. A. (2015). Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: a simulation study. *J. Neurosci. Methods* 250, 85–93. doi: 10.1016/j.jneumeth.2014.08.003
- Pigorini, A., Sarasso, S., Proserpio, P., Szymanski, C., Arnulfo, G., Casarotto, S., et al. (2015). Bistability breaks-off deterministic responses to intracortical stimulation during non-REM sleep. *Neuroimage* 112, 105–113. doi: 10.1016/j.neuroimage.2015.02.056
- Rosanova, M., Gosseries, O., Casarotto, S., Boly, M., Casali, A. G., Bruno, M. A., et al. (2012). Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. *Brain* 135(Pt 4), 1308–1320. doi: 10.1093/brain/awr340
- Sarasso, S., Boly, M., Napolitani, M., Gosseries, O., Charland-Verville, V., Casarotto, S., et al. (2015). Consciousness and complexity during unresponsiveness induced by propofol, xenon, and ketamine. *Curr. Biol.* 25, 3099–3105. doi: 10.1016/j.cub.2015.10.014
- Sarasso, S., Rosanova, M., Casali, A. G., Casarotto, S., Fecchio, M., Boly, M., et al. (2014). Quantifying cortical EEG responses to TMS in (un)consciousness. *Clin. EEG Neurosci.* 45, 40–49. doi: 10.1177/1550059413513723
- Stender, J., Gosseries, O., Bruno, M., Charland-Verville, V., Vanhaudenhuyse, A., Demertzi, A., et al. (2014). Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. *Lancet* 384, 514–522. doi: 10.1016/s0140-6736(14)60042-8
- Thibaut, A., Bruno, M. A., Ledoux, D., Demertzi, A., and Laureys, S. (2014). tDCS in patients with disorders of consciousness: sham-controlled randomized double-blind study. *Neurology* 82, 1112–1118.
- Thibaut, A., Chatelle, C., Vanhaudenhuyse, A., Martens, G., Cassol, H., Martial, C., et al. (2018a). Transcranial direct current stimulation unveils covert consciousness. *Brain Stimul.* 11, 642–644. doi: 10.1016/j.brs.2018.02.002
- Thibaut, A., Chennu, S., Chatelle, C., Martens, G., Annen, J., Cassol, H., et al. (2018b). Theta network centrality correlates with tDCS response in disorders of consciousness. *Brain Stimul.* 11, 1407–1409. doi: 10.1016/j.brs.2018.09.002
- Thibaut, A., Schiff, N., Giacino, J., Laureys, S., and Gosseries, O. (2019). Therapeutic interventions in patients with prolonged disorders of consciousness. *Lancet Neurol.* 18, 600–614. doi: 10.1016/s1474-4422(19)30031-6
- Thibaut, A., Wannez, S., Donneau, A. F., Chatelle, C., Gosseries, O., Bruno, M. A., et al. (2017). Controlled clinical trial of repeated prefrontal tDCS in patients with chronic minimally conscious state. *Brain Inj.* 31, 466–474. doi: 10.1080/02699052.2016.1274776
- Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461. doi: 10.1038/nrn.2016.44
- Wu, M., Yu, Y., Luo, L., Wu, Y., Gao, J., Ye, X., et al. (2019). Efficiency of repetitive transcranial direct current stimulation of the dorsolateral prefrontal cortex in disorders of consciousness: a randomized sham-controlled study. *Neural. Plast.* 2019, 1–11. doi: 10.1155/2019/7089543
- Zhang, Y., Song, W., Du, J., Huo, S., Shan, G., Li, R., et al. (2017). Transcranial direct current stimulation in patients with prolonged disorders of consciousness: combined behavioral and event-related potential evidence. *Front. Neurol.* 8:620. doi: 10.3389/fneur.2017.00620

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Mensen, Bodart, Thibaut, Wannez, Annen, Laureys and Gosseries. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Segregated Co-activation Patterns in the Emergence of Decision Confidence During Visual Perception

Cilia Jaeger<sup>1,2,3</sup>, Sarah Glim<sup>1,2,3</sup>, Cristiana Dimulescu<sup>1,2</sup>, Anja Ries<sup>1,2</sup>, Christian Sorg<sup>1,2,4</sup> and Afra Wohlschläger<sup>1,2\*</sup>

<sup>1</sup>Department of Neuroradiology, Technical University of Munich, Munich, Germany, <sup>2</sup>TUM-Neuroimaging Center, Technical University of Munich, Munich, Germany, <sup>3</sup>Graduate School of Systemic Neurosciences, LMU Munich, Planegg-Martinsried, Germany, <sup>4</sup>Department of Psychiatry, Technical University of Munich, Munich, Germany

Visual metacognition—the introspection and evaluation of one’s own visual perceptual processes—is measured through both decision confidence and “metacognitive efficiency.” Metacognitive efficiency refers to an individual’s ability to accurately judge incorrect and correct decisions through confidence ratings given their task performance. Previous imaging studies in humans and nonhuman primates reported widely distributed brain regions being involved in decision confidence and metacognition. However, the neural correlates of metacognition are remarkably inconsistent across studies concerning spatial outline. Therefore, this study investigates the neural correlates of visual metacognition by examining co-activation across regions that scale with visual decision confidence. We hypothesized that interacting processes of perceptual and metacognitive performance contribute to the arising decision confidence in distributed, but segregable co-activating brain regions. To test this hypothesis, we performed task-fMRI in healthy humans during a visual backward masking task with four-scale, post-decision confidence ratings. We measured blood oxygenation covariation patterns, which served as a physiological proxy for co-activation across brain regions. Decision confidence ratings and an individual’s metacognitive efficiency served as behavioral measures for metacognition. We found three distinct co-activation clusters involved in decision confidence: the first included right-centered fronto-temporal-parietal regions, the second included left temporal and parietal regions, and the left basal forebrain (BF), and the third included cerebellar regions. The right fronto-temporal-parietal cluster including the supplementary eye field and the right basal forebrain showed stronger co-activation in subjects with higher metacognitive efficiency. Our results provide novel evidence for co-activation of widely distributed fronto-parieto-temporal regions involved in visual confidence. The supplementary eye field was the only region that activated for both decision confidence and metacognitive efficiency, suggesting the supplementary eye field plays a key role in visual metacognition. Our results link findings in electrophysiology studies and human fMRI studies and provide evidence that confidence estimates arise from the integration of multiple information processing pathways.

**Keywords:** metacognition, decision confidence, fMRI, supplementary eye field, visual perception

## OPEN ACCESS

### Edited by:

Maria V. Sanchez-Vives,  
Institut de Recerca Biomèdica  
August Pi i Sunyer (IDIBAPS), Spain

### Reviewed by:

Ignasi Cos,  
Pompeu Fabra University, Spain  
Mazyar Fallah,  
York University, Canada

### \*Correspondence:

Afra Wohlschläger  
afra.wohlschlaeger@tum.de

**Received:** 30 April 2020

**Accepted:** 14 October 2020

**Published:** 10 November 2020

### Citation:

Jaeger C, Glim S, Dimulescu C,  
Ries A, Sorg C and Wohlschläger A  
(2020) Segregated Co-activation  
Patterns in the Emergence of  
Decision Confidence During Visual  
Perception.  
Front. Syst. Neurosci. 14:557693.  
doi: 10.3389/fnsys.2020.557693

## INTRODUCTION

This study investigates the neural correlates of visual metacognition by examining co-activation across regions that vary relative to visual decision confidence. Metacognition is defined as the ability to introspect and evaluate the quality of one's decision making (Metcalfe and Shimamura, 1994; Shimamura, 2000; Fleming and Lau, 2014). Metacognition is crucial for guiding behavior, especially in the absence of external feedback, and mitigates future mistakes (Metcalfe and Shimamura, 1994; Yeung and Summerfield, 2012). Visual metacognition is commonly operationalized through subjective confidence ratings about the accuracy of one's visual decision-making processes. Decision confidence has been extensively used to quantify metacognition in humans and primates (Kepecs and Mainen, 2012; Mamassian, 2016; Bang and Fleming, 2018; Vaccaro and Fleming, 2018, for review). Yet, the neural underpinnings of visual metacognition remain unclear due to diverse findings across studies.

Neural correlates of visual metacognition have been reported for widely distributed brain regions for two classes of studies. One class of studies, mostly performing fMRI in humans, primarily identified regions in the prefrontal and cingulo-opercular cortex. Metacognitive ability has been correlated to confidence-related activity in the dorsolateral prefrontal cortex (dlPFC; Lau and Passingham, 2006; Del Cul et al., 2009), rostromedial prefrontal cortex (Fleming et al., 2012; Morales et al., 2018), dorsal anterior cingulate cortex (dACC; Fleming et al., 2012; Bang and Fleming, 2018; Morales et al., 2018), and the ventral striatum (Hebart et al., 2014). These results are conceptualized by a model that suggests that objective decision making and metacognitive performance, measured by decision confidence, occur in coupled but distinct networks (Pasquali et al., 2010; Grimaldi et al., 2015; Fleming and Daw, 2017). In more detail, visual confidence is thought to emerge in prefrontal and frontal areas and reverberate back through recurrent pathways to parietal and early visual areas (Del Cul et al., 2009; Fleming et al., 2010; Fleming and Dolan, 2012; Fleming and Daw, 2017). Behaviorally, objective task performance and subjective evaluation of perceptual decisions through confidence ratings can be dissociated, as the two processes occur independently (Maniscalco and Lau, 2012; Fleming and Daw, 2017; Qiu et al., 2018). For example, disassociation occurs when task performance is poor, yet the subject reports high confidence of being correct. Another class of studies, mostly performing electrophysiology in humans (Gherman and Piliastides, 2015) and non-human primates associated the lateral inferior parietal lobe (LIP), an area known to be involved in decision making (Kiani and Shadlen, 2009), and the supplementary eye fields (SEF; Middlebrooks and Sommer, 2012; So and Stuphorn, 2015) with visual metacognition. These results are conceptualized by a different model, in which task choice and decision confidence arise from the same internal state (Kiani and Shadlen, 2009; Kiani et al., 2014; Van Den Berg et al., 2016). Sensory evidence accumulates until a perceptual decision threshold for one type of stimulus is reached. During the

metacognitive, second-order decision, the amount of confidence is determined by the distance between the decision boundary and additional accumulated sensory evidence for the different choice options (Pleskac and Busemeyer, 2010; Kiani et al., 2014). The question remains whether these distinct regions associated with decision confidence reflect the confidence estimate as one entity, or whether activation in these regions reflect different information subprocesses that contribute to the confidence estimate. Therefore, we are asking the question, whether these distinct sets of brain regions relevant for decision confidence are linked by brain co-activation, which would partially explain incongruent findings. Such a co-activation-focused view on brain activity may help set apart distinct correlates that covary with decision confidence.

Generally, confidence measures serve as good proxies for estimating the degree of decision accuracy in healthy subjects (Kunimoto et al., 2001; Kepecs and Mainen, 2012; Pouget et al., 2016). However, individuals vary in their ability to accurately judge their performance. Overall self-confidence may bias an individual to over- or underestimate discrimination accuracy, which generates a bias in metacognition (Washburn et al., 2005; Maniscalco and Lau, 2012; Fleming and Lau, 2014). For example, subjects experiencing relative blindsight accurately detect and discriminate between visual stimuli, yet they underestimate their task performance and report low confidence ratings (Lau and Passingham, 2006; Silvanto, 2015). The degree by which an observer's confidence ratings distinguish between incorrect and correct decisions is also confounded by the difficulty of the discrimination task (Maniscalco and Lau, 2012; Fleming and Lau, 2014; Boldt et al., 2017). When measuring visual metacognition, it is important to account for individuals' task performance. Maniscalco and Lau (2012) developed a measure, called metacognitive efficiency, which measures the accuracy of the metacognitive process itself. More specifically, metacognitive efficiency quantifies a subject's ability to accurately judge incorrect and correct decisions through confidence ratings given their task performance and reflects intrinsic evaluative processes. Application of the metacognitive efficiency in fMRI studies revealed weak correlations between individual metacognitive efficiency values and BOLD activity in frontal and prefrontal areas as well as left temporal gyri during visual perception (Fleming and Dolan, 2012; McCurdy et al., 2013; Dolan et al., 2018). However, the underlying neural architecture of the association between visual confidence and metacognitive efficiency remains unclear. Therefore, we further ask how the co-activation of confidence correlates with metacognitive efficiency to identify regions that correlate with intrinsic processing of metacognitive judgments independent from task performance.

In the current study, we addressed this question, by implementing a backward-masked visual detection paradigm with post-decision confidence ratings in healthy humans using concurrent task-fMRI. Confidence ratings and metacognitive efficiency values were used as proxies for visual metacognition. Co-varying patterns of blood oxygenation served as a proxy for co-activation of neural regions dependent on decision confidence.

## MATERIALS AND METHODS

### Participants

Thirty-five healthy, young subjects (24 females, mean age = 25.33, SD =  $\pm 3.00$ ) were initially recruited for the study. Testing was stopped pre-emptively in eight subjects who failed to perform above chance during training blocks on the behavioral task. One subject was excluded from the study due to brain abnormalities and two participants were excluded from the analyses because data acquisition could not be completed due to technical problems. Twenty-four subjects completed the study (19 females, mean age = 25.25 years, SD =  $\pm 3.19$  years) and were included in the analyses. All subjects had a normal or corrected-to-normal vision. Written consent was obtained from all subjects. The study was approved by the in-house ethics review committee at the TUM School of Medicine at the Technical University of Munich.

### Behavioral Task

A backward-masked visual perception task (**Figure 1A**) was adapted from (Wohlschläger et al., 2016; Glim et al., 2020; see also, Haynes et al., 2005) to measure visual confidence. fMRI data were concurrently recorded during the task. Participants fixated on a white cross on black background throughout the experiment. In each trial, the target stimuli were presented for 34 ms in the subjects' left visual field. The target appeared at a visual angle of  $9.7^\circ - 13.4^\circ$  from the fixation cross in the left visual field. After an interstimulus interval of 67 ms, a color-inverted mask of the target stimuli was presented for 17 ms. The target consisted of an 18-facet hexagonal honeycomb structure with a hexagonal gap either at the top or bottom of the stimuli. Subjects were asked to indicate the location of the gap (top or bottom) with one button press and give a post-decision confidence rating [very sure (VS), quite sure (QS), slightly sure (SS) and not sure (NS)] with a second button press. Key presses were performed on a four-button response box. Asynchronous inter-trial time intervals were randomized between 6.021 ms, 8.028 ms, and 10.035 ms (being multiples of the TR = 2.007 s) with the number of trials per block being 34/100, 33/100, and 33/100, respectively. To control for target position in the hemifield, trials were randomized between two positions in the subjects upper and lower left visual field, respectively. The behavioral task was delivered through the Presentation Software (Neurobehavioral Systems Inc., Berkeley, CA, USA). To familiarize themselves with the task, subjects underwent two training blocks of 50 trials within the MRI-scanner in which subjects had to indicate the gap location. During the first training run, responses were indicated as either correct or incorrect by a green or red signal. Subjects did not have to rate confidence. The second training block was identical to the experimental runs, in which subjects had to give a confidence rating after the target decision. All training blocks were conducted on a separate day outside of the scanner. Testing was stopped for subjects that failed to perform above chance during training blocks. Four experimental blocks of 100 trials per block were pursued.

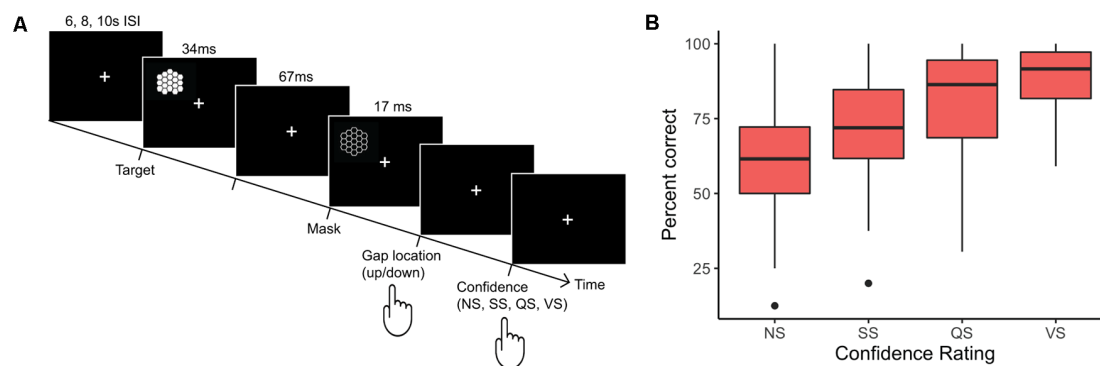
## Behavioral Outcome Measures

### Decision Confidence

Decision confidence was directly inferred from the subjects' second button presses (VS, QS, SS, NS). In all analyses, a response was considered invalid if participants responded before target onset, responded within 200 ms of target presentation, did not respond, or gave more than two answers. Response times were not restricted as subjects were asked to respond as accurately as possible. All trials containing eye movements, blinks, or artifacts within 200 ms of target onset were excluded from the analysis (see **Supplementary Methods** for identification of eye blinks). To validate the feasibility of the paradigm, the behavioral data had to fulfill two criteria: (i) Correct answers on the target property should be significantly above chance level. Using a two-tailed one-sample Student's *t*-test, objective task performance was tested for significance above chance level by comparing the number of correct to the total number of trials. (ii) Correctness should significantly increase with an increasing level of confidence. Trials were grouped by level of confidence rating (VS, QS, SS, NS). A larger proportion of correct trials should correspond to higher confidence ratings and task performance near chance should be associated with low confidence ratings. Again, a two-tailed one-sample Student's *t*-test was performed now for each confidence level. A multiway ANOVA including confidence rating, subjects, and fMRI session as factors and correctness of responses as the dependent variable was also performed across confidence levels, followed by subsequent two-tailed paired-sample Student's *t*-test between each confidence level. A significant increase in correct responses scaling with confidence conveys valid judgments.

### Metacognitive Efficiency

To evaluate how well subjects were able to use confidence ratings to accurately judge their task performance, an individual's metacognitive efficiency was calculated across the entire behavioral task. Task discrimination sensitivity ( $d'$ ), metacognitive sensitivity (meta- $d'$ ), and metacognitive efficiency (meta- $d'/d'$ ) were calculated as described by Maniscalco and Lau (2012). In brief, the decision process is separated into two levels: a first-order performance, which reflects how well the task could be solved (correct vs. incorrect responses), and a second-order performance which reflects the correctness of the subject's judgment on its first-order performance. Therefore, the absolute metacognitive sensitivity value (meta- $d'$ ) measures how much information about task performance is used to make a confidence rating. This measure is the ratio between observed metacognitive sensitivity (derived from an individual's decision confidence ratings across all trials) and expected metacognitive sensitivity, which estimates ideal confidence performance based on task performance. However, the meta- $d'$  measure is dependent on the quality of first-order performance, quantified by the discrimination sensitivity ( $d'$ ). The ratio between the discrimination sensitivity ( $d'$ ) and estimated meta- $d'$  can then be calculated to obtain metacognitive sensitivity value independent of response bias or task performance. The ratio (meta- $d'/d'$ ) is defined as metacognitive efficiency and



**FIGURE 1 |** Study trial design and results of the backward masking task for metacognitive confidence. **(A)** The visual backward masking task consisted of a hexagonal honeycomb with a hexagonal gap at either the top or bottom that was presented for 34 ms. A backward mask was presented 67 ms after the target presentation. Subjects had to indicate the gap location and then rate their decision confidence. Subjects were asked to continuously fixate on a fixation cross-present 2 s before and during the task. Targets were presented in the left visual field. **(B)** The distribution of percent correct trials averaged across subject and sessions for the confidence ratings not sure (NS), slightly sure (SS), quite sure (QS), and very sure (VS) are depicted. The horizontal lines indicate the median (NS = 61.54%, SS = 71.94%, QS = 86.36%, and VS = 91.58), the lower and upper hinges of the boxplots represent the first and third quartile, respectively. The lower and upper whisker indicates the minimum and maximum points less than 1.5 times the interquartile range (IQR). Outliers are indicated by dots and are greater than  $1.5 \times \text{IQR}$ . The mean percent correct for the four confidence ratings were  $59.1 \pm 13.8\%$  (NS),  $73.0 \pm 13.1\%$  (SS),  $80.9 \pm 15.4\%$  (QS), and  $88.42 \pm 9.4\%$  (VS) respectively. All confidence levels were significantly above chance ( $p < 0.01$ ). VS also was significantly higher than all other confidence levels ( $p < 0.01$ ).

measures the accuracy of the metacognitive process independent from the perceptual process (see Maniscalco and Lau, 2012, or <http://www.columbia.edu/bsm2105/type2sdt/archive/index.html> for more detailed information). We used the metacognitive efficiency ratio in our experiment to identify how much decision confidence conveyed a subject's metacognitive performance independent from task performance. Furthermore, a Pearson correlation between meta- $d'/d'$  and  $d'$  was also performed to ensure that indeed these measures were independent of each other in our data and described distinguishable processes.

## Imaging Data Acquisition and Processing

### Imaging Data Acquisition

Imaging data were acquired on a Phillips Ingenia 3T during two appointments. Appointment 1 included a structural MRI scan (MPRAGE, TE = 4 ms, TR = 9 ms, flip angle =  $8^\circ$ , FoV =  $240 \times 240$  mm, 340 slices, voxel size =  $0.5 \times 0.5 \times 0.5$  mm). Appointment 2 was focused on the fMRI scans of the masking paradigm. The fixation of the subjects was assessed from concurrently acquired electro-oculography EOG (see **Supplementary Methods** for more detail). Four-hundred-and-two whole-brain echo-planar imaging (EPI) scans were acquired for each of the four runs (TE = 30 ms, TR = 2.007 s, flip angle =  $80^\circ$ , FoV =  $192 \times 192$  mm, voxel size =  $3 \times 3 \times 3$  mm, 36 slices, slice thickness = 3 mm, no interslice gap).

### Imaging Data Processing

All fMRI data of the masking paradigm was processed and analyzed using statistical parametric mapping (SPM12<sup>1</sup>) and customized Matlab (MATLAB 2016b, The MathWorks, Inc., Natick, MA, USA) scripts. Slices were slice-time corrected,

realigned, and unwarped. The functional images were standardized to anatomical space by co-registering the T1-anatomical scan for each subject. The structural images co-registered to the functional images were segmented and spatially normalized to the ICBM space template of European brains. Functional images were normalized with identical transformations. Images were smoothed with an 8 mm Gaussian Kernel.

## Imaging Data Analysis

### Voxel-Wise Activation Analysis With Decision Confidence, General Linear Model Analysis

A two-stage, mixed-effect analysis was performed in SPM12. A single-subject, fixed effect model was implemented on the first-level using a GLM approach. To model any correlation between confidence and increase in BOLD responses, decision confidence levels were modeled in parametric regressors (paramDC) including only correct trials. Blood-oxygen-level dependent (BOLD) event-related responses were modeled as stick functions convolved with two basic functions in separate regressors: the canonical hemodynamic response function (HRF) and the first derivative in time (TD) of the canonical HRF, to account for possible differences in BOLD response timing. The temporal derivative of the canonical HRF by its shape is better suited to fit an early rise of the BOLD signal after a stimulus. In the following, a strong weight attached to this regressor by the GLM is interpreted as a strong early response. Movement parameters as well as incorrect responses (including eye movement and blinks) were modeled as nuisance regressors. On the second level, contrasts relating to the parametric modulation of BOLD activity dependent on confidence were compared across subjects for statistical significance in one-sample  $t$ -tests for the canonical HRF and the TD HRF, respectively. Clusters were regarded as significant at a threshold of  $p_{\text{FWE}} < 0.05$  family-wise error-

<sup>1</sup> [www.fil.ion.ucl.ac.uk/spm/](http://www.fil.ion.ucl.ac.uk/spm/)



corrected for multiple comparisons. The location of activation maps was determined with the Anatomy Toolbox (Eickhoff et al., 2005) within SPM.

### Voxel-Wise Activation Analysis With Metacognitive Efficiency, General Linear Model Analysis

On the second level, contrasts relating to the parametric modulation of canonical HRF and TD HRF were related to the metacognitive efficiency value for each subject, which was entered as a covariate into separate GLM analyses, respectively. Effects at cluster-level correction  $p_{cc} < 0.05$  with underlying voxel-level of  $p_{unc} < 0.001$ , uncorrected for multiple comparisons were regarded as significant.

### Co-activation Analysis of Decision Confidence, Cluster Analysis of Blood Oxygenation Covariation Patterns

In a second analysis, we focused on those regions of interest (ROIs) displaying a significant correlation with decision confidence. For these regions, defined as significant clusters with voxel numbers above 10 from the respective GLM analysis, we defined masks and extracted the within mask mean contrast values per subject of the contrast “paramDC\_TD”. This contrast quantifies the amount of parametric dependence of the early BOLD response on decision confidence. A spherical ROI (8 mm radius) within the pgACC (MNI:  $x/y/z = -2/44/10$ ), which has previously been shown to be involved in decision confidence (Bang and Fleming, 2018), was included. Within a given ROI, the variation of BOLD activity increase with decision confidence across subjects was quantified in a vector. Subsequent cluster analysis across ROI-vectors groups those assemblies of ROIs varying similarly across subjects and separated those ROIs varying independently across subjects. Cluster analysis was realized as  $k$ -means clustering with  $k = 3$  and 50 repetitions. A silhouette analysis was performed to assess the attribution of each ROI to the resulting ROI-clusters. A value of  $K = 3$  robustly produced identical clusters while higher values of  $K$  did not. Cluster centroids were subsequently correlated with the subjects' metacognitive efficiency values in Pearson correlations to indicate in which cluster the magnitude of the BOLD signal increasing with confidence correlated with metacognitive efficiency. Correlations were regarded as significant at  $p < 0.05$ , Bonferroni corrected for the number of clusters.

ROI-clusters were visualized in two ways based on Fisher-Z transformed correlation coefficients (FZCC) between ROIs across subjects. In more detail, two ROIs from within the same detected ROI-cluster show enhanced correlation (larger FZCC) of BOLD activity increase with decision confidence across subjects, vs. other ROI pairings. (i) These FZCC between any two ROIs placed on a glass brain were visualized through BrainNet Viewer (Xia et al., 2013) thresholded at 0.5. The line thickness of edges between two regions scale with FZCC. (ii) Alternatively, a visualization *via* multi-dimensional scaling is provided. ROIs are displayed as nodes, FZCCs as edges. In this representation, nodes are close to each other if they have a high correlation. It needs to be noted, that generally multi-dimensional scaling projection into the 2D image plane leads

to distortions, so only a general pattern is retained. In our display, node sizes scale with degree centrality, i.e., the average FZCC of one node vs. all others. In both representations, clusters are color-coded.

## RESULTS

### Decision Confidence Behavioral Outcomes

Decision confidence was computed by determining the proportion of correct trials per confidence rating. All of the behavioral statistics were computed by comparing trials of interest to the total number of valid trials. A one-sample  $t$ -test ( $T_{(24)} = 11.49$ ,  $p < 0.001$ ) showed that the mean percent of correct trials were 77.2% (SD  $\pm 18.7\%$ ) and were significantly above chance (corresponding to a mean of 50.0%). The percent of correct to valid trials for each confidence level was: very sure (VS)  $88.42 \pm 9.4\%$ , quite sure (QS)  $80.9 \pm 15.4\%$ , slightly sure (SS)  $73.0 \pm 13.1\%$ , and not sure (NS)  $59.1 \pm 13.8\%$ . One-sample  $t$ -tests demonstrating responses were significantly above chance for each confidence level (VS:  $T_{(24)} = 18.79$ ,  $p < 0.001$ ; QS:  $T_{(24)} = 9.83$ ,  $p < 0.001$ ; SS:  $T_{(24)} = 8.62$ ,  $p < 0.001$ ; NS:  $T_{(24)} = 3.23$ ,  $p = 0.002$ ). A multiway ANOVA including confidence rating, subjects, and fMRI session as factors showed the percent of correct trials across confidence levels was statistically significantly above chance ( $F_{(1,238)} = 94.77$ ,  $p < 0.0001$ ). A significant interaction ( $F_{(23,238)} = 3.7$ ,  $P < 0.0001$ ) was evident between confidence rating and subject indicating that confidence ratings varied significantly across subjects (Figure 1B, Supplementary Figure 1). Furthermore, paired  $t$ -tests were conducted between each confidence level across subjects. A significant increase in mean correct trials was evident with increasing confidence (VS vs. QS,  $p = 0.029$ ; VS vs. SS,  $p < 0.0001$ ; VS vs. NS,  $p < 0.0001$ ; QS vs. SS,  $p = 0.0012$ , QS vs. NS,  $p < 0.0001$ ; SS vs. NS,  $p < 0.0001$ , multiple comparison corrected for six comparisons across confidence pairs). The number of trials for each condition averaged across subjects were: VS = 58.13 (SE  $\pm 10.84$ ), QS = 95.42 (SE  $\pm 8.45$ ), SS = 88.29 (SE  $\pm 7.72$ ), NS = 55.96 (SE  $\pm 8.14$ ). Individual overall confidence responses varied with some subjects responding more conservatively and others more liberally (Figure 1B, Supplementary Figure 1). Four participants responded more conservatively, rating no trials with a very sure response, despite perceiving the target correctly. Since confidence was modeled parametrically and we were interested in looking at confidence variability, the subjects were not excluded from the analysis. No difference in significance in behavioral outcomes was observed when performing a secondary decision confidence behavioral analysis, which excluded the four conservative subjects from the analysis.

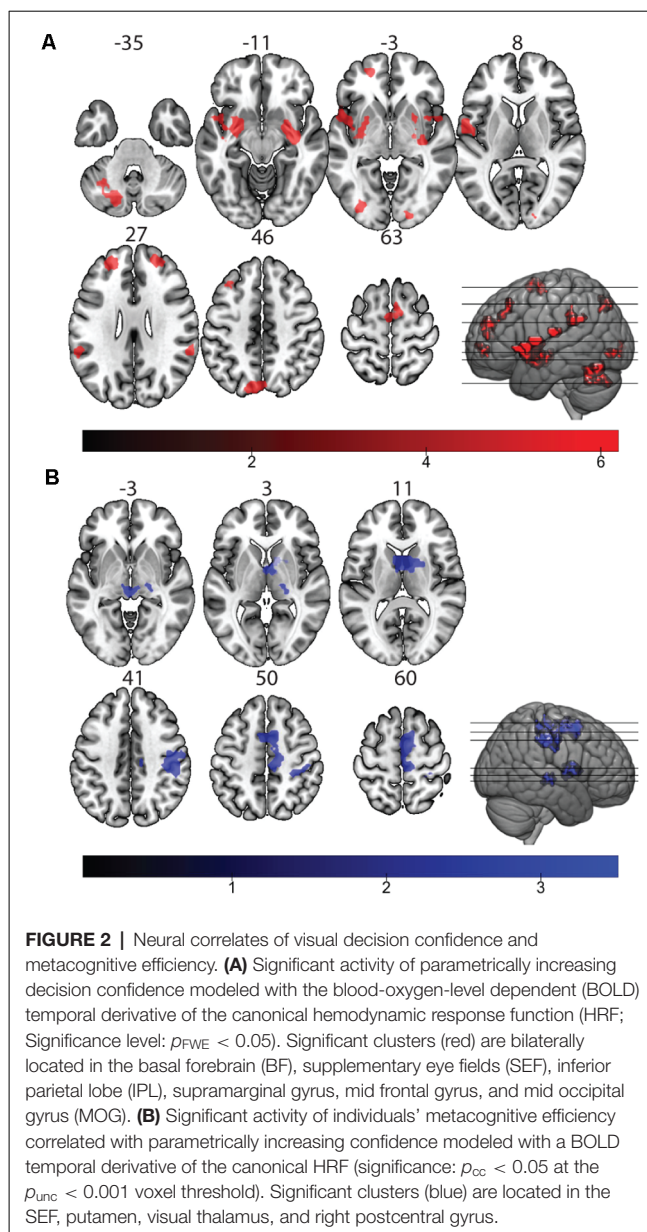
### Metacognitive Efficiency Behavioral Outcomes

To account for inter-subject variability in task discrimination performance, a ratio between metacognitive sensitivity and discrimination sensitivity (meta- $d'$ / $d'$ ) was calculated to measure how efficient trial-by-trial confidence ratings reflect subjects'

ability to accurately judge their performance relative to their type one response (Maniscalco and Lau, 2012; Fleming, 2017). This ratio is called metacognitive efficiency and measures the accuracy of the metacognitive process. Stimulus discrimination sensitivity ( $d'$ ) and metacognitive sensitivity (meta- $d'$ ) were calculated for each subject to compute the metacognitive efficiency ratio (see Maniscalco and Lau, 2012, for methods). Ideal metacognitive efficiency is achieved at a value of 1. A ratio below one suggests that only a portion of the sensory evidence was available for the metacognitive judgment due to sensory signal decay of accumulated noise. Therefore, the confidence rating less accurately predicts task performance accuracy. Whereas a ratio above one indicates subjects perform poorly on task discrimination, yet they are aware of their poor performance and can discriminate accordingly (Maniscalco and Lau, 2012; Fleming and Lau, 2014; Fleming and Daw, 2017). In this study, the metacognitive efficiency ratio was used to account for subject variability in task performance and the subject's overall confidence in the decision confidence measurements to evaluate unconfounded metacognitive performance. Subjects metacognitive efficiency averaged at 0.79 (range: 0.1–1.9; **Supplementary Figure 1**). A Pearson correlation between  $d'$  and meta- $d'/d'$  was performed to further verify that metacognitive efficiency was not biased by task performance. No significant correlation was observed ( $r_{(24)} = -0.3949$ ,  $p = 0.056$ ). The variability in the metacognitive efficiency ratio was then used to correlate inter-subject variability in BOLD activity to metacognitive ability.

## Brain Regions Displaying Significant Activation Dependent on Increasing Decision Confidence

A mixed-effects factorial design, with increasing confidence modeled parametrically, was used to assess differential activity correlating to degree of decision confidence. Significant activity scaling with confidence was observed in the left inferior parietal lobe (IPL), left anterior cingulate cortex (ACC), left middle frontal gyrus, caudate nucleus, and the mid orbital gyrus at a cluster-level corrected threshold of  $p_{cc} < 0.05$  (underlying voxel-level threshold of  $p_{unc} < 0.001$ , uncorrected; **Supplementary Figure 2, Supplementary Table 1**). No clusters survived at the family-wise-error-corrected voxel-level when modeled with the canonical HRF. We predicted that increasing confidence may not only result in increased amplitude of BOLD signal across confidence ratings but also result in a faster onset of the peak BOLD response. Incorporating the temporal derivative of the canonical HRF into the analysis allows for the detection of these fast hemodynamic responses. To also capture shifts in the onset of peak BOLD signal, we incorporated the temporal derivative of the canonical HRF to account for peak responses that slightly deviated in time (up to 1 s) from the canonical response curve (Friston et al., 1998; Henson et al., 2002). By far, the largest voxel clusters significantly scaling with confidence were found in the SEF (MNI:  $x/y/z = 6/2/65$ ) and bilateral nuclei Ch4 of the basal forebrain (BF; MNI:  $x/y/z = 27/-4/13$  and  $x/y/z = 27/-4/-10$ ) at a timing before the expected canonical hemodynamic responses



(Figure 2). Other areas showing a parametric dependence in this early phase of the hemodynamic response at  $p_{FWE} < 0.05$  were left parietal-temporal regions, right parietal areas, bilateral mid frontal gyrus, and Rolandic operculum (**Table 1**).

## BOLD Activity Corresponding to Individual Differences in Metacognitive Efficiency

To isolate metacognitive performance from task performance biases, a measure of metacognitive efficiency (meta- $d'/d'$ ) was calculated across the entire study for each subject. An analysis correlating subjects' metacognitive efficiency with BOLD activity showed that increased metacognitive efficiency is associated with stronger dependence of activity on confidence modeled with the canonical HRF and its temporal derivative in clusters including the SEF, putamen, visual thalamus, and right postcentral gyrus

**TABLE 1 |** Peak voxels of significant clusters for the temporal derivative of the canonical hemodynamic response function (HRF) contrast ( $p_{\text{FWE}} < 0.05$ ) with increasing confidence.

Region	MNI coordinates	Voxel-level z-score
R SEF	6/2/65	6.69
R BF	27/-4/-10	6.54
L BF	-27/-4/-13	6.30
L superior parietal lobe	-15/-82/47	6.08
R mid frontal gyrus	24/50/26	5.95
L mid occipital gyrus	-33/-82/-4	5.84
L cerebellum	-18/-76/-31	5.80
R mid temporal gyrus	57/-25/-10	5.78
L mid frontal gyrus	-18/47/26	5.73
L frontopolar cortex	-18/56/5	5.53
L rolandic operculum	-57/-1/8	5.50
L supramarginal gyrus	-63/-43/29	5.48
R supramarginal gyrus	63/-43/26	5.46
R cerebellum	27/-61/-34	5.41
L inferior parietal lobe	-60/-22/17	5.40
R V1	18/-91/-1	5.32
R rolandic operculum	51/5/-1	5.23
R mid occipital gyrus	36/-82/-4	5.10
R inferior parietal lobe	63/-16/14	5.07

( $p_{\text{cc}} < 0.05$ , cluster-level corrected on an underlying voxel-threshold of  $p_{\text{unc}} < 0.001$ , **Figure 2, Table 2**). Activity associated with increased confidence modeled with the canonical HRF contrast also correlated to increased metacognitive efficiency at the same threshold level in the bilateral inferior temporal lobe ( $p_{\text{cc}} < 0.05$  at  $p_{\text{unc}} < 0.001$ , **Supplementary Figure 2, Supplementary Table 2**). Notably, activity in the SEF scales with both decision confidence and metacognitive efficiency.

## Co-activation of Regions Scaling With Decision Confidence Across Subjects

We further investigated if co-activation between ROIs dependent on the degree of confidence in the early phase of the BOLD signal separated into segregable processes. Here, parametric modulation of the magnitude of the BOLD signal was used to assess co-activation with confidence across ROIs. The magnitude of the BOLD signal was quantified from the parameter estimates of the regressors parametric in confidence. Parameter estimates were obtained from regions significantly activated for parametric confidence modeled with the temporal derivatives of the canonical HRF contrast ( $p_{\text{FWE}} < 0.5$ ) from each subject. Co-activation with confidence was assessed through a correlational analysis between activated ROIs. The underlying model formulates that a high correlation between parameter estimates of two activated ROIs indicates BOLD activity

**TABLE 2 |** Peak voxels of significant clusters for the temporal derivative of the canonical HRF contrast with increasing confidence correlating to an increase in individuals' metacognitive efficiency ( $p_{\text{cc}} < 0.05$  at the  $p_{\text{unc}} < 0.001$  voxel threshold).

Region	MNI coordinates	Z-score
R Post central gyrus	45/-25/41	4.68
R SEF	3/2/59	4.62
L putamen	-6/2/14	4.51
R visual thalamus	21/-25/2	3.81

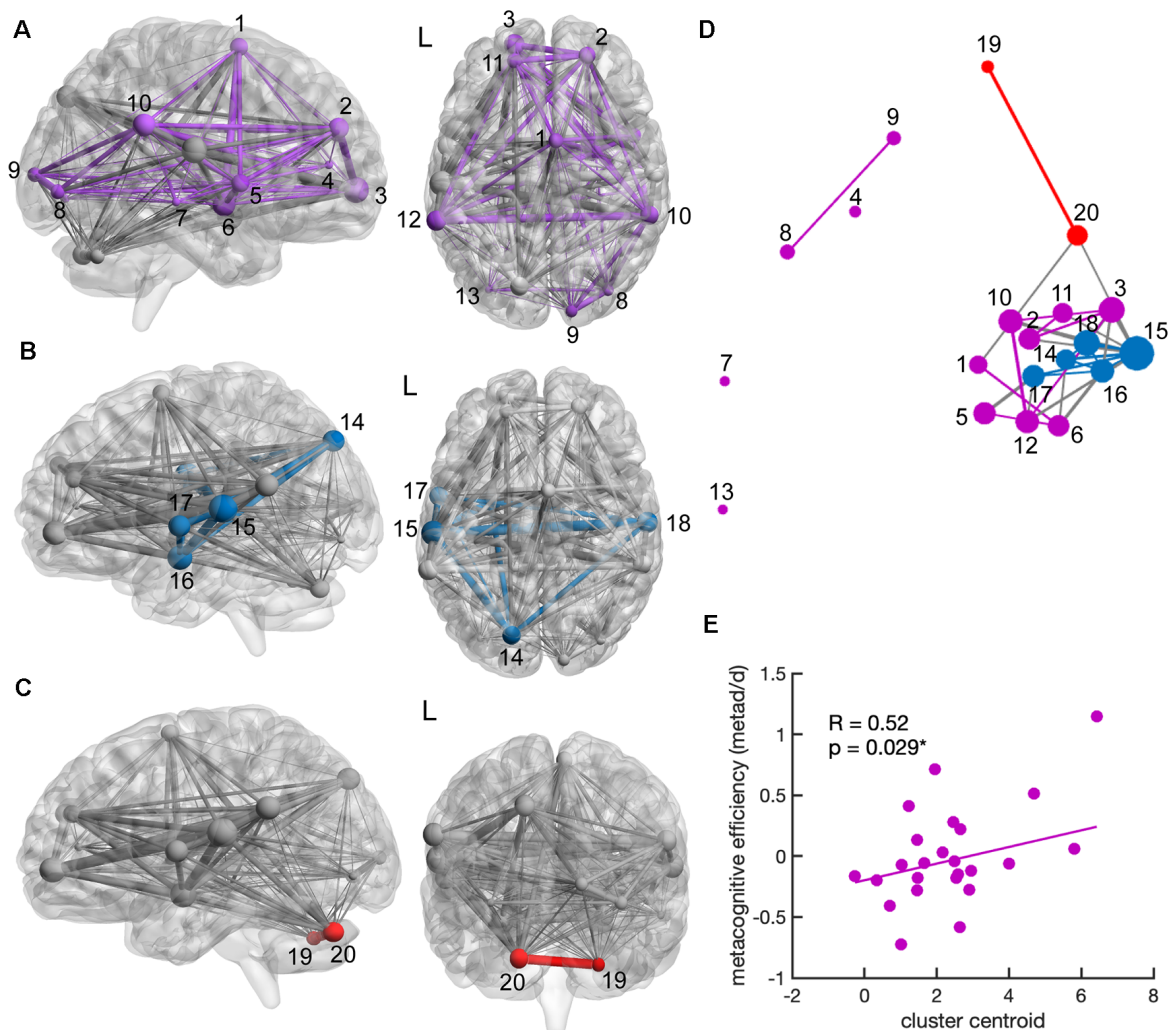
dependent on subjective confidence ratings is linked in both regions and thus suggesting a common process between the two regions. A low correlation coefficient indicates that the activity of region A changes relative to increasing confidence, however, this dependence was unrelated to the dependence on confidence in region B, and thus resulted from a different process. A spherical region around pgACC (MNI:  $x/y/z = -2/44/10$ ) from Bang and Fleming (2018) was included, which is associated with inter-subject decision confidence variability. **Figure 3** shows a visualization of the resulting connectivity between ROIs, and the attribution to three stable clusters (see silhouette analysis; **Supplementary Figure 3**), dissociating separable sets of confidence co-activation clusters: (1) a cluster centered in the right hemisphere (**Figure 3A**) involving the right fronto-temporal-parietal including bilateral frontopolar regions, mid frontal gyrus, supramarginal gyrus, SEF, and right basal forebrain; (2) a cluster centered in the left parieto-temporal hemisphere (**Figure 3B**) including the left Rolandic operculum, bilateral inferior parietal lobe (IPL), left superior parietal lobe (SPL) and left basal forebrain; and (3) a cerebellar cluster (**Figure 3C**). Interestingly, only in co-activation cluster 1 (purple cluster, **Figures 3A,E**) did signal magnitude increasing with confidence correlate with individual metacognitive efficiency ( $R = 0.52$ ,  $p = 0.029$ ).

## DISCUSSION

In this study, we were first to investigate co-activation patterns of BOLD activity associated with visual confidence and individuals' metacognitive efficiency during a backward-masked visual detection task with post-decision confidence ratings. By assessing the parametric dependence of confidence in the early phase of the hemodynamic response, we observed the strongest activation significantly scaling with increasing confidence in the supplementary eye field, bilaterally in the basal forebrain, and bilaterally in the parietal cortex and prefrontal cortex. To disentangle how the encoding of decision confidence gives rise to metacognitive performance, independent of task performance and response bias, we correlated individuals' overall metacognitive efficiency with confidence-dependent BOLD activity. The supplementary eye field, and activated region in the left postcentral gyrus expanding into the inferior parietal lobe, and activated regions in the striatum correlated with increased metacognitive efficiency. Notably, the supplementary eye field was the only region, whose activation correlated with decision confidence and metacognitive efficiency.

We then investigated how the emergence of confidence across distributed regions can be segregated into distinct co-activation processes varying with increasing decision with and correlated these co-activation clusters with individuals' metacognitive efficiency. We found decision confidence scaled in three segregable co-activation clusters: a right-centered fronto-parieto-temporal (**Figure 3A**), a left parieto-temporal (**Figure 3B**), and a cerebellar cluster (**Figure 3C**). In the right-centered cluster, the magnitude of BOLD signal dependent on confidence correlated strongly between the bilateral supramarginal gyrus, the bilateral mid frontal gyrus, and left frontopolar regions, the SEF, and the





1 - SEF(2), 2 - R MFG(3), 3 - L FrP(6), 4 - pgACC(0), 5 - R ROp(2), 6 - R BF(4), 7 - R MTG(1), 8 - RMOG(1), 9 - RV1(1), 10 - R SMG(4), 11 - L MFG(2), 12 - L SMG(5), 13 - L MOG(0), 14 - L SPL(4), 15 - L IPL(10), 16 - L BF(6), 17 - L ROp(4), 18 - R IPL(6), 19 - R cerebell(1), 20 - L cerebell(3)

**FIGURE 3 |** Co-activation clusters relevant to decision confidence and metacognitive efficiency. Co-activation of decision confidence was analyzed across subjects between voxel clusters significantly active at  $p_{FWE} < 0.05$  with increasing confidence in the early hemodynamic response phase [modeled with the derivative in time (TD) contrast]. The region pgACC (MNI:  $-2/44/10$ ) was added, as it has been associated with decision confidence (Bang and Fleming, 2018). Panels (A–C) show the locations of the three clusters with activity similarly covarying with confidence in MNI space. (A) The largest co-activation cluster (purple) was right-lateralized and included many frontoparietal regions: SEF, right BF, right rolandic operculum (ROp), bilateral midfrontal gyrus (MFG), left frontopolar cortex (FrP), right supramarginal gyrus (SMG), right mid temporal gyrus (MTG), as well as right V1 and bilateral mid occipital gyrus (MOG). (B) A second co-activation cluster (blue) was centered in the left temporal parietal region including the: left ROp, bilateral inferior parietal lobe (IPL), left superior parietal lobe (SPL) and left BF. (C) The third co-activation cluster (red) was centered in the cerebellum (cerebell). Edges as displayed were constructed from linear correlational analyses of the amount of dependence on confidence within the regions of interest (ROIs) across subjects. Edge size scale with higher correlation coefficients starting at a threshold of  $R > 0.5$ . Colored edges indicate the correlation degree within a co-activation cluster. Gray edges indicate the strength of the correlation coefficient between two activated regions obtained from the fMRI contrast maps. Node size indicates the degree of centrality. Clusters were visualized with the BrainNet Viewer (Xia et al., 2013, <http://www.nitrc.org/projects/bnv/>). (D) Multidimensional scaling onto a 2D space indicates the degree of co-activation of regions by spatial proximity. Reduction to 2D space leads to inaccuracies but gives a rough illustration of co-activation strength between ROIs. Degree centrality, i.e., the number of edges per node is noted in parentheses within the region labeling below the figure. (E) Only the right-centered co-activation cluster (purple cluster) scaled significantly with individual metacognitive efficiency ( $R = 0.52$ ,  $p_{(c)} < 0.005$ , corrected for multiple comparisons). The line indicates a regression between cluster centroids and metacognitive efficiency values (meta-d'/d').

right basal forebrain. In the left-centered co-activation cluster, signal magnitude relative to decision confidence correlated between the left inferior parietal lobe, the left basal forebrain,

left superior parietal lobe, left Rolandic operculum, and the right inferior parietal lobe. Only the right-centered confidence co-activation cluster including the supplementary eye field scaled



significantly with regions significant for individual metacognitive efficiency. This suggests that confidence arises in the cerebellum and two wide-scale networks in the brain. Yet metacognition, measured by the accuracy of judging one's own perceptual performance through confidence, may be encoded in the supplementary eye field through the emergence of decision confidence in right fronto-temporal-parietal regions.

Although metacognition is crucial in many cognitive processes, such as learning (Elwin et al., 2007; Folke et al., 2017), error monitoring (Boldt and Yeung, 2015; Fitzgerald et al., 2017), and decision making (Fleming, 2016; Van Den Berg et al., 2016; Qiu et al., 2018), it remains debated whether metacognition measured through decision confidence emerges from the same processing mechanism used to make the first-order, task-related decision (Pleskac and Busemeyer, 2010; Kiani et al., 2014; Van Den Berg et al., 2016) or whether second-order, metacognitive decisions occur independently of first-order decisions (Maniscalco and Lau, 2012, 2016; Fleming and Daw, 2017). A divide in the literature has associated parietal and oculomotor decision-making areas with the first theory and prefrontal and frontal executive areas with the latter (Grimaldi et al., 2015). However, the dispute over decision confidence may be explained when examining decision confidence as an entity that arises from the integration of multiple components (Pouget et al., 2016; Bang and Fleming, 2018). During task discrimination, confidence is generally described as a probabilistic computation of a decision being correct given available evidence (Pouget et al., 2016). How different types of evidence are integrated to form a probabilistic outcome remains debated. Our results show that the covariation of activity across three distinct clusters of brain regions becomes more robust with increasing confidence. We speculate that the co-variation of confidence within these three segregable clusters represents three information subprocesses that differentially contribute to the overall confidence estimate. As the activity in these brain regions begins to covary, integration of multiple subprocesses occurs, which influences the probabilistic computation of a decision confidence estimate. Only the right-lateralized frontoparietal network correlated with metacognitive efficiency. This suggests that co-activation of confidence in a distributed frontoparietal network reflects metacognitive processing.

Many previous studies have shown that the quality of sensory information can influence decision confidence (Yeung and Summerfield, 2012; Boldt et al., 2017). Manipulation of stimulus strength and stimulus reliability has been shown to influence metacognitive performance (Boldt et al., 2017), which suggests that first-order and second-order decisions share similar processing mechanisms (Yeung and Summerfield, 2012). In line with this theory, empirical studies in primates and humans showed areas in the parietal cortex to encode both first-order decisions and second-order decisions during a post-decision wagering task (Kiani and Shadlen, 2009; Gherman and Philastides, 2015). In our data, activation in the left and right IPL is significantly scaled with decision confidence. The left and right IPL were also part of a left-centered confidence co-activation cluster that also included the left SPL, basal

forebrain, and Rolandic operculum. The left-centered parietal co-activation cluster may therefore represent confidence arising in decision-making areas (Kiani and Shadlen, 2009; Kiani et al., 2014). However, this co-activation cluster did not correlate with overall metacognitive efficiency, suggesting that metacognition depends on the integration of additional information.

An alternative theory suggests that different sources of information and processing mechanisms lead to first-order and second-order decisions respectively. In a heuristic model, first-order and second-order processes occur separately in a serial manner, with noise from the first-order decision and other internal states or information being evaluated in the second-order decision (Maniscalco and Lau, 2016; Fleming and Daw, 2017). Therefore, metacognitive performance is correlated to but independent of task performance. Disassociation between task performance and metacognitive performance has been identified in blindsight (Lau and Passingham, 2006), and psychiatric disorders (Rouault et al., 2018), but also across healthy individuals (Washburn et al., 2005; Fleming et al., 2010; Maniscalco and Lau, 2012; Bang and Fleming, 2018). Although behaviorally confidence ratings correlated with task accuracy in our data, we found subjects varied in their ability to accurately judge their task performance through confidence ratings. Recent research has tried to identify what types of information and the corresponding brain areas influence metacognitive performance independent of task performance. Action-specific information about first order-decisions in the premotor cortex contributes to metacognitive judgments independently of first-order decisions (Fleming and Lau, 2014; Wokke et al., 2020). Furthermore, functional connectivity between prefrontal and motor areas increased between the first-order response and metacognitive judgment (Wokke et al., 2020). Besides perceptual evidence, global internal states influence metacognitive performance. Global estimates of self-performance seem to be influenced by previous confidence estimates and choices on preceding trials, even during the absence of feedback (Benwell et al., 2019; Rouault et al., 2019). Arousal (Allen et al., 2016; Hauser et al., 2017) and attention (Rahnev et al., 2011) also correlate with disassociation between metacognitive performance and task performance. In human fMRI studies, neural correlates in rostralateral PFC (Fleming and Dolan, 2012; Grimaldi et al., 2015), striatum (Hebart et al., 2014; Gherman and Philastides, 2018), medial prefrontal cortex (De Martino et al., 2013), dorsal anterior cingulate cortex (Fleming and Dolan, 2012), and other frontoparietal regions (Vaccaro and Fleming, 2018) have been associated with metacognitive performance. The right-centered frontoparietal co-activation network (**Figure 3A**) in our results relates to previous findings, showing that confidence covaries across frontal and prefrontal regions along with the SEF and basal forebrain. We incorporated a ROI in the pgACC, which has previously been shown to track expected performance (Bang and Fleming, 2018). However, in our results, the pgACC did not correlate strongly with the other ROIs. Rather inter-subject variability in metacognitive efficiency correlated most strongly with decision confidence in the SEF.

We also identified a third co-activation cluster that consisted of regions in the left and right cerebellum. Traditionally, the cerebellum is thought to monitor and evaluate movement control by dually encoding internal predictions and sensory feedback, and reporting a subsequent error (Schlerf et al., 2012; Schmammann, 2019). The cerebellum may likely have a similar role in other sensory modalities (Popa et al., 2014; Peterburs and Desmond, 2016) and cognitive processes (Stoodley et al., 2012). Error reporting, as well forward modeling, are elements that influence the degree of confidence. In this study, we are unable to differentiate whether the significant activity in the cerebellar co-activation cluster is related to confidence or other performance monitoring processes that contribute to decision confidence.

Our results provide a novel finding in humans in showing converging evidence that the supplementary eye fields are tightly coupled to the emergence of confidence and individuals' metacognitive efficiency in our visual discrimination task. Previously, the SEF has only been shown to be involved in performance monitoring, including error monitoring and conflict-monitoring, in visual-oculomotor tasks in electrophysiological (Stuphorn et al., 2000; Emeric et al., 2010) and fMRI studies (Nachev et al., 2005). Although our study did not include saccade-based decisions, the SEF also correlated with high metacognitive efficiency in our results. A high metacognitive efficiency value ( $\text{meta-}d'/d' > 1$ ) has been suggested to indicate error monitoring, as the individual performs poorly on task discrimination yet is aware of their poor performance (Fleming and Daw, 2017). Higher metacognitive ability may therefore arise from faster, more robust signals in the SEF during a backward-masked discrimination task. Furthermore, we performed a control analysis that excluded trials containing saccades within 200 ms of target presentation (see the **Supplementary Information** for more detail) to exclude possible confounding effects of saccades on our results. The SEF also has been implicated in duration estimates of timing during uncertain tasks (Cui et al., 2009), however, a subsequent analysis of confidence across different inter-stimulus time intervals revealed no significant pattern in our results (see **Supplementary Information**). Therefore, our results suggest the activation of the SEF may be domain-general across different sensory modalities and metacognitive tasks. The SEF may serve as a hub for monitoring and evaluating local estimates of confidence arising from different types of information across distributed brain regions, leading to an overall global estimate of metacognitive ability.

It should be noted that the metacognitive efficiency ratios ( $\text{meta-}d'/d'$ ) varied notably across subjects. A Pearson correlation performed between metacognitive efficiency ( $\text{meta-}d'/d'$ ) and task discrimination sensitivity ( $d'$ ) was trending towards significance ( $p = 0.056$ ). We interpreted the non-significant correlation between metacognitive efficiency and discrimination sensitivity to signify independence in variability between the two measures across subjects. Because of the small subject number, the power of the study may not be

strong enough to fully exclude a possible correlation between metacognitive efficiency ( $\text{meta-}d'/d'$ ) and task discrimination ( $d'$ ). In the case that the correlation would be significant, the value obtained for metacognitive efficiency would still take into account task performance bias, however, metacognitive performance would scale with task performance. We would interpret this to mean that the variability in metacognitive efficiency measures arises from variability in task performance across subjects rather than variability in both task and metacognitive performance. Concerning our findings, the frontoparietal network would still reflect activation patterns correlated with increasing metacognitive efficiency that is distinct from other processes leading to decision confidence. However, we would not be able to differentiate variability in metacognitive performance from the variability in task performance across subjects.

Our study is limited to retrospective confidence ratings in metacognitive perception. Metacognition can be reflected in different introspective processes such as decision uncertainty, reward expectancies, and error-monitoring (Vaccaro and Fleming, 2018). Domain-general and domain-specific regions have also been identified in visual metacognitive memory and visual perception tasks (McCurdy et al., 2013; Morales et al., 2018; Vaccaro and Fleming, 2018). Behavioral measures of metacognitive efficiency also correlate across different sensory modalities (Faivre et al., 2018; Beck et al., 2019), yet the neural correlates of decision confidence and metacognitive efficiency across sensory modalities remain elusive. Future studies are needed to understand how confidence estimates arise across different metacognitive tasks and whether decision confidence estimates covary in domain-general or domain-specific patterns across metacognitive task types.

In conclusion, we found activity correlated significantly with visual confidence in the supplementary eye field, bilateral basal forebrain, prefrontal, and parietal regions. We accounted for individual differences in metacognitive performance by including inter-subject variability in metacognitive efficiency in our fMRI analysis, which further confirmed the involvement of the SEF in visual metacognition. The SEF has been implicated in performance and error monitoring in non-human primate studies. Our results are the first to show the supplementary eye field's role in metacognition in humans during visual perception. We implemented a novel approach, examining confidence dependent co-activation across brain regions that correlated significantly with decision confidence and found three segregable clusters: a right-centered frontoparietal cluster, a left-centered temporal-parietal cluster, and a cerebellar cluster. Only the right-centered co-activation cluster, which included the supplementary eye field, correlated with increased metacognitive efficiency. These findings suggest distinct information processes correlate with decision confidence, yet only the right-centered frontoparietal co-activation cluster reflects metacognitive ability. Our results are in line with the view that confidence estimates arise from the integration of multiple components. Distinct components are thought to be processed through different pathways, with only some processing pathways reflecting metacognitive ability.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The study involving human participants was reviewed and approved by TUM School of Medicine at the Technical University of Munich. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

AW and CS oversaw research. SG, CS, and AW designed research. SG and AR performed experiments. CJ, CD, SG, and AW analyzed data. CJ, CS, and AW interpreted results of experiments. CJ drafted manuscript,

prepared figures, and edited and revised manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

AR was supported by the Studienstiftung des deutschen Volkes.

## ACKNOWLEDGMENTS

We thank Dr. Franziska Knolle for her helpful suggestions.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/10.3389/fnys.2020.557693/full#supplementary-material>.

## REFERENCES

- Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., et al. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *eLife* 5:e18103. doi: 10.7554/eLife.18103
- Bang, D., and Fleming, S. M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. *Proc. Natl. Acad. Sci. U S A* 115, 6082–6087. doi: 10.1073/pnas.1800795115
- Beck, B., Peña-Vivas, V., Fleming, S., and Haggard, P. (2019). Metacognition across sensory modalities: vision, warmth, and nociceptive pain. *Cognition* 186, 32–41. doi: 10.1016/j.cognition.2019.01.018
- Benwell, C. S. Y., Beyer, R., Wallington, F., and Ince, R. A. A. (2019). History biases reveal novel dissociations between perceptual and metacognitive decision-making. *BioRxiv* [Preprint]. doi: 10.1101/737999
- Boldt, A., de Gardelle, V., and Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 1520–1531. doi: 10.1037/xhp0000404
- Boldt, A., and Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *J. Neurosci.* 35, 3478–3484. doi: 10.1523/JNEUROSCI.0797-14.2015
- Cui, X., Stetson, C., Read Montague, P., and Eagleman, D. M. (2009). ReadyandGo: amplitude of the fMRI signal encodes expectation of cue arrival time. *PLoS Biol.* 7:e1000167. doi: 10.1371/journal.pbio.1000167
- Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., and Slachetky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* 132, 2531–2540. doi: 10.1093/brain/awp111
- De Martino, B., Fleming, S. M., Garrett, N., and Dolan, R. J. (2013). Confidence in value-based choice. *Nat. Neurosci.* 16, 105–110. doi: 10.1038/nn.3279
- Dolan, M.-J., Belliart-Guérin, G., Bates, A. S., Frechter, S., Lampin-Saint-Amaux, A., and Aso, Y. (2018). Communication from learned to innate olfactory processing centers is required for memory retrieval in *Drosophila*. *Neuron* 100, 651.e8–668.e8. doi: 10.1016/j.neuron.2018.08.037
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034
- Elwin, E., Juslin, P., Olsson, H., and Enkvist, T. (2007). Constructivist coding: learning from selective feedback. *Psychol. Sci.* 18, 105–110. doi: 10.1111/j.1467-9280.2007.01856.x
- Emeric, E. E., Leslie, M., Pouget, P., and Schall, J. D. (2010). Performance monitoring local field potentials in the medial frontal cortex of primates: supplementary eye field. *J. Neurophysiol.* 104, 1523–1537. doi: 10.1152/jn.01001.2009
- Faivre, N., Filevich, E., Solovey, G., Kühn, S., and Blanke, O. (2018). Behavioral, modeling and electrophysiological evidence for supramodality in human metacognition. *J. Neurosci.* 38, 263–277. doi: 10.1523/JNEUROSCI.0322-17.2017
- Fitzgerald, L. M., Arvaneh, M., and Dockree, P. M. (2017). Domain-specific and domain-general processes underlying metacognitive judgments. *Conscious. Cogn.* 49, 264–277. doi: 10.1016/j.concog.2017.01.011
- Fleming, S. M. (2016). Changing our minds about changes of mind. *eLife* 5:e14790. doi: 10.7554/eLife.14790
- Fleming, S. M. (2017). HMeta-d: hierarchical bayesian estimation of metacognitive efficiency from confidence ratings. *Neurosci. Conscious.* 2017:nix007. doi: 10.1093/nc/nix007
- Fleming, S. M., and Daw, N. D. (2017). Self-evaluation of decision-making: a general bayesian framework for metacognitive computation. *Psychol. Rev.* 124, 91–114. doi: 10.1037/rev0000045
- Fleming, S. M., and Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1338–1349. doi: 10.1098/rstb.2011.0417
- Fleming, S. M., Huijgen, J., and Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *J. Neurosci.* 32, 6117–6125. doi: 10.1523/JNEUROSCI.6489-11.2012
- Fleming, S. M., and Lau, H. C. (2014). How to measure metacognition. *Front. Hum. Neurosci.* 8:443. doi: 10.3389/fnhum.2014.00443
- Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., and Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science* 329, 1541–1543. doi: 10.1126/science.1191883
- Folke, T., Jacobsen, C., Fleming, S. M., and De Martino, B. (2017). Explicit representation of confidence informs future value-based decisions. *Nat. Hum. Behav.* 1:2. doi: 10.1038/s41562-016-0002
- Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. (1998). Event-related fMRI: characterizing differential responses. *NeuroImage* 7, 30–40. doi: 10.1006/nimg.1997.0306
- Gherman, S., and Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *NeuroImage* 106, 134–143. doi: 10.1016/j.neuroimage.2014.11.036
- Gherman, S., and Philiastides, M. G. (2018). Human VMPFC encodes early signatures of confidence in perceptual decisions. *eLife* 7:e38293. doi: 10.7554/eLife.38293
- Glim, S., Ries, A., Sorg, C., and Wohlschläger, A. M. (2020). The temporal evolution of pre-stimulus slow cortical potentials is associated with an upcoming stimulus' access to visual consciousness. *Conscious. Cogn.* 84:102993. doi: 10.1016/j.concog.2020.102993
- Grimaldi, P., Lau, H., and Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not

- know: confidence in decision-making. *Neurosci. Biobehav. Rev.* 55, 88–97. doi: 10.1016/j.neubiorev.2015.04.006
- Hauser, T. U., Allen, M., Purg, N., Moutoussis, M., Rees, G., and Dolan, R. J. (2017). Noradrenaline blockade specifically enhances metacognitive performance. *eLife* 6:e24901. doi: 10.7554/eLife.24901
- Haynes, J.-D., Driver, J., and Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. *Neuron* 46, 811–821. doi: 10.1016/j.neuron.2005.05.012
- Hebart, M. N., Schriever, Y., Donner, T. H., and Haynes, J.-D. (2014). The relationship between perceptual decision variables and confidence in the human brain. *Cereb. Cortex* 26, 118–130. doi: 10.1093/cercor/bhu181
- Henson, R. N. A., Price, C. J., Rugg, M. D., Turner, R., and Friston, K. J. (2002). Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. *NeuroImage* 15, 83–97. doi: 10.1006/nimg.2001.0940
- Kepecs, A., and Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1322–1337. doi: 10.1098/rstb.2012.0037
- Kiani, R., Corthell, L., and Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron* 84, 1329–1342. doi: 10.1016/j.neuron.2014.12.015
- Kiani, R., and Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324, 759–764. doi: 10.1126/science.1169405
- Kunimoto, C., Miller, J., and Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Conscious. Cogn.* 10, 294–340. doi: 10.1006/ccog.2000.0494
- Lau, H. C., and Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc. Natl. Acad. Sci. U S A* 103, 18763–18768. doi: 10.1073/pnas.0607716103
- Mamassian, P. (2016). Visual confidence. *Annu. Rev. Vis. Sci.* 2, 459–481. doi: 10.1146/annurev-vision-111815-114630
- Maniscalco, B., and Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious. Cogn.* 21, 422–430. doi: 10.1016/j.concog.2011.09.021
- Maniscalco, B., and Lau, H. (2016). The signal processing architecture underlying subjective reports of sensory awareness. *Neurosci. Conscious.* 2016:niw002. doi: 10.1093/nc/niw002
- McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., and Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *J. Neurosci.* 33, 1897–1906. doi: 10.1523/JNEUROSCI.1890-12.2013
- Metcalfe, J., and Shimamura, A. P. (1994). *Metacognition: Knowing About Knowing*. Cambridge, MA: The MIT Press.
- Middlebrooks, P. G., and Sommer, M. A. (2012). Neuronal correlates of metacognition in primate frontal cortex. *Neuron* 75, 517–530. doi: 10.1016/j.neuron.2012.05.028
- Morales, J., Lau, H., and Fleming, S. M. (2018). Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. *J. Neurosci.* 38, 3534–3546. doi: 10.1523/JNEUROSCI.2360-17.2018
- Nachev, P., Rees, G., Parton, A., Kennard, C., and Husain, M. (2005). Volition and conflict in human medial frontal cortex. *Curr. Biol.* 15, 122–128. doi: 10.1016/j.cub.2005.01.006
- Pasquali, A., Timmermans, B., and Cleeremans, A. (2010). Know thyself: metacognitive networks and measures of consciousness. *Cognition* 117, 182–190. doi: 10.1016/j.cognition.2010.08.010
- Peterburs, J., and Desmond, J. E. (2016). The role of the human cerebellum in performance monitoring. *Curr. Opin. Neurobiol.* 40, 38–44. doi: 10.1016/j.conb.2016.06.011
- Pleskac, T. J., and Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time and confidence. *Psychol. Rev.* 117, 864–901. doi: 10.1037/a0019737
- Popa, L. S., Hewitt, A. L., and Ebner, T. J. (2014). The cerebellum for jocks and nerds alike. *Front. Syst. Neurosci.* 8:113. doi: 10.3389/fnsys.2014.00113
- Pouget, A., Drugowitsch, J., and Kepecs, A. (2016). Confidence and certainty: distinct probabilistic quantities for different goals. *Nat. Neurosci.* 19, 366–374. doi: 10.1038/nn.4240
- Qiu, L., Su, J., Ni, Y., Bai, Y., Zhang, X., Li, X., et al. (2018). The neural system of metacognition accompanying decision-making in the prefrontal cortex. *PLoS Biol.* 16:e2004037. doi: 10.1371/journal.pbio.2004037
- Rahnev, D., Maniscalco, B., Graves, T., Huang, E., De Lange, F. P., and Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nat. Neurosci.* 14, 1513–1515. doi: 10.1038/nn.2948
- Rouault, M., Dayan, P., and Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nat. Commun.* 10:1141. doi: 10.1038/s41467-019-09075-3
- Rouault, M., Seow, T., Gillan, C. M., and Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biol. Psychiatry* 84, 443–451. doi: 10.1016/j.biopsych.2017.12.017
- Schlerf, J., Ivry, R. B., and Diedrichsen, J. (2012). Encoding of sensory prediction errors in the human cerebellum. *J. Neurosci.* 32, 4913–4922. doi: 10.1523/JNEUROSCI.4504-11.2012
- Schmahmann, J. D. (2019). The cerebellum and cognition. *Neurosci. Lett.* 688, 62–75. doi: 10.1016/j.neulet.2018.07.005
- Shimamura, A. P. (2000). Toward a cognitive neuroscience of metacognition. *Conscious. Cogn.* 9, 313–323. doi: 10.1006/ccog.2000.0450
- Silvanto, J. (2015). Why is 'blindsight' blind? a new perspective on primary visual cortex, recurrent activity and visual awareness. *Conscious. Cogn.* 32, 15–32. doi: 10.1016/j.concog.2014.08.001
- So, N., and Stuphorn, V. (2015). Supplementary eye field encodes confidence in decisions under risk. *Cereb. Cortex* 26, 764–782. doi: 10.1093/cercor/bhv025
- Stoodley, C. J., Valera, E. M., and Schmahmann, J. D. (2012). Functional topography of the cerebellum for motor and cognitive tasks: an fMRI study. *NeuroImage* 59, 1560–1570. doi: 10.1016/j.neuroimage.2011.08.065
- Stuphorn, V., Taylor, T. L., and Schall, J. D. (2000). Performance monitoring by the supplementary eye field. *Nature* 408, 857–860. doi: 10.1038/35048576
- Vaccaro, A. G., and Fleming, S. M. (2018). Thinking about thinking: a coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain Neurosci. Adv.* 2:2398212818810591. doi: 10.1177/2398212818810591
- Van Den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., and Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence. *eLife* 5:e12192. doi: 10.7554/eLife.12192
- Washburn, D. A., David Smith, J., and Taglialetta, L. A. (2005). Individual differences in metacognitive responsiveness: cognitive and personality correlates. *J. Gen. Psychol.* 132, 446–461. doi: 10.3200/GENP.132.4.446-461
- Wohlschläger, A. M., Glim, S., Shao, J., Draheim, J., Köhler, L., Lourenço, S., et al. (2016). Ongoing slow fluctuations in V1 impact on visual perception. *Front. Hum. Neurosci.* 10:411. doi: 10.3389/fnhum.2016.00411
- Wokke, M. E., Achoui, D., and Cleeremans, A. (2020). Action information contributes to metacognitive decision-making. *Sci. Rep.* 10:3632. doi: 10.1038/s41598-020-60382-y
- Yeung, N., and Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1310–1321. doi: 10.1098/rstb.2011.0416
- Xia, M., Wang, J., and He, Y. (2013). BrainNet viewer: a network visualization tool for human brain connectomics. *PLoS One* 8:e68910. doi: 10.1371/journal.pone.0068910

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Jaeger, Glim, Dimulescu, Ries, Sorg and Wohlschläger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Detecting (Un)seen Change: The Neural Underpinnings of (Un)conscious Prediction Errors

Elise G. Rowe<sup>1,2,3,4\*</sup>, Naotsugu Tsuchiya<sup>1,2,5,6,7†</sup> and Marta I. Garrido<sup>3,4,7,8†</sup>

<sup>1</sup> School of Psychological Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, VIC, Australia, <sup>2</sup> Turner Institute for Brain and Mental Health, Monash University, Clayton, VIC, Australia, <sup>3</sup> Queensland Brain Institute, The University of Queensland, Saint Lucia, QLD, Australia, <sup>4</sup> Centre for Advanced Imaging, The University of Queensland, Saint Lucia, QLD, Australia, <sup>5</sup> Center for Information and Neural Networks, National Institute of Information and Communications Technology, Suita, Japan, <sup>6</sup> Advanced Telecommunications Research Computational Neuroscience Laboratories, Kyoto, Japan, <sup>7</sup> ARC Centre of Excellence for Integrative Brain Function, Clayton, VIC, Australia, <sup>8</sup> Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, VIC, Australia

## OPEN ACCESS

### Edited by:

Olivia Gosseries,  
University of Liège, Belgium

### Reviewed by:

Jitendra Sharma,  
Massachusetts Institute  
of Technology, United States  
Robert O'Shea,  
Southern Cross University, Australia

### \*Correspondence:

Elise G. Rowe  
elise.rowe@monash.edu

† These authors have contributed  
equally to this work

### \*ORCID:

Naotsugu Tsuchiya  
orcid.org/0000-0003-4216-8701  
Marta I. Garrido  
orcid.org/0000-0003-0679-4959

**Received:** 10 March 2020

**Accepted:** 08 October 2020

**Published:** 11 November 2020

### Citation:

Rowe EG, Tsuchiya N and  
Garrido MI (2020) Detecting (Un)seen  
Change: The Neural Underpinnings  
of (Un)conscious Prediction Errors.  
Front. Syst. Neurosci. 14:541670.  
doi: 10.3389/fnsys.2020.541670

Detecting changes in the environment is fundamental for our survival. According to predictive coding theory, detecting these irregularities relies both on incoming sensory information and our top-down prior expectations (or internal generative models) about the world. Prediction errors (PEs), detectable in event-related potentials (ERPs), occur when there is a mismatch between the sensory input and our internal model (i.e., a surprise event). Many changes occurring in our environment are irrelevant for survival and may remain unseen. Such changes, even if subtle, can nevertheless be detected by the brain without emerging into consciousness. What remains unclear is how these changes are processed in the brain at the network level. Here, we used a visual oddball paradigm in which participants engaged in a central letter task during electroencephalographic (EEG) recordings while presented with task-irrelevant high- or low-coherence background, random-dot motion. Critically, once in a while, the direction of the dots changed. After the EEG session, we confirmed that changes in motion direction at high- and low-coherence were visible and invisible, respectively, using psychophysical measurements. ERP analyses revealed that changes in motion direction elicited PE regardless of the visibility, but with distinct spatiotemporal patterns. To understand these responses, we applied dynamic causal modeling (DCM) to the EEG data. Bayesian Model Averaging showed visible PE relied on a release from adaptation (repetition suppression) within bilateral MT+, whereas invisible PE relied on adaptation at bilateral V1 (and left MT+). Furthermore, while feedforward upregulation was present for invisible PE, the visible change PE also included downregulation of feedback between right MT+ to V1. Our findings reveal a complex interplay of modulation in the generative network models underlying visible and invisible motion changes.

**Keywords:** consciousness, prediction errors, DCM, vMMN, EEG

## INTRODUCTION

Detecting changes in the environment is fundamental for our survival because these may indicate potential rewards or threats (LeDoux, 1996; Panksepp, 1998; Rensink, 2002). According to predictive coding theory, we detect these irregularities by comparing incoming (bottom-up) sensory information with an internal model based either on prior sensory information or on (top-down) beliefs or expectations (Friston, 2005; Friston and Stephan, 2007; Clark, 2013; Hohwy, 2013). Mismatch between the sensory input and our internal model results in a surprise event, signaled by a prediction error (PE) response in event-related potentials (ERPs). These PE responses represent the difference in the neural activation between expected (frequent or 'standard') and unexpected (rare/surprising or 'deviant') events and have been used extensively in the auditory modality to yield the mismatch negativity (MMN; Näätänen et al., 1978; Näätänen, 1992; Garrido et al., 2007). Over the last decade, protocols for inducing visual PE (or visual MMN, 'vMMN'; more specifically) have been developed, with a focus on situations where individuals are aware of a change (Pazo-Alvarez et al., 2003; Stefanics et al., 2014; Kremláček et al., 2016). In our everyday lives, however, many changes in the sensory environment that evoke PEs may go unnoticed (Czigler, 2014; Stefanics et al., 2014). What remains unclear is how such sensory changes are processed by the brain at the network level.

One of the advantages in investigating non-conscious PE using visual, compared to auditory, stimuli is a richer set of psychophysical methods available to render otherwise salient visual stimuli invisible to participants (Faivre et al., 2017). Employing such tools, researchers have demonstrated PE responses to changes without conscious awareness of these changes. These PE responses include those using backward masking (Czigler et al., 2007; Kogai et al., 2011), binocular rivalry (Jack et al., 2015, 2017), rapid unmasked presentation (1 ms, Bernat et al., 2001, and 10 ms; Brazdil et al., 2001), and attentional blink (Berti, 2011). The researchers isolated PE ERPs by comparisons between standard and deviant trials, finding stronger and more widespread activation associated with conscious than non-conscious visual PE. ERP analyses alone, however, cannot inform us of the underlying neurocircuitry underpinning the PE to changes that do, and do not, emerge into consciousness.

Dynamic causal modeling (DCM) is a technique useful for inferring the underlying causal network of dynamical systems (Friston et al., 2003). Boly et al. (2011), applied DCM to auditory PEs in EEG recorded from healthy participants and two populations of brain damaged patients: minimally conscious and vegetative state patients to show that feedback connectivity (or top-down modulation) was reduced in unconscious vegetative patients compared to minimally conscious patients (and healthy controls). This study suggests a potential involvement of feedback modulations in regulating the level of conscious awareness of PE generating stimuli. However, no studies have performed network level examination of non-conscious processing in fully awake, healthy participants and whether this also abolishes top-down feedback connections.

To understand the differences in the neural circuitry between sensory changes that can and cannot be consciously perceived, we aimed to elicit visual PE using visible and invisible changes in motion direction. To achieve the desired level of visibility of motion stimuli, we manipulated the coherence of random-dot motion. We used DCM to estimate effective connectivity and examined the involvement of feedback connections for visible and invisible changes.

## MATERIALS AND METHODS

### Participants

Twenty-eight healthy university students participated in this study (10 females, aged 18–40,  $M = 22.44$ ,  $SD = 4.30$ ). All participants reported no history of neurological or psychiatric disorder or previous head trauma resulting in unconscious comatose states. All participants gave written informed consent in accordance with the guidelines of The University of Queensland's ethics committee.

### Experimental Design

Continuous EEG data were recorded during the main task using a Biosemi Active Two system with 64 Ag/AgCl scalp electrodes arranged according to the international 10–10 system. Data were recorded at a sampling rate of 1024 Hz. After the EEG was setup, participants practiced the 1-back letter task for 1 min (see below) before being tested in the main task. The main task was followed by psychophysical tests to assess the discriminability of motion stimuli (see below). We did not record EEG during these follow-up psychophysical tests. The entire experiment took under 2 h, including the setup of the EEG.

### Display

The experiment was performed using a Macbook Pro and external screen with a resolution of  $1920 \times 1080$  pixels with a 60 Hz refresh rate. All participants were seated at a viewing distance of 50 cm (without a chinrest). The experiments were programmed using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) for MATLAB (version 2014b).

### Task and Visual Stimuli During EEG Measurement

To achieve the desired level of visibility and to induce visual PE to visible and invisible changes, we manipulated the coherence of random-dot motion (Britten et al., 1992). In short, it is well-known that the direction of motion can be consciously discriminated when the level of motion coherence of random dots is sufficiently high. As the level of coherence approaches zero, participants can no longer consciously see the direction of motion (Watanabe et al., 2011). We exploited this useful feature of motion perception and designed a paradigm to elicit prediction errors from visible and invisible changes in the motion direction of a cloud of random dots (**Figure 1**<sup>1</sup>). These motion stimuli were

<sup>1</sup><https://figshare.com/s/76484519f510ba74891b>

never task-relevant during the main EEG experiment. Instead, participants were instructed to focus on a central 1-back task and be as quick and accurate as possible.

Our motion stimuli comprised 750 white dots enclosed within a 30° diameter circular area on a black background. During each 500 ms trial, each dot moved in a straight trajectory at 14°/s. When any dot went outside of the stimulus window, the dot was redrawn in the opposite location to keep the density of the dots constant. At the first frame of each 500 ms trial, we removed all dots and redrew them at new random locations at the same time (i.e., no blank interval between trials). Please refer to the exemplar video provided for further clarity of the stimuli<sup>1</sup>.

To elicit PE responses, we utilized a roving visual oddball paradigm (i.e., 5–8 repetitions before a change) with dot motion at 5 possible non-cardinal directions (70, 140, 210, 280, and 350°; with 0° representing motion to the right) and 2 possible coherence levels (high: 50% and low 5%). Motion direction was manipulated in a 2 × 2 design comparing the coherence levels: high and low, and *motion direction*: standard (frequent: no change) and deviant (rare/surprising: change of direction). Mean number of low-coherence standard and deviant trials was 829 ( $SD = 65$ ) and 136 ( $SD = 12$ ), respectively. Mean number of high-coherence standard and deviant trials was 832 ( $SD = 64$ ) and 136 ( $SD = 13$ ), respectively.

One experimental block comprised 180 trials. At the beginning of the block, both the direction and the coherence of the motion was chosen randomly from the five possible directions and two coherence levels. The direction and coherence level remained unchanged for 5 to 8 trials before we changed the global motion direction randomly into one of the other four directions. Throughout the paper, we refer to the trials with unchanged motion direction as the standard trials and to the trials in which the direction changed as the deviant trials. Each deviant trial was then followed by another 5 to 8 trials without a change in motion direction, which we considered as the new standard trials. In some deviant trials (once in every 26 to 30 trials), we also changed the coherence level. We did not include any of these ‘double deviant’ trials in our EEG or behavioral analysis.

At the center of the visual display, we presented a 1-back task within a 5° diameter black circle. Here, a sequence of randomly selected white letters (~1.8° visual angle in size) was presented every 450 ms (i.e., 150 ms on and 300 ms off). The first of these letter stimuli was presented at the onset of the first motion trial but became (mostly) desynchronized in time after this (except for every nine motion trials where the onset times realigned, see **Figure 1B**). Participants were required to attend to this letter stream and press the spacebar as quickly as possible whenever they detected the repetition of any letter (which could occur every 10 to 15 letters; **Figure 1B**). On average, the number of letter repeat targets per participant across the experiment was 191 (range from 177 to 196,  $SD = 3.69$ ).

## Behavioral Analysis on the 1-Back Task During the EEG Measurement

We examined participants’ accuracy during high- and low-coherence motion, defined as the hit and false alarm rates

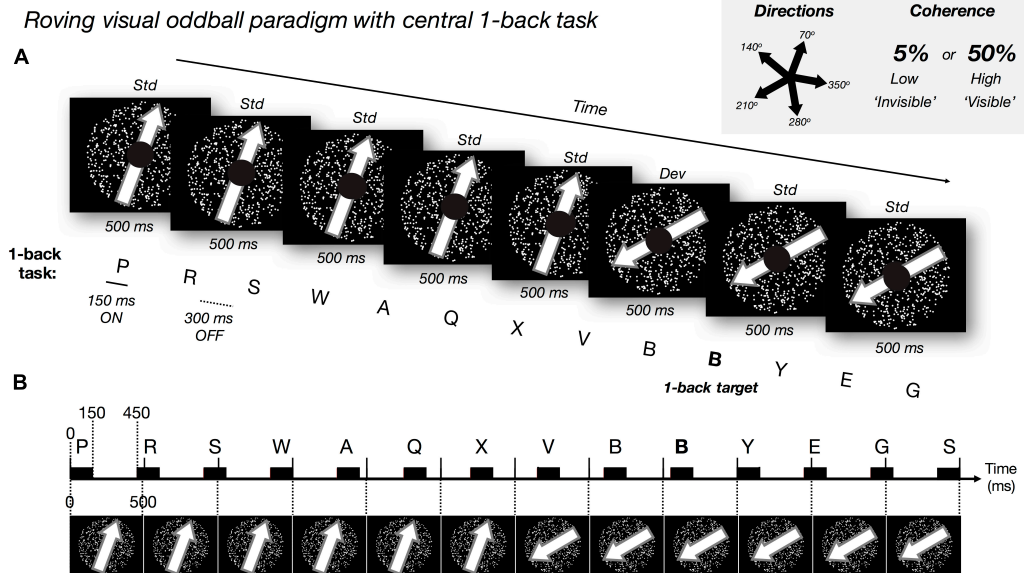
(and the sensitivity measure of  $d'$ ; Green and Swets, 1966). For this, we defined a ‘hit’ as a response made between 200 and 1,000 ms after the onset of a letter repeat. Responses outside of this were considered a false alarm. Due to the fast presentation rate of our target events, we chose to use the method described by Bendixen and Andersen (2013; ‘Method E,’ p. 930) to calculate our false alarm rate. Here, the number of non-target events (the denominator for calculating the false alarm rate) were defined as the duration of the experiment divided by the average time it would take to execute a response (i.e., 300 ms) minus the number of target events. In using this method, we aimed to overcome the bias in using traditional signal detection theory methods in paradigms with high event rates. Two participants failed to correctly detect above 50% of the targets across the entire experiment. We removed these participants from further analyses.

## Follow-Up Psychophysics Tasks

To ensure the visibility and invisibility of high- and low-coherence motion direction changes, respectively, we tested each participant’s performance with two follow-up psychophysics tasks. The first task required participants to make a direction discrimination at high- or low-coherence levels. The second psychophysics task, performed in a subset of participants ( $N = 8$  out of 28 participants), required them to report when they ‘felt’ or ‘sensed’ a change in direction occurred at both high- or low-coherence levels.

### Follow-Up Psychophysics Task 1: Direction Discrimination Task

To estimate the discriminability of the direction of motion, we ran a four alternative-forced choice (4AFC) direction discrimination task based on a study design with similar stimuli to ours (Tsushima et al., 2006). To conservatively estimate the discriminability, we reduced the direction alternatives to four possibilities (80, 160, 240, and 320°), which were different from the motion directions used in the main EEG experiment in order to reduce the chance of perceptual learning which can occur even to sub-threshold dot motion (Tsushima et al., 2006; Watanabe et al., 2011). Furthermore, we aimed to avoid any habituation effects that may have arisen if we had used the same motion directions in both experiments. These effects, in turn, could have improved or reduced task performance. Rather than inducing these uncertain results, we opted for slightly different stimuli as an alternative solution. In each trial, we presented the motion for 915 ms without the central letter task (i.e., the small, black central circle remained but the letters were absent). We selected this extended presentation time based on the study by Tsushima et al. (2006) and in order to conservatively estimate the motion direction discriminability. We did not redraw all the dots after 500 ms as in the EEG experiment but, instead, the dots remained moving for 915 ms (except when it came to the boundary, see above). We randomly selected a motion direction in each trial and pseudo-randomly intermixed four coherence levels (2.5, 5, 25, and 50%) in equal



**FIGURE 1 |** Experimental paradigm. **(A)** Roving visual oddball paradigm with dot motion at five possible directions and two possible coherence levels (low: 5% and high 50%). We manipulated *motion coherence*: high and low, and *motion direction*: standard ('std,' frequent: no change) and deviant ('dev,' rare/surprising: change of direction). The 'roving design' meant each deviant motion direction became the new standard direction over repetitions. We presented the task-irrelevant random-dot motion (500 ms per trial) in the visual periphery. Every 5 to 8 'standard' trials the global direction of the dots changed (the 'deviant'; randomly selected direction). Every 26 to 30 trials (i.e., 13–15 s), we changed both the direction and coherence level of the motion (but discarded EEG events and behavioral responses to these 'double deviants'). Participants focused on a central 1-back letter task (150 ms ON and 300 ms OFF: desynchronized with most of the motion trials, aligning once every nine trials), responding when the same letter was repeated in succession (e.g., bold letter, 'B'). Participants were instructed to ignore the motion stimuli and to be as fast and as accurate in the central task. **(B)** Schematic illustration of the timecourse for motion trials (every 500 ms) and 1-back letter presentation (black rectangles represent 150 ms letter ON).

proportions across 120 trials. We incorporated two additional motion coherence levels to reduce the chance of implicit learning based on the motion coherence level. At the end of every trial, participants reported the perceived direction of motion from the 4 possible alternatives.

### Behavioral Results Psychophysics Task 1: Confirming Visibility and Invisibility of High- and Low-Coherence Motion Direction Changes in Each Individual

We used performance in our follow-up psychophysics direction discrimination task to confirm that high- and low-coherence motion direction changes were visible and invisible, respectively. Based on the performance, we excluded participants from the following EEG analysis based on two criteria: (1) performing above chance in the low (5%) coherence motion condition or (2) performing below chance in the high (50%) coherence motion condition (**Figures 2A,B**). To estimate if the observed discrimination accuracy was above chance, we performed a bootstrap analysis (with replacement) over 1,000 repetitions per participant. For each participant, at a given coherence level, we generated a surrogate sequence of correct vs. incorrect discriminations across 30 trials (e.g., correct, incorrect, correct, correct, correct) which reflected the proportion of correct responses at the number of trials we tested. We then used 2.5%ile and 97.5%ile of the bootstrapped distribution as the confidence interval. That is, for the low (5%) coherence

condition, if the bottom end of the confidence interval was above chance (25% correct), we assumed that the participant may have been aware of the direction of motion during the EEG experiment (**Figure 2B**).

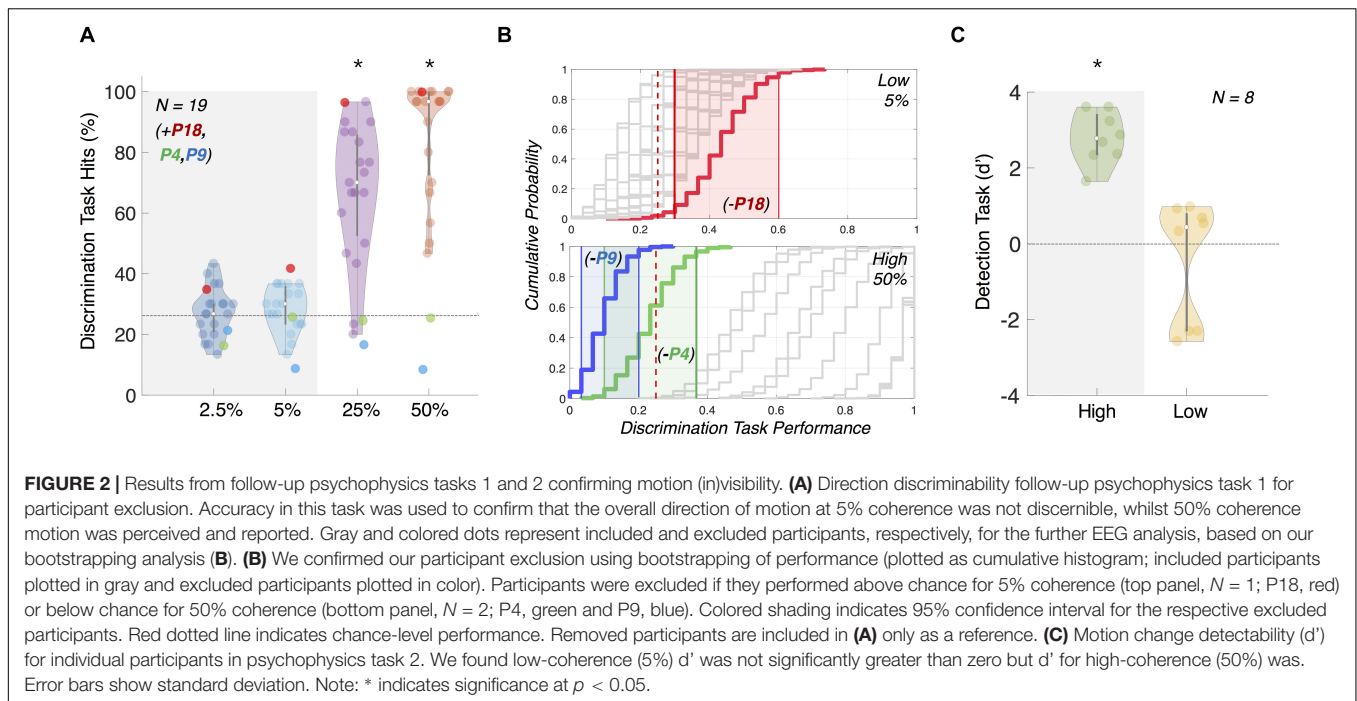
Based on our bootstrapping analyses, we confirmed the visibility and invisibility of the high- and low- coherence motion direction changes, and excluded two participants due to criterion 1 and one due to criterion 2. Due to technical error (corrupted data files), we could not perform bootstrapping analysis for four more participants and we conservatively chose to remove the EEG data of these four participants from further analysis as well.

Overall, we used the data from  $N = 19$  participants for ERP, source, and DCM analysis.

### Follow-Up Psychophysics Task 2: Detectability of Motion Direction Changes

Finally, as another follow-up experiment, we ran a detection task to exclude the possibility that participants were able to 'sense' the change of motion direction even in the absence of the perception of motion direction *per se*, but based on the metacognitive 'feeling of change' (Rensink, 2004). To test this, we repeated one block from the main experiment; after removing the letters for the 1-back task but leaving the small, black central circle (**Figure 1**). The direction of motion and coherence levels were the same as the main experiment. We instructed participants to press the spacebar as quickly as possible whenever they detected a change





in the global motion direction of the cloud of dots. Participants were instructed that even if they were unsure, they should report a feeling of change. We regarded a change of global motion direction within a coherence level (which was either 5 or 50%) as a target event. There were 20 events per coherence level. To analyze this behavioral data, we computed the sensitivity measure of  $d'$  using the same signal detection theory method as our main experiment.

### Behavioral Results Psychophysics Task 2: Confirming Lack of 'Feeling' of Change

We found that  $d'$  was significantly above zero for high-coherence motion direction changes (**Figure 2C**; one-sided one sample  $t$ -test,  $M = 2.77$ ,  $SD = 1.68$ ,  $p > 0.001$ ,  $df = 7$ ) but not for low-coherence motion direction changes (one-sided one sample  $t$ -test,  $M = -0.47$ ,  $SD = 1.60$ ,  $p = 0.781$ ,  $df = 7$ ). We took this as evidence that this subset of participants could not even sense direction changes at low-coherence, while they were able to do so at high-coherence.

### EEG Preprocessing

We used SPM12<sup>2</sup> to pre-process the data. We first re-referenced the raw EEG recordings to the average of all electrodes, down-sampled to 200 Hz and high-pass filtered above 0.5 Hz. We epoched within a trial time window of 100 ms before to 400 ms after the onset of each motion trial. We then removed eyeblink artifacts using the VEOG channels with a bad channel maximum rejection threshold of 20% and by thresholding all channels at 100  $\mu V$ . In the following EEG analysis, we present data without the removal of trials in which a target for the 1-back task

occurred or a button response was made. Given that the task and button presses were uncorrelated with the conditions of interest (i.e., motion trials), these are unlikely to have an effect on our results (which we confirmed; data not shown). To obtain the mean ERP per participant per condition, we averaged the epoched data using robust averaging (Wager et al., 2005) with a built-in function in SPM12. This robust averaging process down-weighted each time-point within a trial according to how different it was from the median across trials. Next, because high frequency noise can be introduced during the robust averaging process, we further low-pass filtered the processed data at 40 Hz. Finally, we baseline corrected the data for each participant and condition by subtracting the (robustly averaged) ERP between  $-100$  and  $0$  ms.

### ERP Analysis

We analyzed ERPs corresponding to each of the experimental conditions across participants (as well as the visual PE, in the form of the 'vMMN': derived from subtracting the standard waveform from the deviant waveform within a condition). We combined channel clusters over left anterior (FC3, F1, F3, F5, AF3), right anterior (FC4, F2, F4, F6, AF4), left central (C3, C5, CP1, CP3, CP5), right central (C4, C6, CP2, CP4, CP6), left posterior (P7, PO3, PO7, PO9, O1) and right posterior (P8, PO4, PO8, PO10, O2) electrodes (see Jack et al., 2015). The significant vMMN time periods were those where the difference waves mean amplitude was below zero across participants (after FDR correction for multiple comparisons at each time point).

### Spatio-Temporal Statistical Maps

We obtained spatiotemporal images of the ERP for each participant and condition across the scalp within the window of

<sup>2</sup><http://www.fil.ion.ucl.ac.uk/spm/>

–100 to 400 ms. To obtain one 2D image, for each of the 101 time bins, we projected the EEG electrode locations onto a plane and interpolated the electrode locations linearly onto the  $32 \times 32$  pixel grid. We then stacked each 2D image over time to obtain a 3D volume ( $32 \times 32 \times 101$ ) and smoothed with a 3D Gaussian kernel of full-width at half maximum:  $12 \text{ mm} \times 12 \text{ mm} \times 20 \text{ ms}$ .

The 3D spatio-temporal image volumes were analyzed, on a participant-by-participant basis, with a mass univariate general linear model (GLM) as implemented in SPM12. We estimated the main effects of surprise (i.e., standards vs. deviants) and coherence (i.e., high- vs. low-coherence), their interaction, and contrasts between standards vs. deviants separately within the high- or low-coherence conditions using between-subject F-contrasts. All spatio-temporal effects are reported at a threshold of  $p < 0.05$  at the cluster-level with family-wise error (FWE) correction for multiple comparisons over the whole spatio-temporal volume. *F*-statistics are reported as the maximum at the peak-level.

## Source Reconstruction

We obtained source estimates on the cortical mesh by reconstructing scalp activity with a single-sphere boundary element method head model, and inverting a forward model with multiple sparse priors (MSP) assumptions for the variance components under group constraints (Friston et al., 2008). One benefit of using MSP for the source reconstruction process is that through the use of empirical Bayes one can select multiple sources to build covariance components for both sparse or distributed source configurations (Friston et al., 2008). Because our main effects of coherence and surprise (and their interaction) spanned the entire epoch we decided to consider the whole peristimulus time window (0 to 400 ms) in the MSP procedure. This allowed for inferences on the most likely cortical regions that generated the sensor-level data across the entire trial time window. We obtained volumes from these reconstructions for each of the four conditions for every participant. These images were smoothed at full-width at half maximum:  $12 \text{ mm} \times 12 \text{ mm} \times 12 \text{ mm}$ . We then computed the main effects of (1) coherence, (2) surprise, (3) the interaction (coherence  $\times$  surprise), as well as the (4) high- and (5) low-coherence PE (standards vs. deviants) using conventional SPM analysis between-subject F-contrasts.

## Dynamic Causal Modeling

We used DCM to estimate a generative causal model that most parsimoniously explained the observed ERPs at the selected source locations with minimal model complexity (Friston et al., 2003). It is important to note, here, that DCM is a statistical inference of causal models, and ‘causality’ is not established through perturbation or other intervention (Pearl, 2000). We used a data-driven approach combined with *a priori* locations drawn from the visual motion processing literature, to explain best the observed PE related signals for visible and invisible motion changes. The data-driven spatial location of left inferior temporal gyrus (ITG) was selected based on our source-level analysis. We also included the sensory input nodes of bilateral primary visual cortices (V1) and middle temporal cortex (MT+), which are likely to be the initial cortical stages for visual motion

processing (Born and Bradley, 2005), and the bilateral posterior parietal cortices (PPC), which are known for higher-level visual motion processing (Anderson, 1989; Ilg et al., 2004). In total, we assumed seven sources: bilateral V1, MT+, PPC and left ITG [see section “Results” for Montreal Neurological Institute (MNI) coordinates].

We connected our candidate nodes using the same architecture and then exhaustively tested all possible combinations for the direction of modulation(s) amongst these nodes (with one exception, see below). Our model architecture comprised: (1) recurrent (forwards and backwards) connections from V1 to MT+ and MT+ to PPC within the left and right hemispheres, and left MT+ to left ITG, (2) each node was laterally connected with the corresponding node in the other hemisphere (e.g., left V1 laterally connected to right V1), (3) intrinsic (or within-region) modulation only at V1 and MT+, which are known to strongly adapt to repeated visual motion in the same direction (Kohn and Movshon, 2004), and (4) left ITG and left PPC laterally connected. We decided to assign the equal hierarchical relationship between ITG and PPC due to the inconsistent relationship between them reported in the literature (e.g., DeYoe and Van Essen, 1988; Felleman and Van Essen, 1991).

Next, although there could be a huge number of possible DCM modulations based on our seven identified nodes, we decided to reduce the possible space for modulations based on anatomical information as much as possible. By anatomical criteria, we decided to examine models that always contained recurrent (forwards and backwards) modulation between the input sources of bilateral V1 and MT+ and intrinsic modulations at these input nodes; we refer to this as the ‘minimal’ model (i.e., a base model for all other models). Using this reduced number of potential modulation directions, we were left with 7 connections that could be modulated beyond the minimal model: 4 connections between MT+ and PPC (forwards and backwards in each hemisphere), 2 connections between MT+ and IT (forwards and backwards in left hemisphere) and 1 connection between PPC and IT (one lateral connection in left hemisphere). Note, that we never modulated the lateral connections between the hemispheres. Thus, in total, we tested 129 models (i.e.,  $2^7 = 128$  + a null model with no modulation at or between any nodes) comprising all combinations of modulation directions (see **Figure 7A**).

We performed DCM analyses to estimate the parameters of each model, separately for each effect of interest: (1) the visible PE and (2) the invisible PE. For the visible PE, we used the between-trial effect (condition weights) of [0, 1] for the high-coherence standard and deviant trials, respectively. For the invisible PE, we used the weights of [0, 1] for the low-coherence standard and deviant trials, respectively. DCM analyses started from estimation of the log model evidence for each of our 129 generative models fitted for every participant. The log model evidence quantified how well each model could generate ERPs similar to the observed (pre-processed) ERPs for that participant. Importantly, the iterative free-energy minimization process used to calculate the log model evidence penalized models with greater complexity to avoid overfitting (Friston et al., 2008). Once estimated, we used a random-effects (RFX)

approach to determine the winning model across participants via Bayesian Model Selection. RFX assumes that a (potentially) different model underpins each participant's responses; making it robust to outliers and best suited for studying perceptual processes whose underlying network structure is likely to be varied across participants (Stephan et al., 2009). To compute a weighted average of the parameter estimates across all models, we employed Bayesian Model Averaging (BMA) (Penny et al., 2010). BMA weights the estimated parameter with the probability of each model associated with that parameter for all participants. In this way, all models contributed to the final connectivity estimate, with the most probable model having the greatest weight and the least probable model contributing the least to the final estimates (Stephan et al., 2010). In a follow-up DCM analyses, we compared the visible and invisible PE by modeling their interaction using the between-trial effect (condition weights) of [0,1] to contrast the mean ERPs for the visible PE and invisible PE per participant.

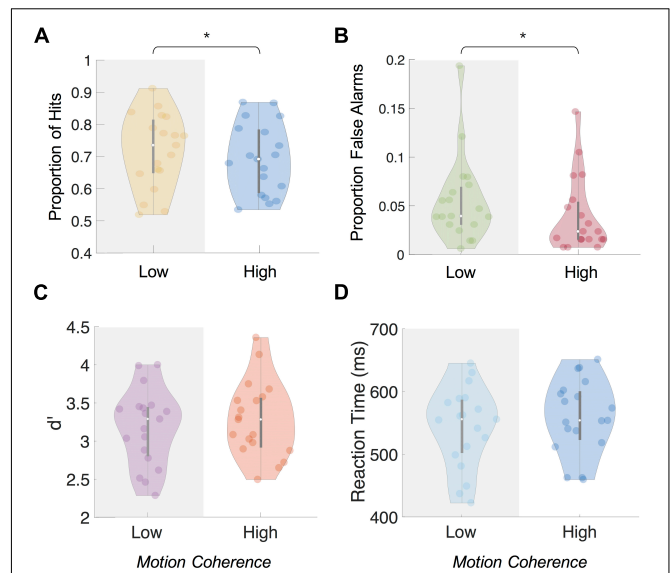
## RESULTS

### No Difference in Distraction by Visible or Invisible Motion Directions

After exclusion based on the follow-up psychophysics task (see Materials and Methods), we analyzed the performance of the remaining  $N = 19$  participants in the 1-back task to check if they appropriately focused on the central letter task and ignored the background motion during the EEG session of the main task. Based on prior studies (e.g., Tsushima et al., 2006), we expected that our low-coherence background motion would distract participants more than high-coherence motion to degrade the performance of the 1-back task.

Between the high- and low-coherence conditions, we observed a difference in the percentage of hits and false alarms but no difference in  $d'$  or reaction times. The proportion of hits during the high-coherence condition ( $M = 69.54\%$ ,  $SD = 11.68\%$ ) was significantly lower than in the low-coherence condition [ $M = 73.87\%$ ,  $SD = 12.34\%$ ; two-tailed paired  $t$ -test,  $t(18) = -3.27$ ,  $p = 0.004$ ; **Figure 3A**]. Mean number of false alarms across participants was also significantly lower in the high-coherence condition ( $M = 4.95$ ,  $SD = 4.71$ ) than the low-coherence condition [ $M = 6.95$ ,  $SD = 5.34$ , two-tailed paired  $t$ -test,  $t(18) = -2.39$ ,  $p = 0.028$ ; proportion shown in **Figure 3B**]. Combining the proportion of hits and false alarms, we computed the  $d'$  measure. Using this measure, we found no significant differences between  $d'$  for performance on the 1-back task during the high- ( $M = 3.32$ ,  $SD = 0.49$ ) and low-coherence conditions [ $M = 3.31$ ,  $SD = 0.49$ , two-tailed paired  $t$ -test,  $t(18) = 0.09$ ,  $p = 0.931$ ] (**Figure 3C**). Furthermore, we found no differences in the reaction times between the high ( $M = 557$  ms,  $SD = 59$  ms) and low [ $M = 555$  ms,  $SD = 64$  ms, two-tailed paired  $t$ -test,  $t(18) = 0.722$ ,  $p = 0.48$ ] coherence conditions (**Figure 3D**).

Overall, we did not find the results to be consistent with our expectation of greater distraction effects during low-coherence motion trials. We suggest this may be due to the differences between our study and that by Tsushima et al. (2006; such as the different behavioral task and the inclusion of the PE



**FIGURE 3 |** Behavioral results from central 1-back task during Main Experiment. Mean proportion of hits (A) and false alarms (B) was significantly higher ( $p < 0.01$  and  $p < 0.05$ , respectively) when background motion was low-coherence than high-coherence. Mean  $d'$  (C) and reaction times (D) showed no significant difference ( $p > 0.05$ ) between the coherence levels. Note: \* indicates significance at  $p < 0.05$ .

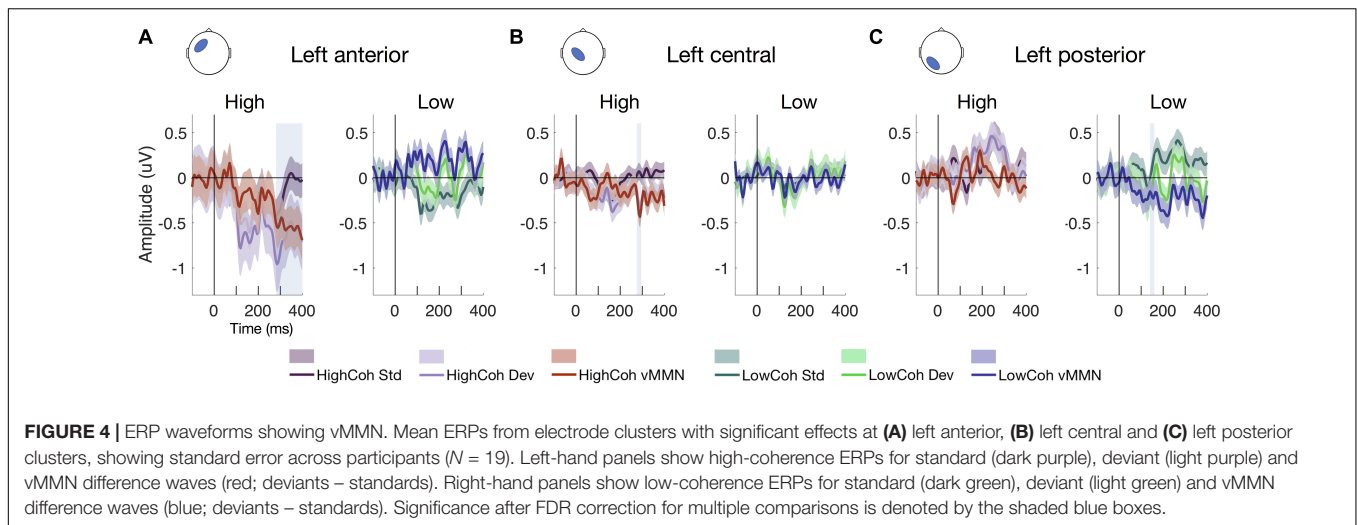
generating stimulus sequence), leading to lower task difficulty and comparable distractibility effects between the two coherence levels. We interpret these results as: (1) participants were able to maintain their focal attention to the 1-back task, (2) participants did not trade off speed for accuracy differently between the two coherence conditions and (3) that the behavioral effects of high- or low-coherence motion were comparable.

## Scalp-Level ERP Analysis

### Non-conscious Prediction Errors Occur Earlier Than Conscious Prediction Errors

We extracted the ERPs from scalp-level electrode clusters to examine the vMMN for visible and invisible PE responses (defined as sustained negativity of vMMN difference waves after correcting for multiple comparisons at each time point,  $q < 0.05$ ). We found that vMMN was evoked from both visible and invisible motion direction deviants. For visible PE responses (**Figures 4A,B**), we observed vMMN ( $p_{FDR} < 0.05$ ) at left central and anterior channels clusters between 285–295 and 275–400 ms, respectively. For the invisible PE responses (**Figure 4C**), we observed vMMN at left posterior channels at 150 ms.

In addition to examining ERPs at the scalp, we used spatio-temporal statistical map analysis in SPM (see section “Materials and Methods”). Here, we obtained the 3D images interpolated from ERP data recorded at the scalp and applied between-subject  $F$ -tests to quantify the effect of surprise [i.e., prediction error (PE)] within the high- or low-coherence conditions (i.e., visible and invisible PE). Importantly, we found that PEs were evoked from both visible and invisible motion direction deviants. PE to visible motion direction changes (**Figure 5A**) disclosed a number



of significant clusters, ranging from 290 to 395 ms observed across widespread channels. The earliest cluster at 290 ms was found at the central channels (peak-level  $F_{max} = 25.44$ , cluster-level  $p_{FWE} = 0.024$ ) followed by a cluster at left front-temporal and central channels at 380 ms (peak-level  $F_{max} = 34.12$ , cluster-level  $p_{FWE} < 0.001$ ) and at 395 ms (peak-level  $F_{max} = 67.29$ , cluster-level  $p_{FWE} < 0.001$ ). Compared to visible PE, invisible PE occurred earlier and were less spatially spread (**Figure 4B**); only at 160 ms in left parietal channels (peak-level  $F_{max} = 25.65$ , cluster-level  $p_{FWE} = 0.024$ ).

Next, we applied between-subject  $F$ -tests to quantify the main effects of coherence (**Figure 5C**), surprise (**Figure 5D**) and their interaction (**Figure 5E**). The main effect of coherence (**Figure 5C**) disclosed three significant clusters. The first cluster at 160 ms was located occipitally (peak-level  $F_{max} = 34.37$ , cluster-level  $p_{FWE} = 0.002$ ), the second at 185 ms occurred in the same location (peak-level  $F_{max} = 35.72$ , cluster-level  $p_{FWE} = 0.004$ ) and the third at 295 ms was found at right occipito-parietal and frontal channels (peak-level  $F_{max} = 32.31$ , cluster-level  $p_{FWE} = 0.001$ ). The main effect of surprise (**Figure 5D**) showed three significant clusters. The first at 80 ms occurred at left parietal channels (peak-level  $F_{max} = 29.53$ , cluster-level  $p_{FWE} = 0.014$ ), the second at 285 ms was located at right central channels (peak-level  $F_{max} = 34.66$ , cluster-level  $p_{FWE} < 0.001$ ) and the third at 375 ms was observed in the same location (peak-level  $F_{max} = 55.29$ , cluster-level  $p_{FWE} < 0.001$ ). Finally, we observed an interaction between surprise and coherence (**Figure 5E**) at central and frontal channels at 395 ms (peak-level  $F_{max} = 35.93$ , cluster-level  $p_{FWE} = 0.002$ ).

## Source-Level Analysis

### Left ITG as a Source for Conscious PE

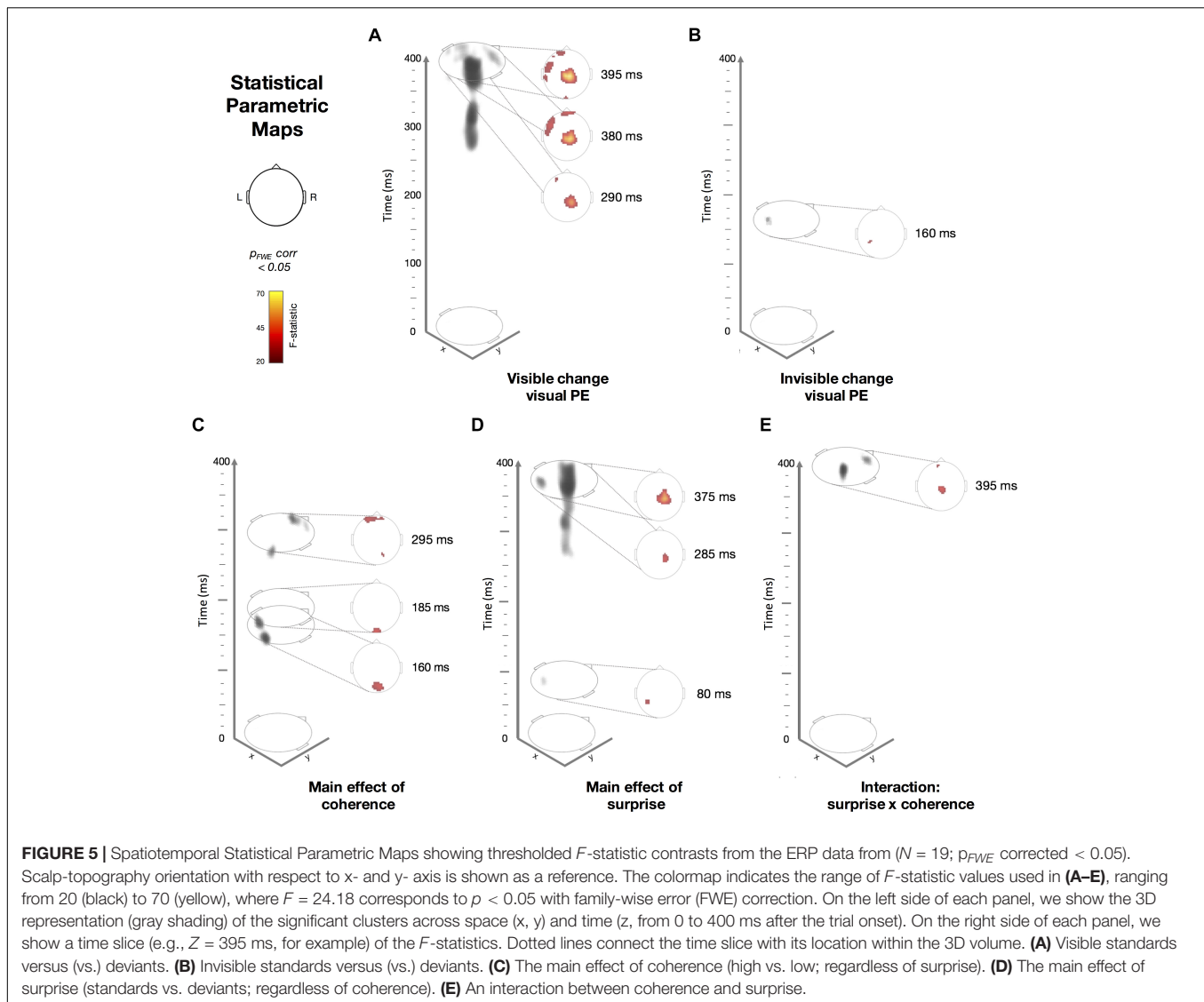
We applied MSP source reconstruction to estimate the cortical regions involved in generating PE to visible and invisible motion direction changes. In **Figure 5** (the scalp-level maps), we found that the main effects of surprise, coherence, and the interaction spanned the whole epoch, thus we decided to use the whole epoch data (0–400 ms) rather than to temporally constrain the data for

source reconstruction (see section “Materials and Methods” for details). **Figure 6** shows the significant source-level results for the main effect of surprise and PE to visible changes at an uncorrected threshold of  $p < 0.001$ . We did not find any significant sources for invisible PE, main effects of coherence or interaction when we corrected for multiple comparisons. For the main effect of surprise, we found one significant cluster in the left ITG ( $[-52, -26, -30]$ , peak-level  $F_{max} = 22.19$ , cluster-level  $p_{FWE} = 0.003$ , **Figure 6A**). For the PE to visible changes, we found a similar cluster in the left ITG ( $[-48, -12, -32]$ , peak-level  $F_{max} = 17.91$ , cluster-level  $p_{FWE} = 0.039$ , **Figure 6B**). The effect of surprise for the invisible conditions revealed a cluster on the right hemisphere (**Figure 6A**) which did not survive correction.

## Dynamic Causal Modeling

We used DCM to examine how the source location identified in the previous step, interacted with cortical locations known to specialize in motion processing, and how the strengths of these interactions are modulated by the visibility of motion changes. Specifically, we identified one region at the FWE corrected ( $p < 0.05$ ) threshold (between-subject  $F$ -tests) (Anatomy Toolbox; Eickhoff et al., 2005): the left ITG (MNI coordinate:  $[-48, -12, -32]$ , **Figure 6**). We included the sensory input nodes of bilateral V1 (MNI coordinates: left  $[-14, -100, 7]$  and right  $[17, -97, 9]$ ) and bilateral MT+ /V5 (MNI coordinates: left  $[-48, -69, 7]$  and right  $[50, -66, 11]$ ) as these regions are essential for visual motion processing (Born and Bradley, 2005; Plomp et al., 2015). We also included the bilateral PPC (MNI coordinates: left  $[-46, -46, 54]$  and right  $[52, -42, 50]$ ) because these regions are known for higher-level visual motion processing (Ilg et al., 2004). Based on our 7 identified nodes, we tested 129 models comprising all combinations of modulation directions (building on the fixed ‘minimal model’; see section “Materials and Methods” for more information). **Figure 7A** shows the defined model space, including: the fixed model architecture (white arrows) and how each model was allowed to vary in terms of the direction of modulation (black arrows).





**FIGURE 5 |** Spatiotemporal Statistical Parametric Maps showing thresholded  $F$ -statistic contrasts from the ERP data from ( $N = 19$ ;  $p_{FWE}$  corrected  $< 0.05$ ). Scalp-topography orientation with respect to  $x$ - and  $y$ -axis is shown as a reference. The colormap indicates the range of  $F$ -statistic values used in (A–E), ranging from 20 (black) to 70 (yellow), where  $F = 24.18$  corresponds to  $p < 0.05$  with family-wise error (FWE) correction. On the left side of each panel, we show the 3D representation (gray shading) of the significant clusters across space ( $x$ ,  $y$ ) and time ( $z$ , from 0 to 400 ms after the trial onset). On the right side of each panel, we show a time slice (e.g.,  $Z = 395$  ms, for example) of the  $F$ -statistics. Dotted lines connect the time slice with its location within the 3D volume. (A) Visible standards versus (vs.) deviants. (B) Invisible standards versus (vs.) deviants. (C) The main effect of coherence (high vs. low; regardless of surprise). (D) The main effect of surprise (standards vs. deviants; regardless of coherence). (E) An interaction between coherence and surprise.

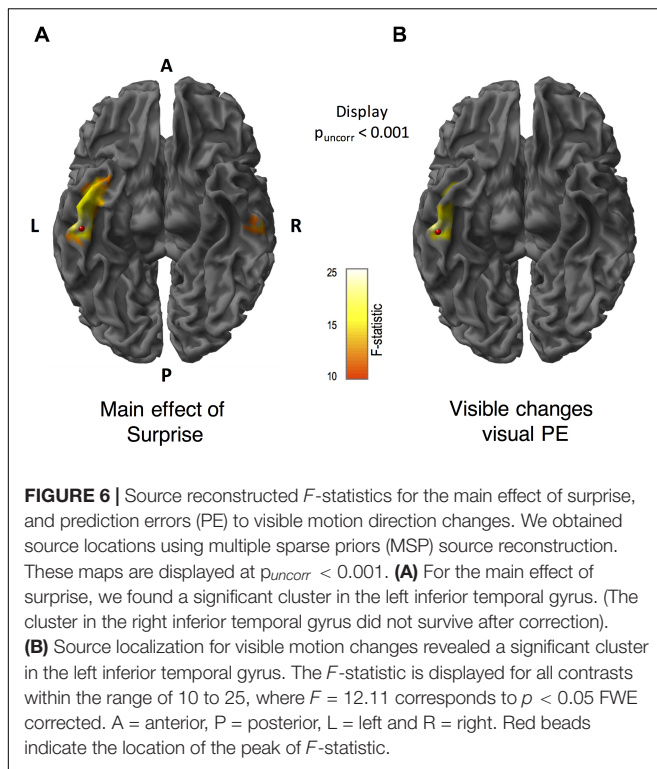
## Optimized DCM Modulation Parameters Explaining Differential ERPs for Visible and Invisible PE

Once the log model evidence for every model and participant was estimated, we used RFX Bayesian Model Selection to determine the winning model at the group-level. **Figure 7B** shows, for both the visible and invisible PE, the exceedance probability of all models at the group-level. The exceedance probability is the probability that a particular model is more likely than any other model given the group data.

For the visible PE, there was no single clear winning model amongst our 129 models. This is unsurprising when using RFX analysis as effects can be diluted if a large number of models are compared in a small sample size. In this situation, it is recommended not to single out the best architecture alone, but rather to use BMA to consider all models, taking into account the strength of the evidence for each model. After applying BMA,

we found significant positive modulation of the intrinsic self-connections of left MT+ (mean parameter outputs = + 0.2004, two-tailed  $t$ -test against 0:  $p_{FDR} = 0.0486$ ,  $df = 18$ ), and right MT+ (+ 0.2644,  $p_{FDR} = 0.0011$ ,  $df = 18$ ). This result confirms what we expected from the nature of our oddball paradigm; the visible change induced PE relies, in part, on intrinsic (self) modulation at MT+ due to a release from adaptation (i.e., repetition suppression) following the onset of deviant motion. In addition to this, we also found two more significantly modulated connections. We found a *backward* connection from right MT+ to right V1 was significantly downregulated (−0.1310,  $p_{FDR} = 0.0453$ ,  $df = 18$ ) and a *forward* connection from left MT+ to left ITG was significantly positively modulated (+ 0.0875,  $p_{FDR} = 0.0060$ ,  $df = 18$ ).

For the invisible PE, as with our visible PE DCM analyses, in the absence of one single winning model we again applied BMA and found that, similarly to the visible PE DCM results, we observed significant positive modulations for intrinsic



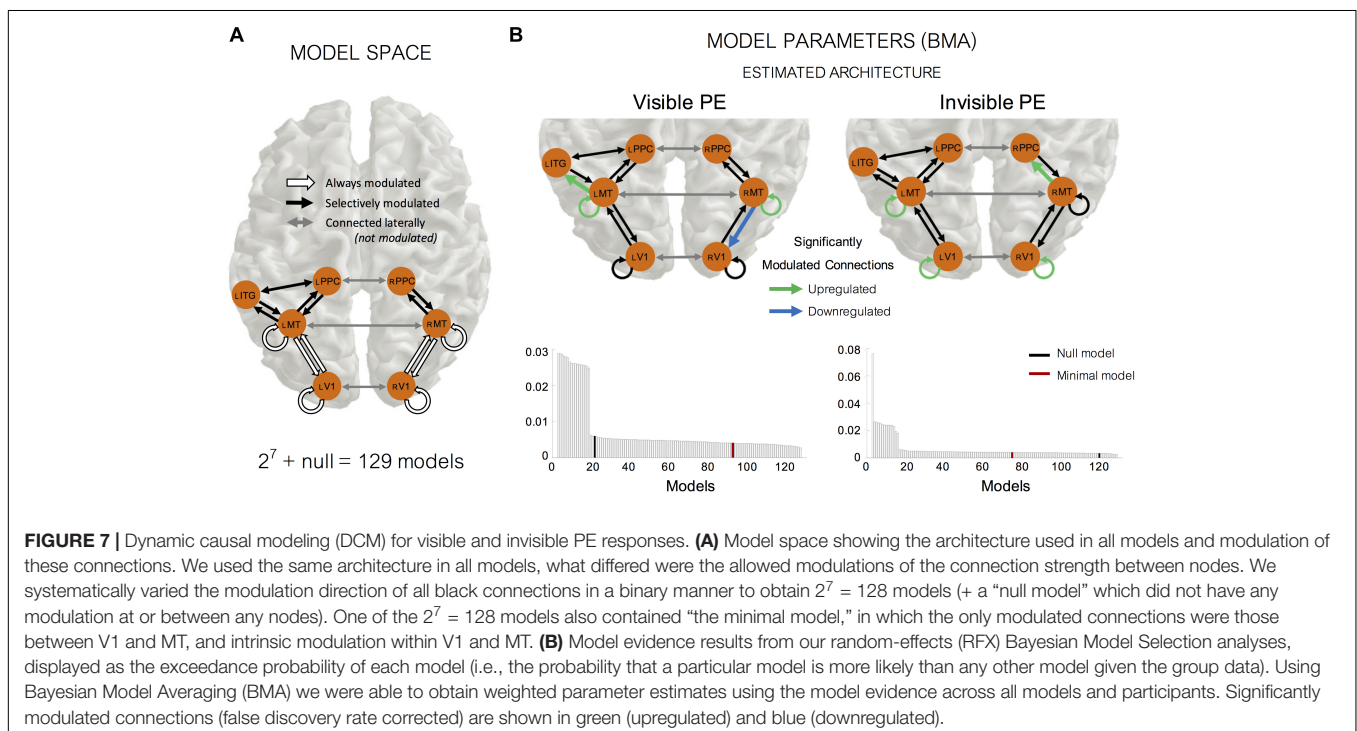
modulations within left V1 (+ 0.1501,  $pFDR < 0.0001$ ,  $df = 18$ ), right V1 (+ 0.1400,  $pFDR = 0.0081$ ,  $df = 18$ ) and left MT+ (+ 0.2821,  $pFDR < 0.0001$ ,  $df = 18$ ), reflecting a release from adaptation upon a change in motion direction, despite no

awareness of this change occurring. One difference, compared to the visible PE, was the direction of one other significantly positively modulated connection *forwards* from right MT+ to PPC (+ 0.2048,  $pFDR = 0.0094$ ,  $df = 18$ ). We found no significantly modulated backward connections within our DCM analyses of the invisible PE.

Finally, we directly compared differences in effective connectivity between the visible and invisible PE by modeling their interaction. We applied RFX BMA to determine the model parameters in the absence of one single winning model and found two significant connections (prior to FDR correction) between left ITG and left PPC and at a self-connection of left MT+. However, these connections did not survive correction for multiple comparisons. We interpret these findings as weak but consistent with our source results showing left-hemisphere localization of the visible PE.

## DISCUSSION

In this study, we aimed to elicit PE related neural activity via changes in stimulus statistics that were consciously visible or not. As planned, we successfully manipulated the awareness of the stimulus changes using supra- and sub-threshold motion coherence of a dynamic random-dot display and confirmed the visibility of the motion direction changes with a follow-up psychophysics task. Our ERP analyses (**Figures 4, 5**) confirmed robust visible PE responses, which were widespread in space and time and invisible PE responses, which were more confined in space and time. Our source level analysis located the source of the visible PE response in the left ITG. Using the identified location of



the left ITG and the cortical nodes known to be critically involved in motion processing (V1, MT+, and PPC), we performed DCM analyses, hoping to reveal the network underpinning visible and invisible PE responses.

## Confirmation of Motion Direction (In)visibility

It is important that we first acknowledge that our method of rendering our stimuli invisible (or weaker stimulus paradigms in general, e.g., Dehaene et al., 1998; Dehaene et al., 2001; Del Cul et al., 2007; Gaillard et al., 2009; van Vugt et al., 2018) confounds the issue of conscious perception with the issue of stimulus change. We also point out that alternative approaches which eschew this issue (such as binocular-rivalry like paradigms), do suffer from other issues such as the effects of the report. That is, when the physical stimulus input is equated, the effects of reports can be falsely interpreted as the neural correlates of consciousness (Wilke et al., 2009; Frässle et al., 2014; Tsuchiya et al., 2015). Thus, we foresee the combination of both approaches (Tsuchiya et al., 2016) will be important for future studies.

Secondly, we must highlight the factors leading to the design of our follow-up psychophysics task 1 that was used to confirm (in)visibility of motion direction at high- and low-coherence. The design of this task was motivated by a number of concerns with presenting motion stimuli identical to that used in the main experiment. We opted to change the motion parameters during this task primarily because of our concern on perceptual learning and habituation during the visibility check. To alleviate this issue, we made four critical changes in the stimulus design: we lengthened the visual motion presentation time, removed the distracting central letter task, intermixed 4 coherence levels, and reduced the direction alternatives. All of these changes, in addition to the fact that during the main experiment participants did not pay attention to the motion, should have made our visibility assessment rather conservative; that is, we should have detected any participants who were aware of the low-coherence visual motion direction changes during the main task using this visibility task. Furthermore, while our choice of 30 trials was on the smaller end in terms of the number trials as a visibility check, we performed non-parametric bootstrapping analysis on an individual basis, which is the most sensitive method to detect aware participants. Overall, we are confident in our method of confirming that the motion direction changes were visible or invisible.

## Temporal Features of the Prediction Error (PE) Responses With Visible and Invisible Motion Changes

We were able to elicit PE responses for both visible and invisible motion direction changes. Scalp ERPs (Figure 4) disclosed a significant early component at 150 ms for invisible PE (i.e., the low-coherence vMMN difference wave) at left posterior channels. In contrast, we observed two significant later components for the visible PE (high-coherence vMMN difference waves) between 285–295 and 275–400 ms at left central and left anterior clusters, respectively. Our spatio-temporal scalp-level statistical mapping

analysis (Figure 5) supported these findings. Here, after FWE correction, PE evoked by invisible motion direction changes peaked at parietal channels earlier in time (<160 ms) than PE to visible motion direction changes (>290 ms), which were observed at central and fronto-temporal channels.

Our finding is consistent with some studies that elicited non-conscious visual PE. In an MEG study using vertical gratings and backward masking of rapidly presented deviants, Kogai et al. (2011) found similar latencies in the non-conscious PE response from 143 to 154 ms in striate regions. Further, Czigler et al. (2007), backward masked colored checkerboards at different stimulus onset asynchronies (SOAs) and found mask SOAs at 40 and 53 ms elicited vMMN at 124 and 126 ms at occipital channels (but that participants had a level of awareness of the deviant stimuli). Other studies, however, report later PE to invisible oddballs. For example, Jack et al. (2017) found slightly later non-conscious PE (to deviant stimuli presented monocularly to the non-dominant eye) that peaked around 250 ms (as did the conscious PE). This longer latency for the non-conscious PE was also echoed in another study from the same group (Jack et al., 2015). We suggest the discrepancies in the latencies arise from both stimulus differences and in the methods used to render stimuli non-conscious. Future paradigms in which both the standard and deviant stimuli can be rendered invisible may help disentangle conscious and non-conscious PE effects.

But why does the invisible PE occur earlier than visible PE? Here we offer two possible explanations. The first is related to the effects of attention. Previous visual PE studies that have used consciously *perceivable motion direction* changes found an earlier PE at 142–198 ms regardless of the attentional load on the irrelevant task (Kremláček et al., 2013). A separate study found only a later PE component (>275 ms) when attention was directed to the PE generating stimuli (Kuldkepp et al., 2013). Indeed, in a binocular rivalry vMMN study (using grating stimuli), van Rhijn et al. (2013) observed two early components from 140 to about 220 ms both when attention was directed to the rivalry stimuli and when attention was diverted. But a second late negativity (270 to 290 ms) was observed only when attention was directed to the rival stimuli. In our paradigm, participants may have attended to the visible motion resulting in a delayed PE. Conversely, invisible motion may not have deployed attention, which resulted in an earlier PE. However, this explanation does not speak to the neural mechanisms. An alternative explanation lies on a putative difference in adaptation (or repetition suppression) for visible and invisible stimuli. It is plausible that stronger adaptation in MT+ to the visible stimuli may mask the earlier PE component in V1. This adaptation effect may be weaker in MT+ for invisible stimuli, although this interpretation does not fully concur with our DCM results showing adaptation in V1 (but not right MT+) for invisible stimuli (see below for further discussion). Additionally, the later PE component for visible stimuli may be more related to the awareness of prediction violation. Further studies that orthogonally manipulate adaptation and prediction might be able to distinguish between these possibilities (Summerfield et al., 2006; Summerfield and Egner, 2009; Kok et al., 2012) and disentangle the two components underlying the ('classical')

vMMN; namely, the ‘genuine’ vMMN PE response and the effects of neural adaptation (O’Shea, 2015).

## Spatial Features of the Prediction Error (PE) Responses With Visible Motion Changes

In terms of spatial characteristics of visible PE, and based on previous studies by Tsushima et al. (2006), we expected to detect prefrontal source activity to our supra-threshold visible motion stimuli. Instead, our source-level analyses (which pooled responses over time) provided evidence that visible PE, as well as the main effect of coherence (regardless of the level of surprise), were generated by cortical sources within the left ITG. We suggest these source-level differences might have arisen from differences, most likely, in our measurement modalities (fMRI in Tsushima et al., 2006, and EEG in our study), as well as differences in experimental paradigms, namely, our inclusion of the roving oddball design instead of coherence differing on a trial-by-trial basis and our 1-back task.

According to our literature search, we note one study of visual consciousness that supports our findings of left hemispheric lateralization (O’Shea et al., 2013). In this study, the authors used binocular rivalry to seek brain activity that could predict visual consciousness and found early activity (around 180 ms) confined to left parietal-occipital-temporal regions. Beyond this, literature on the lateralization of conscious awareness of specific stimuli to the left hemisphere is scarce; reported only in a handful of studies of emotional processing (Gazzaniga, 2000; Kimura et al., 2004; Williams et al., 2006; Meneguzzo et al., 2014; Schepman et al., 2016). In these studies, subliminally presented face stimuli activated right amygdala (via a subcortical route) in response to fear but left amygdala for supraliminal fearful stimuli (Williams et al., 2006), and using masked or unattended affective-stimuli activated the right hemisphere in both the visual (Kimura et al., 2004) and auditory domains (Schepman et al., 2016). However, unlike these studies, we used visual motion stimuli and examined the difference between PE rather than responses to visual stimuli more generally. Another possible explanation for the observed lateralization of conscious perception of motion changes to left hemisphere comes from work in split-brain patients showing the left hemisphere is more adept at monitoring probabilities to infer causal relationships based on series of events over time (Gazzaniga, 2000; Roser et al., 2005) and at creating internal models to predict future events (Wolford et al., 2000). This suggests that when stimuli are consciously perceivable, the left ITG generates and updates predictions and PE. However, due to the lack of literature, specifically on visual PE lateralization, further work is needed to fully understand the role of left ITG source for conscious PE processing.

## Network Level Model of Causal Connections Investigated by DCM

Using DCM, we aimed to extend earlier visual PE studies by investigating the network level properties underlying the generation of PE to visible and invisible changes. The primary question of our DCM analysis was whether there was evidence

for top-down modulations for PE to visible or invisible change. According to the predictive coding framework, when an unexpected stimulus occurs, the PE signal is propagated from lower to higher brain areas, resulting in upregulation of forward connectivity. This, in turn, is followed by the revised prediction from high- to low-level brain areas, resulting in increased feedback connectivity. Previous studies are consistent with this theory that consciously perceived PE lead to increases in both feedforward and backward connectivity (e.g., Boly et al., 2011). Our finding of increased forward connectivity from left MT+ to ITG for visible PE is consistent with the first part of the theoretical prediction. What is puzzling is that conscious PE was accompanied by significant decreases in top-down feedback connectivity from right MT+ to right V1 (**Figure 7B**). This means that the neural prediction from MT+ to V1 decreased when the motion direction was unexpected, which appears inconsistent with the general framework of predictive coding. One possible explanation is that when the prediction is violated, the system suspends prediction, corresponding to the down regulated prediction from the high- to low-level area. Invisible PE, on the other hand, only induced enhanced feedforward connectivity from right MT+ to PPC.

Common to both the visible and invisible PE were significant modulations of the self-connections within lower-level visual areas of V1 and MT+. This suggests that both types of PE rely (in part) on a release from adaptation (i.e., repetition suppression) upon deviant motion onset (i.e., change in stimulus statistics). Whether the adaptation effects were observed at V1 or MT+ depended upon whether this change was consciously perceived. That is, visible change PE relied on adaptation at bilateral MT+, whilst invisible change PE relied on adaptation at bilateral V1 (and left MT+). One possible explanation for the stronger effects observed for visible PE at MT+ than V1 could be related to the subcortical visual motion pathway that carries visual information directly from the lateral geniculate nucleus to MT+ (Sincich et al., 2004). It is possible that when the motion signal is more coherent (i.e., stronger), this subcortical pathway is also activated (in addition to that between V1 and MT+), leading to a stronger motion signal in MT+. Subsequently, if MT+ is more highly activated compared to V1, this may lead to greater adaptation effects upon deviant motion presentation. Alternatively, when the motion signal is weak (i.e., low-coherence), this subcortical route is not activated, and thus, V1 and MT+ may have comparable levels of adaptation (as observed in our invisible PE DCMs). Findings that the generators of visual PE to motion changes are located in motion centers or the dorsal pathway itself have been suggested previously by other studies of visible motion direction changes (Pazo-Alvarez et al., 2003; Kremláček et al., 2006). We add to these findings by showing that this still holds when the changes are invisible.

## CONCLUSION

We provide new insights into the brain mechanisms underpinning visual change detection, even in the absence of awareness, when task reporting is not required. We lend



support for visual PE in response to both consciously and non-consciously perceivable changes; with the former evidenced as stronger and more widespread cortical activity. Our findings suggest hemispheric lateralization within the left hemisphere when motion changes were visible. Using DCM, we found that both types of PE were generated via a release from adaptation in sensory areas responsible for visual motion processing. The overall pattern emerging from our study reveals a complex picture of down- and up-regulation of feedforward and feedback connectivity in relation to conscious awareness of changes. To test the generality of our findings, further investigations are necessary, especially with techniques that explicitly manipulate conscious awareness under comparable task conditions testing for the neuronal effects on prediction and surprise.

## DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of ethical restrictions. Requests to access the datasets should be directed to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The University of Queensland Ethics Committee.

## REFERENCES

- Anderson, R. A. (1989). Visual and eye movement functions of the posterior parietal cortex. *Annu. Rev. Neurosci.* 12, 377–403. doi: 10.1146/annurev.ne.12.030189.002113
- Bendixen, A., and Andersen, S. K. (2013). Measuring target detection performance in paradigms with high event rates. *Clinic. Neurophysiol.* 124, 928–940. doi: 10.1016/j.clinph.2012.11.012
- Bernat, E., Shevrin, H., and Snodgrass, M. (2001). Subliminal visual oddball stimuli evoke a P300 component. *Clinic. Neurophysiol.* 112, 159–171. doi: 10.1016/s1388-2457(00)00445-4
- Berti, S. (2011). The attentional blink demonstrates automatic deviance processing in vision. *Neuro Rep.* 22, 6647–6667.
- Boly, M., Garrido, M. I., Gosseries, O., Bruno, M.-A., Boveroux, P., Schnakers, C., et al. (2011). Preserved feedforward but impaired top-down processes in the vegetative state. *Science* 332, 858–862. doi: 10.1126/science.1202043
- Born, R. T., and Bradley, D. C. (2005). Structure and function of visual area MT. *Annu. Rev. Neurosci.* 28, 157–189. doi: 10.1146/annurev.neuro.26.041002.131052
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision* 10, 433–436. doi: 10.1163/156856897x00357
- Brazdil, M., Rektor, I., Daniel, P., Dufek, M., and Jurak, P. (2001). Intracerebral event-related potentials to subthreshold target stimuli. *Clinic. Neurophysiol.* 112, 650–661. doi: 10.1016/s1388-2457(01)00463-1
- Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* 12, 4745–4765. doi: 10.1523/jneurosci.12-12-04745.1992
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 1–73.
- Czigler, I. (2014). Visual mismatch negativity and categorization. *Brain Topogr.* 27, 590–598. doi: 10.1007/s10548-013-0316-8
- Czigler, I., Weisz, J., and Winkler, I. (2007). Backward masking and visual mismatch negativity: Electrophysiological evidence for memory-based detection of deviant stimuli. *Psychophysiology* 44, 610–619. doi: 10.1111/j.1469-8986.2007.00530.x
- Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.-F., Poline, J.-P., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.* 4, 752–758. doi: 10.1038/89551
- Dehaene, S., Naccache, L., Le Clec'h, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., et al. (1998). Imagining unconscious semantic priming. *Nature* 395, 597–600.
- Del Cul, A., Baillet, S., and Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol.* 5:e260. doi: 10.1371/journal.pbio.0050260
- DeYoe, E. A., and Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends Neurosci.* 11, 19–226.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034
- Faivre, N., Arzi, A., Lungchi, C., and Salomon, R. (2017). Consciousness is more than meets the eye: call for a multisensory study of subjective experience. *Neurosci. Conscious.* 2017:nix003.
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1
- Frässle, S., Sommer, J., Jansen, A., Naber, M., and Einhäuser, W. (2014). Binocular rivalry: Frontal activity related to introspection and action but not to perception. *J. Neurosci.* 34, 1738–1747. doi: 10.1523/jneurosci.4403-13.2014
- Friston, K. (2005). A theory of cortical responses. *Philosophic. Trans. R. Soc. B* 360, 815–836.
- Friston, K. J., and Stephan, K. E. (2007). Free-energy and the brain. *Syntheses* 159, 417–458. doi: 10.1007/s11229-007-9237-y
- Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *NeuroImage* 19, 1273–1302. doi: 10.1016/s1053-8119(03)00202-7
- Friston, K. J., Trujillo-Barreto, N., and Daunizeau, J. (2008). DEM: a variational treatment of dynamic systems. *NeuroImage* 41, 849–885. doi: 10.1016/j.neuroimage.2008.02.054

The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

ER collected and analyzed data and wrote the first draft of the manuscript. All authors designed the study and edited the manuscript.

## FUNDING

This work was funded by a The University of Queensland Fellowship (2016000071) and the ARC (Australian Research Council) Centre of Excellence for Integrative Brain Function (ARC Centre Grant CE140100007) to MG, an ARC Future Fellowship (FT120100619) to NT, as well as ARC Discovery Projects: DP180104128 to MG and NT and DP180100396 to NT.

## ACKNOWLEDGMENTS

We thank the volunteers for participating in this study. We also thank both reviewers for their insightful comments, which helped improve the manuscript.

- Gaillard, R., Dehaene, S., Adam, C., Clémenceau, S., Hasboun, D., Baulac, M., et al. (2009). Converging intracranial markers of conscious access. *PLoS Biol.* 7:e1000061. doi: 10.1371/journal.pbio.1000061
- Garrido, M., Kilner, J. M., Kiebel, S. J., and Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proc. Natl. Acad. Sci.* 104, 20961–20966. doi: 10.1073/pnas.0706274105
- Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication. Does the corpus callosum enable the human condition? *Brain* 123, 1293–1326. doi: 10.1093/brain/123.7.1293
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics Vol. 1*, New York, NY: Wiley.
- Hohwy, J. (2013). *The Predictive Mind*. New York, NY: Oxford University Press.
- Ilg, U. J., ASchumann, S., and Thier, P. (2004). Posterior parietal cortex neurons encode target motion in world-centered coordinates. *Neuron* 43, 145–151. doi: 10.1016/j.neuron.2004.06.006
- Jack, B. N., Roeber, U., and O'Shea, R. P. (2015). We make predictions about eye of origin of visual input: Visual mismatch negativity from binocular rivalry. *J. Vision* 15, 1–19. doi: 10.1075/aicr.90.01mil
- Jack, B., Widmann, A., O'Shea, R. P., Schroger, E., and Roeber, U. (2017). Brain activity from stimuli that are not perceived: visual mismatch negativity during binocular rivalry suppression. *Psychophysiology* 54, 755–763.
- Kimura, Y., Yoshino, A., Takahashi, Y., and Nomura, S. (2004). Interhemispheric difference in emotional response without awareness. *Physiol. Behav.* 82, 727–731. doi: 10.1016/j.physbeh.2004.06.010
- Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception* 36, 1–16.
- Kogai, T., Aoyama, A., Amano, K., and Takeda, T. (2011). Visual mismatch response evoked by a perceptually indistinguishable oddball. *Neuro Rep.* 22, 535–538. doi: 10.1097/wnr.0b013e328348ab76
- Kohn, A., and Movshon, J. A. (2004). Adaptation changes the direction tuning of macaque MT neurons. *Nat. Neurosci.* 7, 764–772. doi: 10.1038/nn1267
- Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., and de Lange, F. P. (2012). Attention reverses the prediction in silencing sensory signals. *Cerebral Cortex* 22, 2197–2206. doi: 10.1093/cercor/bhr310
- Kremláček, J., Kreegipuu, K., Tales, A., Astikainen, P., Pöldver, N., Näätänen, R., et al. (2016). Visual mismatch negativity (vMMN): A review and meta-analysis of studies in psychiatric and neurological disorders. *Cortex* 80, 76–112. doi: 10.1016/j.cortex.2016.03.017
- Kremláček, J., Kuba, M., Kubová, Z., and Langrová, J. (2006). Visual mismatch negativity elicited by magnocellular system activation. *Vision Res.* 46, 485–490. doi: 10.1016/j.visres.2005.10.001
- Kremláček, J., Kuba, M., Kubová, Z., Langrová, J., Szanyi, J., Vit, F., et al. (2013). Visual mismatch negativity in the dorsal stream is independent of concurrent visual task difficulty. *Front. Hum. Neurosci.* 7:411. doi: 10.3389/fnhum.2013.00411
- Kuldkepp, N., Kreegipuu, K., Raidvee, A., Näätänen, R., and Allik, J. (2013). Unattended and attended visual change detection of motion as indexed by event-related potentials and its behavioural correlate. *Front. Hum. Neurosci.* 7:476. doi: 10.3389/fnhum.2013.00476
- LeDoux, J. E. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York, NY: Simon and Schuster.
- Meneguzzo, P., Tsakiris, M., Schioth, H. B., Stein, D. J., and Brooks, S. J. (2014). Subliminal versus supraliminal stimuli activate neural responses in anterior cingulate cortex, fusiform gyrus and insula: a meta-analysis of fMRI studies. *BMC Psychol.* 2, 1–11. doi: 10.1186/s40359-014-0052-1
- Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale, NJ: Lawrence Erlbaum.
- Näätänen, R., Gaillard, A. W. K., and Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313–329. doi: 10.1016/0001-6918(78)90006-9
- O'Shea, R. P. (2015). Refractoriness about adaptation. *Front. Hum. Neurosci.* 9:38. doi: 10.3389/fnhum.2015.00038
- O'Shea, R. P., Kornmeier, J., and Roeber, U. (2013). Predicting visual consciousness electrophysiologically from intermittent binocular rivalry. *PLoS One* 8:e76134. doi: 10.1371/journal.pone.0076134
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York: Oxford University Press.
- Pazo-Alvarez, P., Cadaveira, F., and Amenendo, E. (2003). MMN in the visual modality: a review. *Biol. Psychol.* 63, 199–236. doi: 10.1016/s0301-0511(03)00049-8
- Pearl, J. (2000). *Causality*, Cambridge: Cambridge University Press.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* 10, 437–442. doi: 10.1163/156856897x00366
- Penny, W., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., et al. (2010). Comparing families of Dynamic Causal Models. *PLoS Comput. Biol.* 6:e1000709. doi: 10.1371/journal.pcbi.1000709
- Plomp, G., Hervais-Adelman, A., Astofli, L., and Michel, C. M. (2015). Early recurrence and ongoing parietal driving during elementary visual processing. *Sci. Rep.* 5:18733.
- Rensink, R. A. (2002). Change detection. *Annu. Rev. Psychol.* 53, 245–277.
- Rensink, R. A. (2004). Visual sensing without seeing. *Psychol. Sci.* 15, 27–32. doi: 10.1111/j.0963-7214.2004.01501005.x
- Roser, M., Fugelsang, J. A., Dunbar, K. N., Corballis, P. M., and Gazzaniga, M. S. (2005). Dissociating processes supporting causal perception and causal inference in the brain. *Neuropsychology* 19, 591–602. doi: 10.1037/0894-4105.19.5.591
- Schepman, A., Rodway, P., and Pritchard, H. (2016). Right-lateralised unconscious, but not conscious, processing of affective environmental sounds. *Laterality* 21, 606–632. doi: 10.1080/1357650x.2015.1105245
- Sincich, L. C., Park, K. R., Wohlgenuth, M. J., and Horton, J. C. (2004). Bypassing V1: A direct geniculate input to area MT. *Nat. Neurosci.* 7, 1123–1128. doi: 10.1038/nn1318
- Stefanics, G., Kremláček, J., and Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Front. Hum. Neurosci.* 8:666. doi: 10.3389/fnhum.2014.00666
- Stephan, K. E., Penny, W., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage* 46, 1004–1017. doi: 10.1016/j.neuroimage.2009.03.025
- Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., and Friston, K. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage* 49, 3099–3109. doi: 10.1016/j.neuroimage.2009.11.015
- Summerfield, C., and Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends Cognit. Sci.* 13, 403–409. doi: 10.1016/j.tics.2009.06.003
- Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., and Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science* 314, 1311–1314. doi: 10.1126/science.1132028
- Tsuchiya, N., Frässle, S., Wilke, M., and Lamme, V. (2016). No-report and report-based paradigms jointly unravel the NCC: Response to Overgaard and Fazekas. *Trends Cognit. Sci.* 20, 242–243. doi: 10.1016/j.tics.2016.01.006
- Tsuchiya, N., Wilke, M., Frässle, S., and Lamme, V. A. F. (2015). No-report paradigms Extracting the true neural correlates of consciousness. *Trends Cognit. Sci.* 19, 757–770. doi: 10.1016/j.tics.2015.10.002
- Tsushima, Y., Sasaki, Y., and Watanabe, T. (2006). Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science* 314, 1786–1788. doi: 10.1126/science.1133197
- van Rhijn, M., Roeber, U., and O'Shea, R. P. (2013). Can the eye of origin serve as a deviant? Visual mismatch negativity from binocular rivalry. *Front. Hum. Neurosci.* 7:190. doi: 10.3389/fnhum.2013.00190
- van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., et al. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science* 360, 537–542. doi: 10.1126/science.aar7186
- Wager, T. D., Keller, M. C., Lacey, S. C., and Jonides, J. (2005). Increased sensitivity in neuroimaging analyses using robust regression. *NeuroImage* 26, 99–113. doi: 10.1016/j.neuroimage.2005.01.011
- Watanabe, T., Nanez, J. E., and Sasaki, Y. (2011). Perceptual learning without perception. *Nature* 413, 844–848. doi: 10.1038/35101601
- Wilke, M., Mueller, K.-M., and Leopold, D. A. (2009). Neural activity in the visual thalamus reflects perceptual suppression. *Proc. Natl. Acad. Sci.* 106, 9465–9470. doi: 10.1073/pnas.0900714106
- Williams, L. M., Liddell, B. J., Kemp, A. H., Bryant, R. A., Meares, R. A., Peduto, A. S., et al. (2006). Amygdala-prefrontal dissociation of subliminal and supraliminal fear. *Hum. Brain Map.* 27, 652–661. doi: 10.1002/hbm.20208

Wolford, G., Miller, M. B., and Gazzaniga, M. (2000). The left hemisphere's role in hypothesis formation. *J. Neurosci.* 20, 1–4.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Rowe, Tsuchiya and Garrido. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Up and Down States During Slow Oscillations in Slow-Wave Sleep and Different Levels of Anesthesia

Melody Torao-Angosto<sup>1</sup>, Arnau Manasanch<sup>1</sup>, Maurizio Mattia<sup>2</sup>  
and Maria V. Sanchez-Vives<sup>1,3\*</sup>

<sup>1</sup>Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain, <sup>2</sup>National Center for Radioprotection and Computational Physics, Istituto Superiore di Sanità, Rome, Italy, <sup>3</sup>Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

## OPEN ACCESS

### Edited by:

Heiko J. Luhmann,  
Johannes Gutenberg University  
Mainz, Germany

### Reviewed by:

Tomasz Błasiak,  
Jagiellonian University, Poland  
Andrea Piarulli,  
University of Pisa, Italy

### \*Correspondence:

Maria V. Sanchez-Vives  
msanche3@clinic.cat

**Received:** 23 September 2020

**Accepted:** 12 January 2021

**Published:** 09 February 2021

### Citation:

Torao-Angosto M, Manasanch A, Mattia M and Sanchez-Vives MV (2021) Up and Down States During Slow Oscillations in Slow-Wave Sleep and Different Levels of Anesthesia. *Front. Syst. Neurosci.* 15:609645. doi: 10.3389/fnsys.2021.609645

Slow oscillations are a pattern of synchronized network activity generated by the cerebral cortex. They consist of Up and Down states, which are periods of activity interspersed with periods of silence, respectively. However, even when this is a unique dynamic regime of transitions between Up and Down states, this pattern is not constant: there is a range of oscillatory frequencies (0.1–4 Hz), and the duration of Up vs. Down states during the cycles is variable. This opens many questions. Is there a constant relationship between the duration of Up and Down states? How much do they vary across conditions and oscillatory frequencies? Are there different sub regimes within the slow oscillations? To answer these questions, we aimed to explore a concrete aspect of slow oscillations, Up and Down state durations, across three conditions: deep anesthesia, light anesthesia, and slow-wave sleep (SWS), in the same chronically implanted rats. We found that light anesthesia and SWS have rather similar properties, occupying a small area of the Up and Down state duration space. Deeper levels of anesthesia occupy a larger region of this space, revealing that a large variety of Up and Down state durations can emerge within the slow oscillatory regime. In a network model, we investigated the network parameters that can explain the different points within our bifurcation diagram in which slow oscillations are expressed.

**Keywords:** Up states, Down states, slow oscillations, sleep, anesthesia, cerebral cortex, cortical model, slow-wave sleep

## INTRODUCTION

Slow oscillations are an emergent pattern of the cortical network that also recruit subcortical nuclei and, in particular, the cortico-thalamocortical loop. Slow oscillations are the hallmark of slow-wave sleep (SWS), and much research has been carried out to try to understand their role in sleep-induced plasticity and memory consolidation (Diekelmann and Born, 2010; Klinzing et al., 2019). However, this emergent pattern of activity consistent in slow oscillations is generated by the cerebral cortex not only during SWS but also in a variety of pharmacologically induced states (e.g., following administration of propofol, ketamine,



urethane, or isoflurane; for a review, see Brown et al., 2010) and also in pathological conditions such as stroke or traumatic brain injury (Sarasso et al., 2019; Cassidy et al., 2020). For this reason, it is important to understand the genesis and dynamics of slow waves (Sanchez-Vives, 2020), as well as the brain state transitions that lead the cerebral cortex in and out of these dynamic regimes.

In this study, we investigated the characteristics of Up and Down states in SWS, deep anesthesia, and lighter anesthesia, in an attempt to generate a “map” of the slow oscillatory space which acts as a common framework for the slow oscillations that are generated under different physiological, pathological, pharmacological, or experimental conditions.

## MATERIALS AND METHODS

### Surgery and Chronic Implants

All experiments were carried out following Spanish regulatory laws (BOE-A-2013-1337), which comply with European Union guidelines and were supervised and approved by the Animal Experimentation Ethics Committee of the Universitat de Barcelona (287/17 P3).

To obtain local field potential (LFP) long-term recordings in the freely moving rat (Lister-Hooded, 6–10 months old), we carried out chronic implants of electrodes that were inserted 600 microns deep in the cerebral cortex using a stereotaxic apparatus (Kopf Instruments, Tujunga, CA, USA). The recording electrodes were handmade by twisting stainless steel Teflon-insulated wires of 100  $\mu\text{m}$  diameter (California Fine Wire Co., CA, USA). For the recording of the EMG from the neck muscle, a single electrode was made with a 125  $\mu\text{m}$  tungsten-insulated wire (Advent Research Materials Limited, Oxford, UK). The EMG was connected to one of the channels of the preamplifier (Multi Channel Systems, Germany) and acquired at 500 Hz. All signals were amplified  $\times 1000$ .

The surgery to chronically implant the electrodes was performed with the animal deeply anesthetized with isoflurane (2%). Five to six anchoring stainless steel screws were placed in the skull. A screw located over the cerebellum and connected using a soldered wire was used as ground. The craniotomy was made following the stereotaxic coordinates for the primary visual cortex (V1;  $-7.3$  mm AP,  $2.2$  mm ML,  $-0.6$  mm DV; Paxinos and Watson, 2007).

After the placement of the electrodes, these were fixed to the skull with an initial application of glue (Loctite 406, Henkel, Germany) and then a second layer of glass ionomer luting cement (Medicaline, Geestland, Germany). Once fixed, the electrodes were welded to the contacts of the case blank connector with crimping contacts (Molex, IL, USA) and in the final step of the surgery, the case was attached to the skull with glass ionomer luting cement. Buprenorphine (0.06 mg/kg) and enrofloxacin (25 mg/kg) were administered for a minimum of 5 days after surgery for analgesia and the prevention and treatment of possible infections. After the post-surgical treatment period, 5 days of handling and habituation to the recording chamber

were performed before the initiation of brain recordings, to minimize stress and abrupt movements of the animal during the experimental sessions.

### Recording Protocols

LFPs were recorded from the primary visual cortex of the chronically implanted rats during different anesthesia levels and their natural sleep-wake cycle. For this purpose, the subject was first connected to the headstage preamplifier (Multi Channel Systems, Germany) using a custom-made adapter (IMB-CNM, CSIC) to the implanted case. Then, the animal was placed in a plastic recording cage ( $57 \times 39 \times 42$  cm), while being videotaped and recorded. The animals were free during all the recordings (not in a stereotaxic). Recordings were acquired, digitized at 5 kHz using a data acquisition interface and Spike 2 software (Cambridge Electronic Design, Cambridge, UK).

### Slow-Wave Sleep Recordings

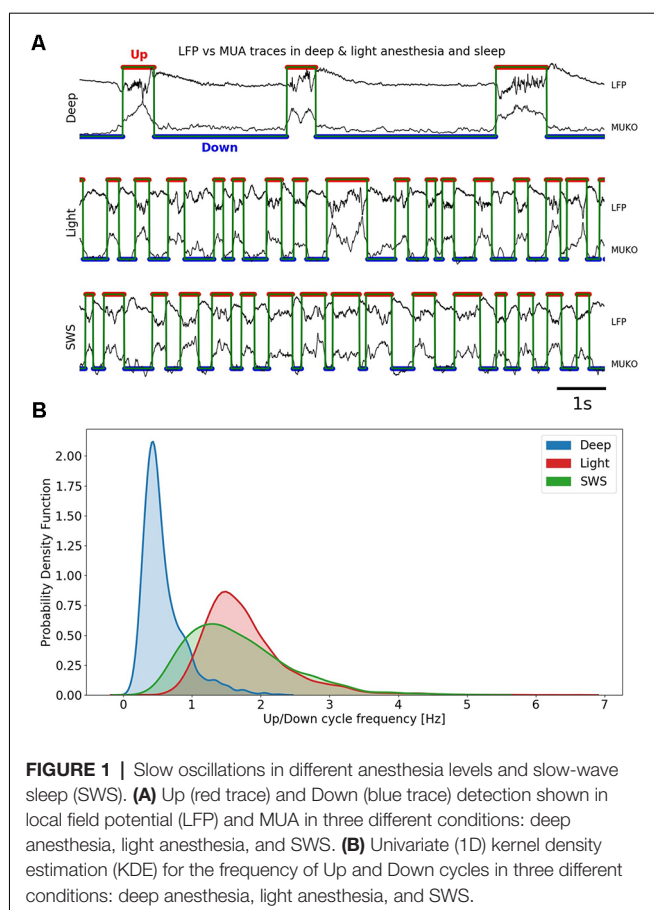
After the post-surgical recovery period, the animals ( $n = 2$ ) were recorded daily for several hours during their natural sleep-wake cycle of sleep for a minimum of 3 days. SWS periods were classified based on LFP, EMG, and video following the scale for sleep scoring by Iber et al. (2007). Only periods of SWS were included in the current study.

### Anesthesia Recordings

First, a baseline recording was obtained with the awake animal for a minimum of 30 min. Afterward, a mixture of ketamine (Ketolar 50 mg/ml) and medetomidine (Domtor 1 mg/ml) was administered intraperitoneally. The administration of a single subcutaneous injection of atropine immediately after the anesthetic induction (0.05 mg/kg) was part of the anesthetic routine, to reduce bronchial secretions (Sanchez-Vives et al., 2000; Bettinardi et al., 2015; Tort-Colet et al., 2019; Redinbaugh et al., 2020) preventing respiratory obstructions in the deepest phase of the anesthesia. A dose of 0.6 ml of saline was subcutaneously injected every 2 h during the anesthesia for hydration. The rectal temperature was monitored during anesthesia and maintained at  $37^\circ\text{C}$  during the recording using a probe and an electric blanket. Two different doses of anesthesia were administered: “light” anesthesia in three animals (20 mg/kg of ketamine and 0.15 mg/kg of medetomidine), and “deep anesthesia” in four animals (40 mg/kg of ketamine and 0.3 mg/kg of medetomidine). Notice that the animals were free during all the recordings (not in a stereotaxic), therefore the anesthesia doses were given only to study the cortical effects and without the need to reach a surgical state of anesthesia. Cortical activity was recorded beginning at injection (induction), throughout the slow oscillatory period of anesthesia until the complete fade-out of the anesthetic effect, all the way to wakefulness.

### Data Analysis

Raw signals were first downsampled to 3 kHz for computational time reduction purposes. For each recording belonging to an anesthetic condition, the first 10–20 min after a regular establishment of the slow oscillation were selected to compute the Up and Down state duration in the SO cycles. SWS periods in sleep recordings were selected along the entire experiment.



**FIGURE 1 |** Slow oscillations in different anesthesia levels and slow-wave sleep (SWS). **(A)** Up (red trace) and Down (blue trace) detection shown in local field potential (LFP) and MUA in three different conditions: deep anesthesia, light anesthesia, and SWS. **(B)** Univariate (1D) kernel density estimation (KDE) for the frequency of Up and Down cycles in three different conditions: deep anesthesia, light anesthesia, and SWS.

Periods of approximately 4 min of SO cycles were extracted per recording. To quantify the durations of Up and Down states, we used the same method as described in *Dasilva et al. (2021)*. Three different characteristics were extracted from the raw signal (LFP) to perform the detection of the Up and Down states (MUKO method, see **Figure 1A**): the Slow Oscillation deflection (SO), the gamma rhythm, and the firing of the local network (MUA). These characteristics were obtained as time-series: the decimated raw signal, the envelope of the variance of the gamma-filtered signal (*Mukovski et al., 2007*) and the estimation of the MUA signal from the power of the frequencies in the band between 200 and 1,500 Hz computed in 5 ms windows (*Mattia et al., 2010; Reig et al., 2010; Sanchez-Vives et al., 2010; Ruiz-Mejias et al., 2011*). To compensate for the high fluctuations in the firing of neurons that are close to the electrode, MUA signal values were logarithmically scaled, thus obtaining the LogMUA signal (*Ruiz-Mejias et al., 2011*).

The multivariate time series composed by the three above signals individually z-score normalized were processed relying on a principal component analysis (PCA). Projections on the first principal component resulted in a bimodal distribution, such that the two peaks of the distribution corresponded to samples of network activity belonging to either Up (higher activity) or Down (lower activity) states. Such segregation allowed us to choose a threshold that best separated the Up states from the Down states

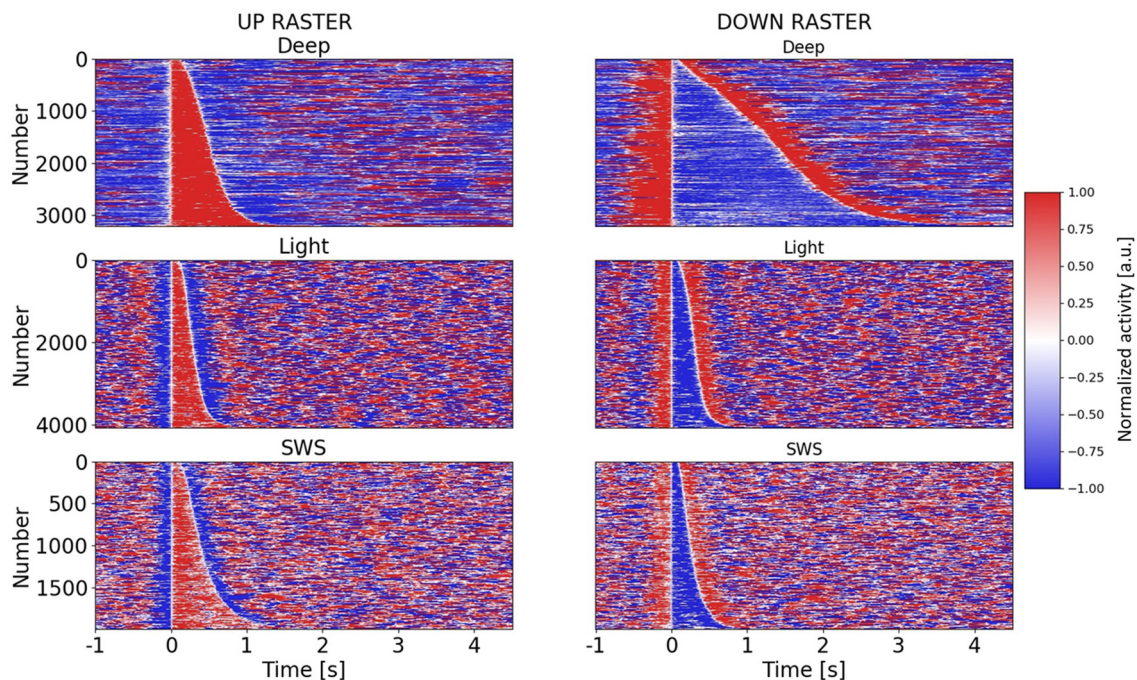
(**Figure 2**). To avoid the detection of random fluctuations of the signal that could be detected as the Up States, a minimum duration for the Up and Down durations of 80 ms was set. This threshold was heuristically set after visual inspection of the signal.

For each Up and Down cycle, the duration for both the Up and Down states was computed (**Figure 2**). Once extracted, they were scattered one against the other to construct the 2D space of points coloring by the condition. Kernel density estimation (KDE) was used to obtain and construct univariate (1D histogram estimate) and bivariate (2D histogram estimate) plots of the Up and Down state durations.

## Simulations

Networks of leaky integrate-and-fire (LIF) neurons (6,880 excitatory and 2,580 inhibitory, E and I, respectively) were simulated relying on an event-based numerical integration (*Mattia and Del Giudice, 2000*). The network parameters were chosen to model a self-consistent network of Layer 5 neurons each having an average number of pre-synaptic connections computed considering the connection probabilities described in *Markram et al. (2015)*, the cellular densities reported in *Beaulieu (1993)* and the exponential decay of the cortico-cortical connectivity across layers reported in *Schnepel et al. (2015)* and in *Kätzel et al. (2011)* for excitatory and inhibitory pre-synaptic sources, respectively. As result, excitatory neurons received on average 2,910 (129) synaptic contact from E (I) neurons, while the inhibitory ones received 746 (45) connections from E (I) pre-synaptic cells. To model a 1 mm<sup>2</sup>-patch of the cortex only 2% of excitatory connections were considered to be recurrent. Synaptic efficacies were set to have—under mean-field approximation—a fixed point with a firing rate of 0.75 Hz and 4.375 Hz for excitatory and inhibitory neurons (*Watson et al., 2016*), respectively.

More specifically, for all neurons, the emission threshold was 20 mV and the reset potential was 15 mV, reached by the membrane potential after a refractory period of 2 (1) ms for E (I) neurons following the emission of each spike. Membrane capacitance  $C_m$  was arbitrarily set to 1 leading to express currents in units of mV/s. Decay constant  $\tau_m$  of the membrane potential was 20 and 10 ms for E and I neurons, respectively. Each neuron received as background synaptic noise a Poisson spike train with frequency  $\nu_{\text{ext}}$  (2,296 Hz and 586.4 Hz for E and I neurons, respectively) transduced in current by instantaneous synaptic transmission with efficacies  $J_{\text{ext}}$  of 0.48 mV and 2.2 mV. The mean external synaptic current was  $I_{\text{ext}} = J_{\text{ext}} \nu_{\text{ext}}$ . The probability  $c_{\alpha\beta}$  of having a synapse between pre- and post-synaptic neurons  $\alpha$  and  $\beta \in \{E, I\}$ , respectively, was  $\{c_{EE}, c_{EI}, c_{IE}, c_{II}\} = \{0.6, 5, 0.2, 1.7\}\%$ . Synaptic transmission was instantaneous with efficacies  $\{J_{EE}, J_{EI}, J_{IE}, J_{II}\} = \{1.9, -1.1, 2.2, -1.1\}$  mV. Spikes from E(I) neurons were delivered with a randomly chosen axonal delay sampled from an exponential distribution with a mean of 22.6 ms (5.7 ms), respectively. All synaptic efficacies were randomly sampled from a Gaussian distribution with mean  $J_{\alpha\beta}$  and a relative standard deviation of 25%. In addition to synaptic currents, each excitatory neuron received an additional potassium current  $-g_a a(t)$  modeling spike-frequency adaptation with strength  $g_a$  and adaptation level  $\dot{a} = -\frac{a}{\tau_a} + \sum_k \delta(t - t_k)$



**FIGURE 2 |** Raster plots of Up and Down states in different anesthesia levels and SWS. Raster plots for all Up and Down states in three different conditions: deep anesthesia, light anesthesia, and SWS. Up and Down states are ordered by duration. For each row in the raster plot, a color-based visualization of the normalized activity (see “Materials and Methods” section) is shown. A 5.5 s (–1 to 4.5 s) window is displayed, with Up or Down states starting at  $t = 0$  s.

(Gigante et al., 2007; Mattia and Sanchez-Vives, 2012) having unitary jumps at each spike emission time  $t_k$  of the neuron and decaying with the characteristic time  $\tau_a = 0.15$  s. For each parameter set ( $I_{\text{ext}}, g_a$ ) we simulated five networks with randomly extracted synaptic coupling for a time span of 100 s.

## RESULTS

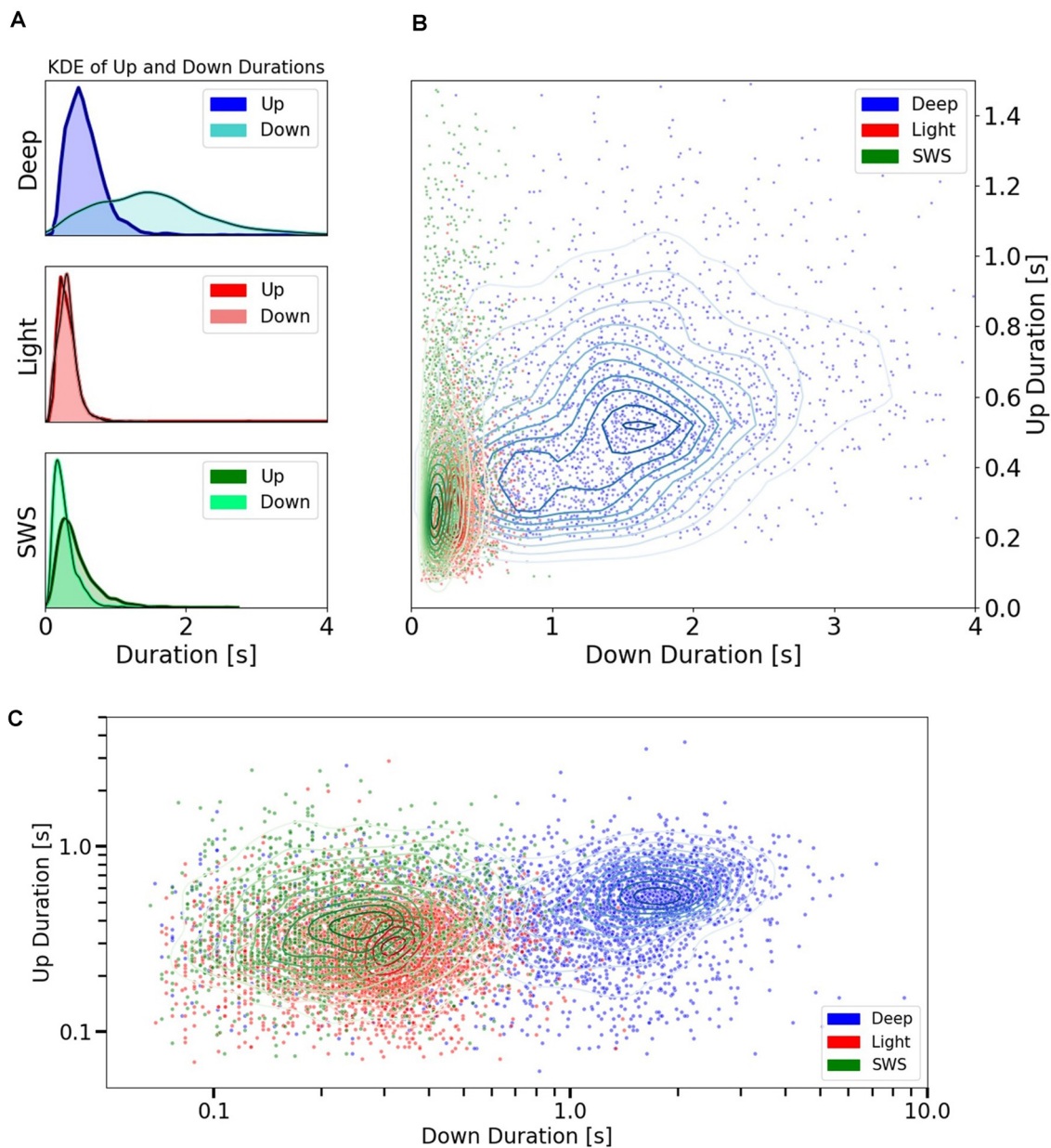
To investigate the detailed dynamics of cortical slow oscillations, we studied Up and Down state durations during the slow oscillatory regime *in vivo*. The recordings were obtained from the visual cortex of chronically implanted rats under three conditions: SWS, deep anesthesia, and light anesthesia. The data included here comprises six recording periods in SWS ( $n = 2$  rats), five recording periods in deep anesthesia ( $n = 5$  rats), and four recording periods in light anesthesia ( $n = 2$  rats). In total, 9,271 cycles were analyzed.

The population firing rate was obtained from the recordings of the local field potential (LFP; see “Materials and Methods” section). In **Figure 1A**, we illustrate traces corresponding to deep anesthesia (top), light anesthesia (middle), and SWS (bottom). The detection of Up and Down states were automatically performed and based on the Slow Oscillation deflection (SO), the gamma rhythm, and the firing of the local network (MUA; see “Materials and Methods” section; MUKO method, **Figure 1A**); Up and Down states were identified by the red and blue traces, respectively). The frequency of the slow oscillations ranged between 0.1 and 4 Hz, and **Figure 1B** represents the distribution

of frequencies in the three studied conditions. In deep anesthesia, the frequency of oscillation ranged between 0.1 and 2 Hz, but the peak was at 0.45 Hz. In lighter anesthesia, the range of oscillatory frequency was displaced towards higher frequency values, as we have previously described (Tort-Colet et al., 2019; Dasilva et al., 2021). The range of frequencies for light anesthesia was 0.5–3.5 Hz, with a peak at 1.5 Hz. The frequency of slow oscillations in SWS largely overlapped with those in light anesthesia, although expanded towards lower frequencies, ranging between 0.2 and 3.5 Hz, and peaking at 1.3 Hz.

Therefore, including the three experimental conditions, the whole range of frequencies described for slow oscillations (0.1–4 Hz) was covered, opening the door to the investigation of the relationships between Up and Down state durations. The raster plots of the Up and Down states for all the cycles included in this study: 3,205 for deep anesthesia, 4,073 for light anesthesia, and 1,993 for SWS, are displayed in **Figure 2**. As can be seen in the figure, while the duration of Up (left) and Down (right) states was similar for light anesthesia and SWS, in deep anesthesia both Up and Down states were longer (see exact values next). In particular, Down states could last up to 4 s, and in a few cycles even 9 s (not displayed, see panel **Figure 3C**). In light anesthesia, the duration of Up and Down states overlapped (**Figure 3A**), with an average value of  $0.31 \pm 0.18$  s for Up states and  $0.31 \pm 0.13$  s for Down states, therefore the network spent a similar time in the firing periods (Up) and in silence (Down states). However, when anesthesia became deeper, the silent periods became longer, exceeding the periods of firing. The average duration





**FIGURE 3 |** Up and Down state durations in different anesthesia levels and SWS. **(A)** Univariate (1D) KDE (Kernel Density Estimation) for Up and Down state durations in three different conditions: deep anesthesia, light anesthesia, and SWS. **(B)** Scatter plot showing the space of durations for Up and Down cycles. Each point represents the duration of an Up vs. the duration of the subsequent Down. They are colored by conditions: Deep and Light anesthesia or SWS. Irregular ellipses stand for the bivariate (2D) KDE for both Up and Down durations in each of these conditions. **(C)** Scatter plot of Up vs. Down durations in the three conditions (deep and light anesthesia and SWS) represented with a logarithmic scale.

of Up states in deep anesthesia was  $0.51 \pm 0.33$  s, while that of Down states was  $1.37 \pm 0.83$  s (**Figure 3A**). SWS had similar durations to those in light anesthesia, with an average of  $0.45 \pm 0.28$  s for Up states and  $0.28 \pm 0.16$  s for Down states, with a slight tendency towards longer Down states (**Figure 3A**). For a statistical evaluation, a Mann-Whitney test comparing Up and Down state durations in the different conditions was performed. A comparison between Light ( $n = 4,073$ ) and Deep

( $n = 3,205$ ) anesthesia showed significantly different Up and Down durations, with a  $p = 0$  (given the high number of samples). The same was the case for a comparison between SWS ( $n = 1,993$ ) and Deep ( $n = 3,205$ ) anesthesia, with a  $p$ -value that was effectively 0 for Up state durations ( $p = 8.8 \cdot 10^{-29}$ ) and  $p = 0$  for Down state durations. A comparison between Up and Down durations in Light anesthesia ( $n = 4,073$ ) and SWS ( $n = 1,993$ ) also revealed a significant difference with a  $p$ -value



effectively 0 ( $p = 0$  and  $p = 2.6 \cdot 10^{-37}$ , respectively). Even though the formal statistical test showed that the center of location of each distribution was different—particularly given the large number of samples—the distribution of SWS and light anesthesia presented a large overlap (**Figures 3A–C**), suggesting that the dynamics in light anesthesia and SWS are fairly similar and different from deeper anesthesia. But what are the characteristics of those dynamics? What is the relationship between Up states and the subsequent Down states?

To explore the relationship between Up and Down states, we represented every cycle with their Up state and subsequent Down state (**Figure 3B**) to obtain cartography of the space covered by each condition. Deep anesthesia, in blue, covers a large area of the space of relation between Up states and Down states, while the area occupied by light anesthesia and SWS is more restricted, mainly in the sense of never displaying Down states longer than 0.80 s (99th percentile). This data reveals that in these two specific conditions, one physiological (SWS) and one pharmacologically induced (light anesthesia), there was a relatively tight relationship between Up state and Down state duration. However, this is not always the case, and the regime of slow oscillations can be expressed in a larger variety of Up and Down state durations. In **Figure 3C**, the same data as in **Figure 3B** is represented using a logarithmic scale, which allows better visualization of the distribution of cycles for SWS and light anesthesia.

What mechanisms can support such a variety of Up and Down state durations? To answer this question, we resorted to the cortical network model similar to Mattia and Sanchez-Vives (2012); (**Figure 4A**), by varying two parameters: (i) the cortico-cortical synaptic input ( $\Delta I_{\text{ext}}$ ) associated with changes of glutamatergic synaptic transmission; and (ii) the strength of spike-frequency adaptation ( $g_a$ ) related to activity-dependent hyperpolarizing  $K^+$  currents. These two key features shape the dynamical regime of the model networks giving rise to four different phases (**Figure 4B**). Two of them are single-attractor asynchronous states, one with high firing (similar to wakefulness) and the other with low firing (e.g., barbiturates, burst suppression-like). In the third bistable regime the network displayed simultaneously two possible stable asynchronous Up-(on) and Down-like (off) states, while in the fourth one, slow oscillations were spontaneously generated. In the parameter subspace where such oscillations were produced, the Up and Down state durations widely varied as shown in **Figure 4C**, where only the synaptic input changed. This in principle can highlight anticorrelations in the Up and Down state durations (Mattia and Sanchez-Vives, 2012; Levenstein et al., 2019; Nghiem et al., 2020) giving rise to “banana”-shaped distributions at fixed  $\Delta I_{\text{ext}}$  or  $g_a$  (**Figures 4E,F**). Hence, the slow oscillation phase of the cortical network is not an invariant regime; rather, it expresses a wide spectrum of timescales (Mattia and Sanchez-Vives, 2012), which in turn allows us to infer the effective values  $\Delta I_{\text{ext}}$  and  $g_a$  needed to reproduce the state duration statistics we measured in different experimental conditions (**Figure 4D**, as in **Figure 3C** but averaging chunks of 10 Up-Down cycles). As result, deep anesthesia statistics were produced by networks in the slow oscillation phase close to the boundary with the burst suppression-like regime

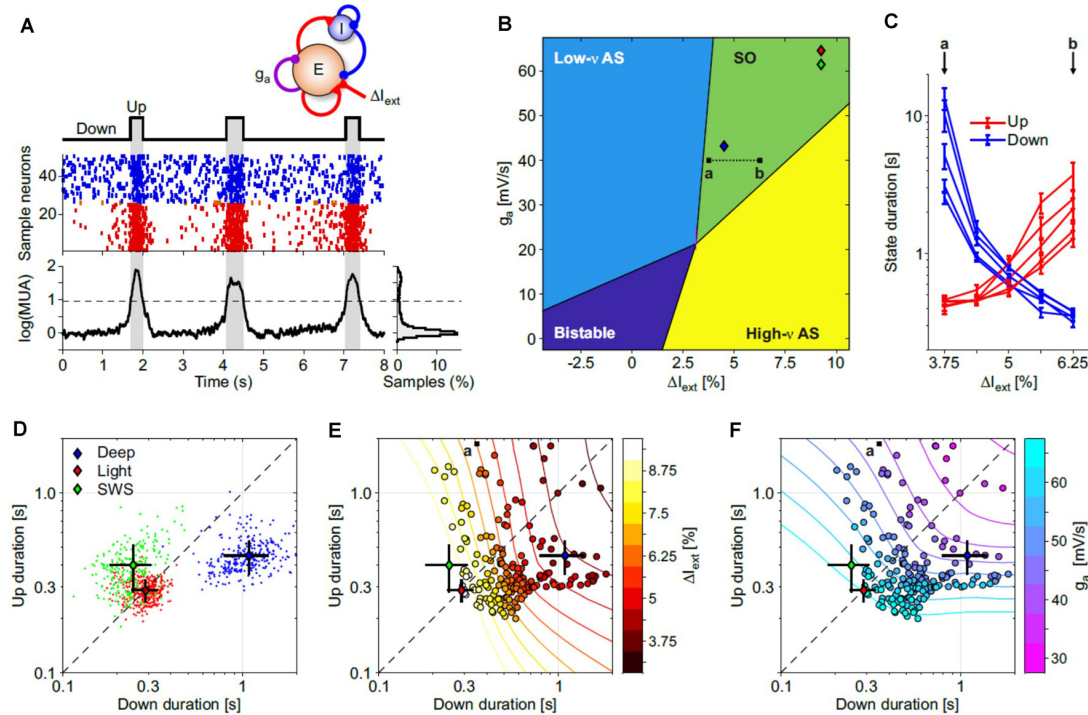
(**Figure 4B**, blue diamond): here the Down state was more preferable and as such had longer durations. Light anesthesia and natural SWS (red and green diamonds, respectively) had similar Up/Down duration statistics (**Figure 4D**) resulting in cortical networks with increased adaptation and cortico-cortical synaptic input. Under this condition the model network was closer to the other boundary separating the SO phase from the awake-like asynchronous state (green and yellow regions, respectively).

## DISCUSSION

In this article, we described slow oscillations as a single dynamic regime that comprises a range of frequencies and a range of Up and Down state durations. Therefore, the regime of slow oscillations is a wide region, which here we investigated with two levels of anesthesia—deep and light—and a physiological condition—SWS. We found that light anesthesia and SWS display similar dynamical features occupying a relatively narrow range of the Up and Down state duration space. Deep anesthesia instead occupies a region of this space displaying a much wider range of possible Up and Down state durations.

All these features can be reproduced in models of cortical networks composed of integrate-and-fire neurons (Gigante et al., 2007; Mattia and Sanchez-Vives, 2012; Jercog et al., 2017). Mean-field approximation and simulations allow us to work out a relatively rich bifurcation diagram for these models relying on two key “forces”: cortico-cortical synaptic strength and adaptation. By varying these two parameters, four spontaneous activity regimes emerge; slow oscillatory regime, bistability, asynchronous state (awake-like), and silent state (coma or barbiturate-like state). In this bifurcation diagram, our light anesthesia and SWS data lie relatively close to the border separating slow oscillations from the awake-like regime. This is a dynamical condition in which microarousals start to emerge (Tort-Colet et al., 2019), interspersed with the periods of coherent slow oscillations on the way to wakefulness. Therefore, light anesthesia, with a tight balance between Up and Down states durations, is a more excitable state which is, therefore, closer to wakefulness. Longer Down states are the expression of a less excitable network, corresponding therefore to a deeper state, further away from wakefulness. This is the case in deep anesthesia, which lies near the border of the state of silence. And in between, a subspace in which different combinations of Up and Down state durations, and thus different frequencies, can be expressed.

Why investigate the parameter space of the slow oscillatory regime? We consider this to be fundamental to understand the multiscale mechanisms that generate these patterns, how the multiple variables in the network move the emergent pattern in different directions or eventually induce a transition towards other global states. Like creating a cartographical map of cortical dynamics, in which we can place the different states induced by physiological (SWS) and pathological states (disorders of consciousness, lesions) or drugs (various anesthetics and doses) that result in slow oscillations, this allows a comparison of results from different studies under



**FIGURE 4 |** Exploring the Up and Down state durations in *in silico* cortical networks. **(A)** Spontaneous slow oscillations in a network of excitatory and inhibitory leaky integrate-and-fire (LIF) model neurons with spike-frequency adaptation (see “Materials and Methods” section for details). Adaptation is implemented only in excitatory neurons with the strength of self-inhibition  $g_a$  (see top sketch). Middle: emitted spikes by a subset of excitatory (red) and inhibitory (blue) neurons. From the firing rate  $v_E(t)$  of the excitatory neurons is extracted the multi-unit activity shown at the bottom [ $\log(MUA) = \log(v_E + 1)$  + offset, offset is set to have a lower peak of the distribution on the right at 0]. Up and Down states are detected by thresholding  $\log(MUA)$ ; dashed line. **(B)** Bifurcation diagram of the dynamical regimes expressed by the network model, as in Mattia and Sanchez-Vives (2012). For each point ( $\Delta I_{ext}$ ,  $g_a$ ), 100 s of simulations in five independent networks (randomly selected synaptic matrix).  $\Delta I_{ext}$  is the relative change in mean current received by excitatory neurons obtained by changing the external spike rate  $v_{ext}$ . Low- $v$  and high- $v$  AS (asynchronous state), neuronal spiking is irregular and  $v_E(t)$  fluctuates around a fixed value at relatively low and high firing rate, respectively. Bistable, the coexistence of both low- $v$  and high- $v$  AS. SO, the quasi-periodic slow alternation between Up and Down metastable states. Colored diamonds, networks with Up and Down state statistics as in deep and light anesthesia, and sleep experiments (see panel D). **(C)** Average Up and Down state durations (red and blue, respectively) computed in five independent networks with parameter changes depicted by the a-b dotted line in panel (B). **(D)** Distributions of average Up and Down durations from experiments under deep (blue) and light (red) anesthesia, and during natural NREM sleep (SWS, green). Colored diamonds, grand-averages (centroids) of the state durations for these three distributions. Black lines, standard deviations of the state durations. **(E,F)** Distribution of average Up and Down state durations from simulated networks (circles) used to work out the diagram in panel (B). Symbol colors represent  $\Delta I_{ext}$  and  $g_a$  in panels (E,F), respectively. Contour lines, isolevel curves of  $\Delta I_{ext}$  and  $g_a$  obtained by fitting with a smoothed surface the mean state durations from simulations. Diamonds as in panel (D). Fitted surfaces were used to infer the effective ( $\Delta I_{ext}$ ,  $g_a$ ) needed to reproduce experimental state duration statistics in simulations (diamonds in panel B).

the same framework. Furthermore, it is important to interpret the results from different experimental manipulations of slow oscillations (Deco et al., 2009; Sancristóbal et al., 2016; D’Andola et al., 2018) or different species, in a comprehensive dynamical framework.

Our study is not the first to compare SWS and anesthesia. A systematic comparison of intracellular and extracellular patterns during SWS vs. anesthesia was carried out by Chauvette et al. (2011). In this study, it was reported that silent (Down) states were longer under anesthesia than during SWS. These findings are compatible with the data included here since anesthesia induces longer Down states when it is deep, but we also find that light anesthesia induces Down states as short as SWS. Therefore, we find that it is not the anesthesia *per se* that creates different dynamics, but the level of anesthesia, which can result in emergent patterns very similar to SWS.

Interestingly, Nghiem et al. (2020) investigated Up and Down state durations in different species and conditions (SWS, anesthesia) and found a positive correlation between Up and Down state durations while in anesthesia. While in SWS however, Jercog et al. (2017) reported a negative correlation between Up and Down state durations. Our results also show a positive correlation between Up and Down state duration in anesthesia, but we also find a positive correlation in SWS. We previously found a negative correlation between the duration of Up and Down states in cortical slices (Mattia and Sanchez-Vives, 2012), in particular in networks in which inhibition is decreased and spike firing adaptation acquires a prominent role in the emergent dynamics (Sanchez-Vives et al., 2010). This is an interesting illustration of the fact that different mechanisms shape this emergent activity simultaneously, some of which have radically different effects. We propose here that a slight

change towards the dominance of short range or long range connections (Bettinardi et al., 2015; Dasilva et al., 2021), or in neuromodulators affecting spike firing adaptation (Barbero-Castillo et al., 2019; Nghiem et al., 2020) can tilt the ongoing slow oscillations in such a way that direct or inverse correlations between Up and Down state durations can be expressed. Indeed, in our network model we find simultaneously a direct correlation due to the simultaneous change of the two key parameters, while within each level an inverse correlation between Up and Down states is embedded (**Figures 4E,F**).

Slow oscillations have been thoroughly studied for the last quarter of the century [for a review, see Neske (2016), since the characterization made by Steriade et al. (1993)] of an activity known since the first days of EEG. It has been proposed to be the default activity model of the cortical circuit (Sanchez-Vives and Mattia, 2014), that acts as an attractor whenever the cortex becomes structurally or functionally disconnected (Sanchez-Vives et al., 2017) and plays an important role in the functional disruption caused by brain lesions (Sarasso et al., 2019). Still, this apparently simple highly synchronized pattern is difficult to understand in detail and leaves many questions regarding how local and global dynamics in the brain are generated, some of which we have highlighted in this study.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in Torao-Angosto et al. (2021).

## REFERENCES

- Barbero-Castillo, A., Riefolo, F., Matera, C., Caldas-Martínez, S., Mateos-Aparicio, P., Weinert, J. F., et al. (2019). Control of brain state transitions with light. *bioRxiv* [Preprint]. doi: 10.1101/793927
- Beaulieu, C. (1993). Numerical data on neocortical neurons in adult rat, with special reference to the GABA population. *Brain* 609, 284–292. doi: 10.1016/0006-8993(93)90884-p
- Bettinardi, R. G., Tort-Colet, N. N., Ruiz-Mejias, M., Sanchez-Vives, M. V., and Deco, G. (2015). Gradual emergence of spontaneous correlated brain activity during fading of general anesthesia in rats: Evidences from fMRI and local field potentials. *NeuroImage* 114, 185–198. doi: 10.1016/j.neuroimage.2015.03.037
- Brown, E. N., Lydic, R., and Schiff, N. D. (2010). General anesthesia, sleep and coma. *N. Engl. J. Med.* 363, 2638–2650. doi: 10.1056/NEJMra0808281
- Cassidy, J. M., Wodeyar, A., Wu, J., Kaur, K., Masuda, A. K., Srinivasan, R., et al. (2020). Low-frequency oscillations are a biomarker of injury and recovery after stroke. *Stroke*, 1442–1450. doi: 10.1161/STROKEAHA.120.028932
- Chauvette, S., Crochet, S., Volgushev, M., and Timofeev, I. (2011). Properties of slow oscillation during slow-wave sleep and anesthesia in cats. *J. Neurosci.* 31, 14998–15008. doi: 10.1523/JNEUROSCI.2339-11.2011
- D'Andola, M., Weinert, J. F., Mattia, M., and Sanchez-Vives, M. V. (2018). Modulation of slow and fast oscillations by direct current stimulation in the cerebral cortex *in vitro*. *bioRxiv* [Preprint]. doi: 10.1101/246819
- Dasilva, M., Camassa, A., Navarro-Guzman, A., Pazienti, A., Perez-Mendez, L., Zamora-López, G., et al. (2021). Modulation of cortical slow oscillations and complexity across anesthesia levels. *NeuroImage* 224:117415. doi: 10.1016/j.neuroimage.2020.117415
- Deco, G., Martí, D., Ledberg, A., Reig, R., and Sanchez Vives, M. V. (2009). Effective reduced diffusion-models: a data driven approach to the analysis of

## ETHICS STATEMENT

The animal study was reviewed and approved by Animal Experimentation Ethics Committee of the Universitat de Barcelona (287/17 P3).

## AUTHOR CONTRIBUTIONS

MT-A performed the surgeries and obtained the electrophysiological recordings. AM performed the data analysis. MS-V designed and supervised the study. MM performed the modeling work. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Spanish Ministry of Science and Innovation BFU2017-85048-R to MS-V and from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3) to MS-V and MM. MT-A was funded by an FPI predoctoral fellowship from the Spanish Ministry of Science and Innovation.

## ACKNOWLEDGMENTS

We thank Tony Donegan for language editing.

- neuronal dynamics. *PLoS Comput. Biol.* 5:e1000587. doi: 10.1371/journal.pcbi.1000587
- Diekelmann, S., and Born, J. (2010). The memory function of sleep. *Nat. Rev. Neurosci.* 11, 114–126. doi: 10.1038/nrn2762
- Gigante, G., Mattia, M., and Del Giudice, P. (2007). Diverse population-bursting modes of adapting spiking neurons. *Phys. Rev. Lett.* 98:148101. doi: 10.1103/PhysRevLett.98.148101
- Iber, C., Ancoli-Israel, S., Chesson, A. L., and Quan, S. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Westchester, IL: American Academy of Sleep Medicine.
- Jercog, D., Roxin, A., Barthó, P., Luczak, A., Compte, A., and De La Rocha, J. (2017). UP-DOWN cortical dynamics reflect state transitions in a bistable network. *eLife* 6:e22425. doi: 10.7554/eLife.22425
- Kätzel, D., Zemelman, B. V., Buetfering, C., Wölfel, M., and Miesenböck, G. (2011). The columnar and laminar organization of inhibitory connections to neocortical excitatory cells. *Nat. Neurosci.* 14, 100–107. doi: 10.1016/j.jhazmat.2021.125045
- Klinzing, J. G., Niethard, N., and Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nat. Neurosci.* 22, 1598–1610. doi: 10.1038/s41593-019-0467-3
- Levenstein, D., Buzsáki, G., and Rinzel, J. (2019). NREM sleep in the rodent neocortex and hippocampus reflects excitable dynamics. *Nat. Commun.* 10:2478. doi: 10.1038/s41467-019-10327-5
- Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell*. 163, 456–492. doi: 10.1016/j.cell.2015.09.029
- Mattia, M., Ferraina, S., and Del Giudice, P. (2010). Dissociated multi-unit activity and local field potentials: A theory inspired analysis of a motor decision task. *NeuroImage* 52, 812–823. doi: 10.1016/j.neuroimage.2010.01.063

- Mattia, M., and Del Giudice, P. (2000). Efficient event-driven simulation of large networks of spiking neurons and dynamical synapses. *Neural Comput.* 12, 2305–2329. doi: 10.1162/089976600300014953
- Mattia, M., and Sanchez-Vives, M. V. (2012). Exploring the spectrum of dynamical regimes and timescales in spontaneous cortical activity. *Cogn. Neurodyn.* 6, 239–250. doi: 10.1007/s11571-011-9179-4
- Mukovski, M., Chauvette, S., Timofeev, I., and Volgushev, M. (2007). Detection of active and silent states in neocortical neurons from the field potential signal during slow-wave sleep. *Cereb. Cortex* 17, 400–414. doi: 10.1093/cercor/bhj157
- Neske, G. T. (2016). The slow oscillation in cortical and thalamic networks: mechanisms and functions. *Front. Neural Circuits* 9:88. doi: 10.3389/fncir.2015.00088
- Nghiem, T.-A. E. A. E., Tort-Colet, N., Górski, T., Ferrari, U., Moghimi-firoozabad, S., Goldman, J. S., et al. (2020). Cholinergic switch between two types of slow waves in cerebral cortex. *Cereb. Cortex* 30, 3451–3466. doi: 10.1093/cercor/bhz320
- Paxinos, G., and Watson, C. (2007). *The Rat Brain in Stereotaxic Coordinates*. 6th Edn. New York, NY: Academic Press.
- Redinbaugh, M. J., Phillips, J. M., Kambi, N. A., Mohanta, S., Andryk, S., Dooley, G. L., et al. (2020). Thalamus modulates consciousness via layer-specific control of cortex. *Neuron* 106, 66–75. doi: 10.1016/j.neuron.2020.01.005
- Reig, R., Mattia, M., Compte, A., Belmonte, C., and Sanchez-Vives, M. V. (2010). Temperature modulation of slow and fast cortical rhythms. *J. Neurophysiol.* 103, 1253–1261. doi: 10.1152/jn.00890.2009
- Ruiz-Mejias, M., Ciria-Suarez, L., Mattia, M., and Sanchez-Vives, M. V. (2011). Slow and fast rhythms generated in the cerebral cortex of the anesthetized mouse. *J. Neurophysiol.* 106, 2910–2921. doi: 10.1152/jn.00440.2011
- Sanchez-Vives, M. V., Mattia, M., Compte, A., Perez-Zabalza, M., Winograd, M., Descalzo, V. F., et al. (2010). Inhibitory modulation of cortical up states. *J. Neurophysiol.* 104, 1314–1324. doi: 10.1152/jn.00178.2010
- Sanchez-Vives, M. V., Nowak, L. G., and McCormick, D. A. (2000). Membrane mechanisms underlying contrast adaptation in cat area 17 in vivo. *J. Neurosci.* 20, 4267–4285. doi: 10.1523/JNEUROSCI.20-11-04267.2000
- Sanchez-Vives, M. V. (2020). Origin and dynamics of cortical slow oscillations. *Curr. Opin. Physiol.* 15, 217–223. doi: 10.1016/j.cophys.2020.04.005
- Sanchez-Vives, M. V., and Mattia, M. (2014). Slow wave activity as the default mode of the cerebral cortex. *Arch. Ital. Biol.* 152, 147–155. doi: 10.12871/000298292014239
- Sanchez-Vives, M. V., Massimini, M., and Mattia, M. (2017). Shaping the default activity pattern of the cortical network. *Neuron* 94, 993–1001. doi: 10.1016/j.neuron.2017.05.015
- Sancristóbal, B., Rebollo, B., Boada, P., Sanchez-Vives, M. V., and Garcia-Ojalvo, J. (2016). Collective stochastic coherence in recurrent neuronal networks. *Nat. Phys.* 12, 881–887. doi: 10.1038/nphys3739
- Sarasso, S., D'Ambrosio, S., Fecchio, M., Casarotto, S., Viganò, A., Landi, C., et al. (2019). Local sleep-like cortical reactivity in the awake brain after focal injury. *Brain* 143, 3672–3684. doi: 10.1093/brain/awaa338
- Schnepel, P., Kumar, A., Zohar, M., Aertsen, A., and Boucsein, C. (2015). Physiology and impact of horizontal connections in rat neocortex. *Cereb. Cortex* 25, 3818–3835. doi: 10.1093/cercor/bhu265
- Steriade, M., Contreras, D., Curró Dossi, R., Nuñez, A., Houtkooper, R. H., Auwerx, J., et al. (1993). The slow (< 1 Hz) oscillation in reticular thalamic and thalamocortical neurons: scenario of sleep rhythm generation in interacting thalamic and neocortical networks. *J. Neurosci.* 13, 3284–3299. doi: 10.1523/JNEUROSCI.13-08-03284.1993
- Torao-Angosto, M., Manasanch, A., Mattia, M., and Sanchez-Vives, M. V. (2021). Data associated with “Up and Down states during slow oscillations in slow wave sleep and different levels of anesthesia” [Data set]. *Zenodo* doi: 10.5281/zenodo.4456700
- Tort-Colet, N., Capone, C., Sanchez-Vives, M. V., and Mattia, M. (2019). Attractor competition enriches cortical dynamics during awakening from anesthesia. *bioRxiv* [Preprint]. doi: 10.1101/517102
- Watson, B. O., Levenstein, D., Greene, J. P., Gelineas, J. N., and Buzsáki, G. (2016). Network homeostasis and state dynamics of neocortical sleep. *Neuron* 90, 839–852. doi: 10.1016/j.neuron.2016.03.036

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Torao-Angosto, Manasanch, Mattia and Sanchez-Vives. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Analyzing the Loss and the Recovery of Consciousness: Functional Connectivity Patterns and Changes in Heart Rate Variability During Propofol-Induced Anesthesia

**Davide Sattin<sup>1,2,\*†</sup>, Dunja Duran<sup>3†</sup>, Sergio Visintini<sup>4†</sup>, Elena Schiaffi<sup>5</sup>, Ferruccio Panzica<sup>6</sup>, Carla Carozzi<sup>7</sup>, Davide Rossi Sebastiano<sup>4</sup>, Elisa Visani<sup>3</sup>, Eleonora Tobaldini<sup>8,9</sup>, Angelica Carandina<sup>8,9</sup>, Valeria Citterio<sup>8,9</sup>, Francesca Giulia Magnani<sup>1</sup>, Martina Cacciatore<sup>1</sup>, Eleonora Orena<sup>7</sup>, Nicola Montano<sup>8,9</sup>, Dario Caldiroli<sup>7</sup>, Silvana Franceschetti<sup>5</sup>, Mario Picozzi<sup>10</sup> and Leonardi Matilde<sup>1\*</sup>**

## OPEN ACCESS

### Edited by:

Maurizio Mattia,  
National Institute of Health (ISS), Italy

### Reviewed by:

Axel Hutt,  
Inria Nancy - Grand-Est Research  
Centre, France  
Uncheol Lee,  
University of Michigan, United States

### \*Correspondence:

Leonardi Matilde  
matilde.leonardi@istituto-besta.it  
Davide Sattin  
davide.sattin@istituto-besta.it

<sup>†</sup>These authors have contributed  
equally to this work

<sup>‡</sup>Deceased

**Received:** 11 January 2021

**Accepted:** 15 March 2021

**Published:** 06 April 2021

### Citation:

Sattin D, Duran D, Visintini S, Schiaffi E, Panzica F, Carozzi C, Rossi Sebastiano D, Visani E, Tobaldini E, Carandina A, Citterio V, Magnani FG, Cacciatore M, Orena E, Montano N, Caldiroli D, Franceschetti S, Picozzi M and Matilde L (2021) Analyzing the Loss and the Recovery of Consciousness: Functional Connectivity Patterns and Changes in Heart Rate Variability During Propofol-Induced Anesthesia. *Front. Syst. Neurosci.* 15:652080. doi: 10.3389/fnsys.2021.652080

<sup>1</sup>Neurology, Public Health, Disability Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, <sup>2</sup>Clinical and Experimental Medicine and Medical Humanities-PhD Program, Insubria University, Varese, Italy, <sup>3</sup>Clinical and Experimental Epileptology Division, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, <sup>4</sup>Department of Neurosurgery, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, <sup>5</sup>Neurophysiology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, <sup>6</sup>Clinical Engineering Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, <sup>7</sup>Department of Anaesthesia, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, <sup>8</sup>Department of Internal Medicine, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan, Italy, <sup>9</sup>Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy, <sup>10</sup>Center for Clinical Ethics, Biotechnology and Life Sciences Department, Insubria University, Varese, Italy

The analysis of the central and the autonomic nervous systems (CNS, ANS) activities during general anesthesia (GA) provides fundamental information for the study of neural processes that support alterations of the consciousness level. In the present pilot study, we analyzed EEG signals and the heart rate (HR) variability (HRV) in a sample of 11 patients undergoing spinal surgery to investigate their CNS and ANS activities during GA obtained with propofol administration. Data were analyzed during different stages of GA: baseline, the first period of anesthetic induction, the period before the loss of consciousness, the first period after propofol discontinuation, and the period before the recovery of consciousness (ROC). In EEG spectral analysis, we found a decrease in posterior alpha and beta power in all cortical areas observed, except the occipital ones, and an increase in delta power, mainly during the induction phase. In EEG connectivity analysis, we found a significant increase of local efficiency index in alpha and delta bands between baseline and loss of consciousness as well as between baseline and ROC in delta band only and a significant reduction of the characteristic path length in alpha band between the baseline and ROC. Moreover, connectivity results showed that in the alpha band there was mainly a progressive increase in the number and in the strength of incoming connections in the frontal region, while in the beta band the parietal region showed mainly a significant increase in the number and in the strength of outgoing connections values. The HRV analysis showed that the induction

of anesthesia with propofol was associated with a progressive decrease in complexity and a consequent increase in the regularity indexes and that the anesthetic procedure determined bradycardia which was accompanied by an increase in cardiac sympathetic modulation and a decrease in cardiac parasympathetic modulation during the induction. Overall, the results of this pilot study showed as propofol-induced anesthesia caused modifications on EEG signal, leading to a “rebalance” between long and short-range cortical connections, and had a direct effect on the cardiac system. Our data suggest interesting perspectives for the interactions between the central and autonomic nervous systems for the modulation of the consciousness level.

**Keywords:** consciousness, anesthesia, connectivity, heart rate variability, propofol

## INTRODUCTION

General anesthesia (GA) is a pharmacologically induced reversible state in which the alterations of the Central and the Autonomic Nervous Systems (CNS, ANS) functioning cause behavioral unresponsiveness and decreased arousal and awareness. The changes induced in consciousness by anesthesia are often used as a theoretical model to study consciousness itself and deepen some issues related to it (e.g., Disorders of Consciousness; Pinsker, 2012; Sanders et al., 2012; Bonhomme et al., 2019). Anesthesiologists perform their work to obtain safe and rapid Loss of Consciousness (LOC) and to realize a proper Recovery of Consciousness (ROC) after surgery to prevent both too deep sedations as well as accidental awareness during GA. However, what exactly happens in both the CNS and ANS during the induction of anesthesia and the recovery is still poorly understood. In the next two paragraphs, we rough out some evidence that emerged in studies using Propofol administration during GA, a sedative-hypnotic agent which involves a positive modulation of the inhibitory function of the neurotransmitter gamma-aminobutyric acid (GABA) through GABA-A receptors (Sebel and Lowdon, 1989), that defined the empirical background of the present article.

### EEG Relative Power, Global and Local Network Patterns During Anesthesia

Regarding the study of the changes in the brain networks, Electroencephalography (EEG) is one of the most used techniques to explore the activity of CNS due to its high temporal resolution and its feasible use in the GA environment. Recent articles focused on exploring the changes in the spectral analysis or in the global or local connectivity during GA showing very interesting data. Spectral analysis, and in particular the analysis of the Relative Power (RP) distribution in various frequency bands, is the most EEG variable studied in the GA environment. It is well-known that arousal changes due to anesthesia are related to significant EEG modifications, namely a decrease in posterior alpha and an increase in fronto-central beta power during the induction phase and an increase in frontal power predominance in alpha, theta, and delta frequencies in deep sedation. These dynamics invert during the ROC and all these effects seemed to be common for different sedative and hypnotic agents (Gugino et al., 2001). Again, Purdon et al. (2013) suggested that the

low-frequency phase moderates alpha amplitude to be maximal at low-frequency peaks during deep sedation, whereas maximal at low-frequency nadirs during the transitional phases of loss and recovery of consciousness (Purdon et al., 2013).

Recent studies on global and local network patterns in the brain during GA suggested that the parietal-frontal network, one of the most long-range brain network studied in the scientific literature about consciousness (Lee et al., 2013a,b; Ryu et al., 2017), seemed to reverse the phase relationship between the two regions during propofol administration. A decrease in spectral Granger causality (a probabilistic measure investigating causality between two variables in a time series), for example, from the frontal to parietal areas was observed in the low-frequency bands conversely associated with an increase from the parietal to frontal areas of the high-frequency bands (Kim et al., 2017). In line with this result, propofol administration seemed to cause a disruption of hub-structures in the parietal region, an area that seems to play a pivotal role in information integration and transmission in the brain (Lee et al., 2013a). Lee et al. (2010) stated that changes of feedback and feedforward functional connections in the frontoparietal network are associated with changes in states of human consciousness. Moreover, a recent study on functional networks highlighted the relationship between individual alpha rhythms and unresponsiveness during sedation further reporting as the phase-amplitude coupling between slow and alpha oscillations correlated with propofol concentrations in blood (Chennu et al., 2016).

### Autonomic Nervous System and Heart Rate Variability Changes During Anesthesia

On the other side, the study of ANS signals during general anesthesia is quite difficult, due to the complex systems that regulate sympathetic and parasympathetic balance as well as the limitations regarding the ANS activity measurement due to the extreme fluctuations of the system (Acharya et al., 2006). However, some techniques based on Heart Rate (HR), acquired with electrocardiography (ECG) and the tone of vascular smooth muscle showed good reliability in monitoring the ANS activity even in overstimulating environments such as surgery rooms. One of the commonly used techniques monitoring the ANS activity is the Heart Rate Variability (HRV) which is focused on

the HR fluctuation over time (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996).

The application of HRV for the study of the role of ANS for consciousness evaluation is quite recent (Riganello et al., 2019). Some studies showed an increase in High-Frequency component power indicating an increase in vagal control of HR before the loss of consciousness in the GA environment: propofol seemed to reduce the total power as well as Low-Frequency power with an increase in the proportional part of HF power (Storella et al., 1999; Tarvainen et al., 2012). However, the results on the role of ANS on the modifications of consciousness level during GA are still lacking.

## Open Questions and Aims of the Present Study

As pointed out, CNS and ANS could provide important information on both the loss and the recovery of consciousness, using the analysis of both brain networks as well as the balance between sympathetic and vagal systems at different stages of sedation, in the same patients. Indeed, the study of how neural systems generate consciousness implies a profound analysis of what happens in that systems during situations in which consciousness is altered by different agents.

However, despite the results reported above, there are many open questions and limitations regarding CNS and ANS analysis during GA. For example, in most studies, only the baseline and period after LOC are compared using EEG data without considering the changes within the induction or the recovery periods, so giving a partial analysis of the activities of the cortical circuits (Kreuzer et al., 2014; Schneider et al., 2014). Moreover, apart from the frontal and parietal areas, the other cortical areas are less explored, although the role of global and local brain areas connections as well as the characteristics of these connections are highlighted as really important by some recent theories on consciousness (Tononi and Edelman, 1998; McFadden, 2007; Prakash et al., 2008; Yu and Blumenfeld, 2009). Moreover, other theories (Bosse et al., 2008; Damasio, 2012) underline the importance of the ANS as the pivotal systems fundamental for consciousness generation connecting information from the internal changes of the vegetative system to the cortical elaboration of stimuli and *vice versa*. HRV is a real chance to analyze how ANS information changed during LOC and ROC and its analysis during GA, although its analysis is often limited to pain perception (Logier et al., 2006; Sawaguchi et al., 2016; Garret-Bernardin et al., 2017) and not as an opportunity to study the balance between sympathetic and parasympathetic activity during the administration of anesthetic agents considering the changes within the induction or the recovery periods as for EEG analysis.

Therefore, the present pilot study aims to analyze CNS and ANS features through EEG and HRV analysis, in a sample of patients during propofol-induced unconsciousness for spinal surgical treatments. In detail, we investigated EEG spectral and connectivity properties in different bands and brain regions to explore changes during different phases of the

induction and recovery processes. In parallel, we performed a symbolic and complexity analysis of the HRV signals to detect non-reciprocal changes of sympathetic and parasympathetic modulation and to evaluate cardiac autonomic complexity. All the changes in EEG and HRV parameters during different phases of the induction and recovery processes were evaluated and results will be discussed in light of recent theories on consciousness.

## MATERIALS AND METHODS

### Participants

We enrolled 19 adult patients undergoing spinal surgery (for herniation or stenosis) in the neurosurgical unit. Exclusion criteria included age less than 18 years old, CNS disease, cardiovascular disease, or any disease that is known to affect the ANS, or patients taking drugs affecting the ANS, CNS, or cardiovascular system. Exclusion criteria also included a history of alcohol, drug abuse, or major psychiatric diseases. Data on eight patients were excluded from the analysis due to clinical problems during anesthesia that caused the changing of the anesthetic protocol ( $n = 5$ ) or low quality of EEG and HRV collected data during the induction or recovery phases ( $n = 3$ ).

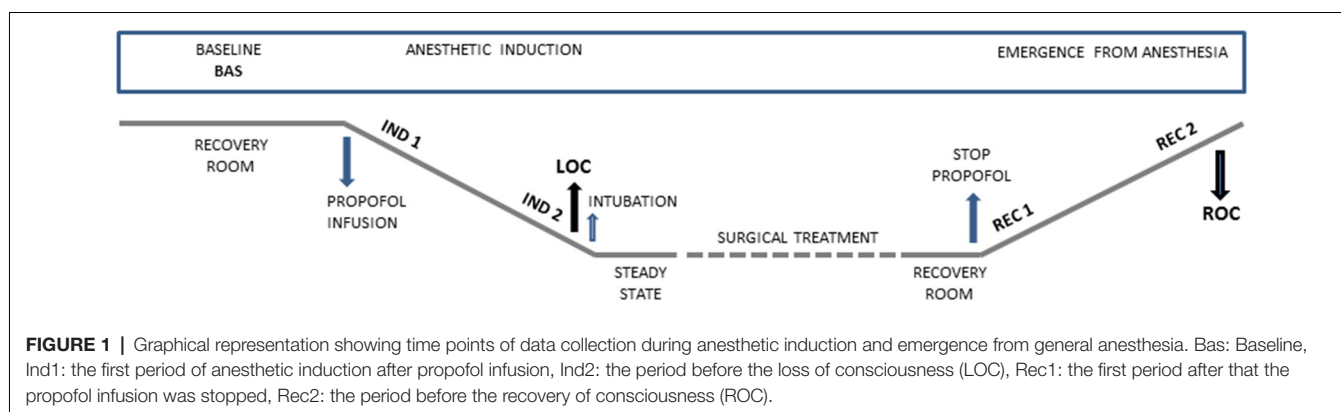
### Study Design and Procedure

All patients were enrolled in this prospective observational study during the pre-hospitalization visit (around 15 days before the surgery). During the preparation for the neurosurgical treatment, electrodes for EEG and ECG were applied. The data collection was made in five different conditions: at baseline (Bas), during the induction of anesthesia from the start of propofol infusion (Ind1) and immediately before the LOC (Ind2), and during the recovery phase immediately after the interruption of the propofol infusion (Rec1) and until the complete ROC (Rec2). LOC was defined according to the Observer Assessment of Alertness/Sedation Scale (OAA/S; Chernik et al., 1990). The timeline of the acquisition is reported in **Figure 1**.

The study was approved by the local Ethics Committee and written informed consent was obtained from all the participants.

### General Anesthesia Protocol and Monitoring

The following protocol is routinely applied by the anesthesiologist's staff to all patients who undergo surgery in our center. The study was conducted in a fully-equipped recovery room and in the operating theatre with three-lead ECG, pulse oximetry, capnography, and non-invasive blood pressure monitoring. An intravenous cannula was inserted in the forearm for fluids and intravenous drug administration. Intravenous GA was administered through a Bispectral Index (BIS)-guided anesthesia protocol where hypnotic/analgesic dosages were continuously adjusted according to the EEG depressive effect of drugs and to clinical levels of sedation. A BIS-xp electrode was placed on the subject's forehead, according to the manufacturer's recommendations, and was connected to the BIS Vista monitor (BIS Vista, Medtronic, MN, USA).



Each BIS index range, between 100 and 0, is clinically correlated to a sedation level as depicted by OAA/S. The interval 100–80 correlates to fully awake/minimally sedated, i.e., a patient able to open spontaneously the eyes, the 80–60 correlates to light/moderate sedation, i.e., a patient that opens his eyes after the command to open. The 60–40 correlates to LOC, i.e., the general anesthesia ideal plane where the patient does not open the eyes after mild prodding (level 1 of the OAA/S scale) and is associated with the appearance of a showed delta rhythm plus spindle-like waves with the absence of fast waves, features of general anesthesia (Bennett et al., 2009). Below 40 BIS number corresponds to over sedation.

Anesthesia was induced by mean of TCI delivery system (Injectomat® TIVA Agila, Fresenius Kabi, France) settled at Effect-site Concentration (Ce), i.e., at the Cerebral Concentration desired. According to the pharmacokinetic model for propofol (Schnider et al., 1999) and Remifentanyl (Minto et al., 1997), the software pump calculates the infusion rate of drugs to achieve and maintain the chosen Ce at any time, avoiding overshoot or under dosage during dose adjustments and drug accumulation.

Starting from a propofol Ce of 2 µg/ml, the Ce was increased by 1 µg/ml after 1 min of equilibration at the reached Ce, until the appearance of LOC. The Remifentanyl infusion started after the LOC to prevent pain for intubation, while at ROC the Ce Remi was recorded (range Ce 1–1.5 µg/ml). Then curare (cisatracurium) was administered, and the patient was intubated (0.1–0.2 mg/kg of Cisatracurium is the usual dosage for intubation with the routine use of Glidescope at our Institution). In the present study, the cisatracurium was administered only once after LOC for intubation purposes. Since the Recovery time of Curarization is around 50–60 min for Cisatracurium, any effect could be largely negligible at ROC. All patients were extubated without reversal drugs of curarization (mean anesthesia duration exceeded 60 min). The surgery phase of anesthesia was managed for maintaining stable standard clinical parameters (BIS 40–60, blood pressure, heart rate, and absence of body movement). At the end of the surgical procedure, the patient was then transferred to the recovery room for the awakening, and propofol infusion was stopped, while Remifentanyl was reduced at Ce 1 ng/ml for additional analgesia and tolerance to the orotracheal tube. The propofol and Remifentanyl decay were continuously monitored on the

pumps. After reaching the ROC, the patients were extubated, and Remifentanyl stopped.

## Signal Acquisition

### EEG

EEG was recorded using 19 Ag/AgCl surface electrodes (impedance <5 kΩ) placed following the 10–20 International System and acquired at a sampling rate of 256 Hz (Micromed SpA, Mogliano Veneto, Treviso, Italy). In four patients, the signal acquired on Fz had a very low signal-to-noise ratio thus only 18 common channels were included in the analysis. Data were average re-referenced and band-pass filtered in 0.5–48 Hz band. The analysis was performed on 1-min artifact-free EEG representative epochs acquired in the five conditions (Bas, Ind1, Ind2, Rec1, and Rec2; **Figure 1**).

### ECG

To evaluate cardiac autonomic control during the different stages of anesthesia, the ECG signals were recorded (SystemPLUS, Micromed S.p.A., Treviso, IT, USA) in the different conditions following the scheme reported in **Figure 1**. ECG recordings were processed through a specific software (HeartScope II; AMPS-LLC, New York, NY, USA). The R-R intervals between adjacent R peaks were identified to provide the tachogram. Ectopic beats were deleted from the resulting R-R series and replaced by a virtual beat placed by interpolating adjacent R peaks as recommended (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996). Finally, segments of  $250 \pm 50$  consecutive beats were selected for each period. We applied two different methods for the analysis of HRV on the selected segments, symbolic analysis and complexity analysis as described in the next sections.

## EEG Analysis

### Spectral Power Analysis

The EEG epochs were filtered (1–40 Hz, 12 db/octave) and, for each condition, divided into 30 non-overlapping 2-s segments and analyzed using the fast Fourier transform (FFT). Relative power (RP) was calculated in the delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta bands (13–30 Hz), and averaged within each EEG channel.



## Connectivity Analysis

The connectivity pattern was studied using Partial Directed Coherence (PDC), a frequency-domain measure derived from the multivariate autoregressive (MVAR) modeling of multichannel EEG signals (Baccalá and Sameshima, 2001). The characteristic of the PDC is to capture the direction of the influence and to be less affected by the volume conduction problem than other methods.

For each condition, EEG data were epoched in non-overlapped short segments of 3 s which were de-trended and normalized by subtracting the mean value and dividing the result by the variance. The length of the epoch was chosen to balance the stationarity of the data and the model fit (better parameter estimation for longer segments). All of the EEG channels were simultaneously used as inputs for the MVAR model. The optimal order was chosen after calculating for each epoch the Bayesian information criterion. All segments showed the optimal model order ranging from 25 to 30, so order 26 was chosen and applied to all the epochs. Once the MVAR coefficients had been adequately estimated, the PDC spectra were estimated. The statistical significance of non-zero PDC values, at each frequency, was assessed by a bootstrap approach using phase randomization. Only significant connections were considered and averaged over segments. PDC values were calculated in the same bands considered for the spectral power analysis. To evaluate local connectivity patterns, in addition to the mean global PDC value, regional PDC values were calculated by averaging values into five regions of interest (ROIs): frontal (F3, F4, Fp1, and Fp2), central (C3, C4, and Cz), parietal (P3, P4, Pz), occipital (O1 and O2) and temporal (T4, T6, F8, T5, T3, F7).

Graph theory was applied to describe connectivity results by calculating indices describing the topological properties of the networks. To compare connectivity indices, a fixed number of connections was set for each band as the minimum level of connections to have a connected network in all subjects.

Global efficiency, local efficiency, the characteristic path length, and the clustering coefficient measured the level of integration and segregation of the whole network. Moreover, the number of connections (degree) and the strength of connections as well as the directional values in/out-degree, in/out-strength (number and strength of incoming and outgoing connections), and in/out-degree fraction were calculated by averaging values of the electrodes within ROIs.

The signals were pre-processed and analyzed using a custom-written toolbox in MATLAB (MathWorks Inc., Natick, MA, USA) using open-source toolboxes (Fieldtrip, BCT, eMVAR). A detailed description of the indexes used was reported in the **Supplementary Materials**.

## ECG Analysis

### Symbolic Analysis

Cardiac autonomic control was evaluated through a non-linear approach to detect non-reciprocal changes of sympathetic and parasympathetic modulation (Porta et al., 2001, 2007a,b; Guzzetti et al., 2005). After the transformation of the tachogram into a sequence of symbols (i.e., words), four patterns were identified and quantified as to their rate of occurrence considering

sequences of three consecutive symbols: 0 V%, when a sequence of three symbols shows no variations; 1 V%, pattern with only one variation; 2 LV%, when the symbols show two like variations (ascending or descending slope); 2 UV%, the pattern with two unlinked variations. Experimental and clinical studies involving pharmacological blockade and autonomic tests reveal that 0 V% index is a marker of cardiac sympathetic modulation while 2 LV% and 2 UV% indexes are markers related to cardiac vagal modulation (Porta et al., 2001, 2007a,b; Guzzetti et al., 2005).

## Complexity Analysis

The heart rate variability is the result of the activity of several mechanisms acting over similar but not coincident frequencies (e.g., vagal and sympathetic modulation, baroreflex and chemoreflex regulation, respiratory sinus arrhythmia, hormonal influences, local factors). All these mechanisms contribute to the complexity of the signal and non-linear tools, such as entropy-derived measures, provide information on the complexity of the autonomic cardiac control in both physiological and pathophysiological conditions (Porta et al., 2001). In this study, we applied Corrected Conditional Entropy (CCE) to evaluate cardiac autonomic complexity. CCE is derived from Conditional Entropy that assesses the amount of information carried by the current RR sample. CCE is bounded between zero when the new sample is fully predictable, and Shannon entropy, when the new sample is unpredictable. By dividing CCE by the Shannon entropy, it is possible to derive an index of regularity (Ro) which is bounded between 1, maximum regularity and lowest complexity, to 0, lowest regularity, and maximum complexity (Porta et al., 2001).

## Statistical Analysis

Statistical analyses were performed using SPSS (version 17, SPSS Inc. Chicago, IL, USA) and SigmaStat software (2016 Systat Software, Inc., Chicago, IL, USA). Repeated measures analysis of variance (RM-ANOVA) was applied to assess the effects of the anesthesia conditions on RP, PDC, connectivity values, in each band separately, and on cardiovascular autonomic measures. The sphericity assumption was evaluated using Mauchly's test and the Greenhouse-Geisser correction was applied when appropriate. The Shapiro-Wilk test was used to assess the normality of data distribution. When a non-normal distribution was found, differences were examined by the Friedman Test for repeated measures analysis of variance on ranks. *Post-hoc* analyses were performed using paired *t*-test (Bonferroni corrected). Statistical significance was always set at  $p < 0.05$  for all the analyses.

## RESULTS

The characteristics of the patients are shown in **Table 1**. No intraoperative awareness or recall was reported by any patient in the study.

### EEG Relative Power

In **Figure 2A** was reported the EEG RP distribution for delta, theta, alpha, and beta bands in each condition. During the induction and recovery period, a decrease in occipital alpha RP was observed, whereas an increase in delta RP was noted

**TABLE 1** | Demographic and clinical data.

Patient id	Gender	Age (years)	Weight (Kg)	Height (cm)	Diagnosis
1	F	66	57	160	Spinal disc herniation L4, L5
2	F	53	57	160	Spinal disc herniation L4, L5
3	M	49	80	175	Spinal disc herniation L5, S1
4	M	67	70	168	Spinal disc herniation L3, L4
5	F	56	59	160	Spinal disc herniation L5, S1
6	M	62	71	169	Spinal stenosis L3, L4
7	M	79	95	180	Spinal stenosis L2, L3
8	F	74	93	175	Spinal stenosis L3, L4
9	F	45	49	160	Spinal stenosis L5, S1
10	M	30	78	170	Spinal disc herniation L4, L5
11	M	76	91	169	Spinal stenosis L3-L5

in all regions. Central beta RP increased during the beginning period of the induction and then decreased in the other intervals. The statistical comparison of RP in different periods among frequency bands and ROIs was reported in **Figure 2B**.

RM-ANOVA for ROIs and intervals indicated significant changes in delta ( $p < 0.001$ ), alpha ( $p < 0.001$ ), and beta ( $p = 0.009$ ) bands. *Post-hoc* investigation of delta band in all regions showed that relative power in delta band increased in all ROIs during the induction intervals, with significant differences between Bas and Ind2 (pre-LOC) in all regions. The occipital region only showed a significant difference between baseline and light sedation (Ind1) in the delta ( $p < 0.001$ ), whereas differences between the light (Ind1) and deep sedation (Ind2) were seen in temporal ( $p = 0.027$ ) and parietal ( $p = 0.037$ ) ROIs. During the recovery step, it is worth noting that this increase in delta activity did not recover to the initial values observed during the baseline step as high relative power persisted in the pre-ROC (Rec2) interval compared to pre-anesthesia state for central ( $p = 0.004$ ), parietal ( $p = 0.005$ ), and occipital ( $p = 0.001$ ) ROIs.

Alpha and beta band revealed, on the other side, a significant reduction in relative power during the induction intervals. In the alpha band this pattern was observed mainly in posterior ROIs during the induction phase (Bas vs. Ind2; parietal,  $p = 0.008$  and occipital,  $p = 0.004$ ) and persisted with a significant difference in the same areas between the Bas and Rec2 period ( $p < 0.05$  for both areas). In the beta band, this decrease during the induction phase was evident especially in Ind1 and Ind2 for almost all ROIs, except the occipital region.

## Graph Global Properties

Significant results were found mainly in the lower frequency bands (**Figure 3**). Global efficiency changed significantly in the delta ( $p = 0.007$ ), theta ( $p < 0.001$ ), and alpha band ( $p = 0.013$ ). *Post-hoc* pairwise comparison showed an increasing trend during the sedation compared to the baseline, not recovering to the initial baseline value in the awakening ( $p = 0.03$ ) for the theta band.

Similarly, significant changes were observed also in Local Efficiency in delta ( $p < 0.001$ ) theta ( $p < 0.001$ ), and alpha band ( $p < 0.01$ ). *Post-hoc* pairwise comparisons showed a significant increase from Bas to Ind2 interval in delta ( $p = 0.04$ ) and alpha band ( $p = 0.03$ ) indicating a decrease in the distance among neighboring nodes and so an increased efficacy in the

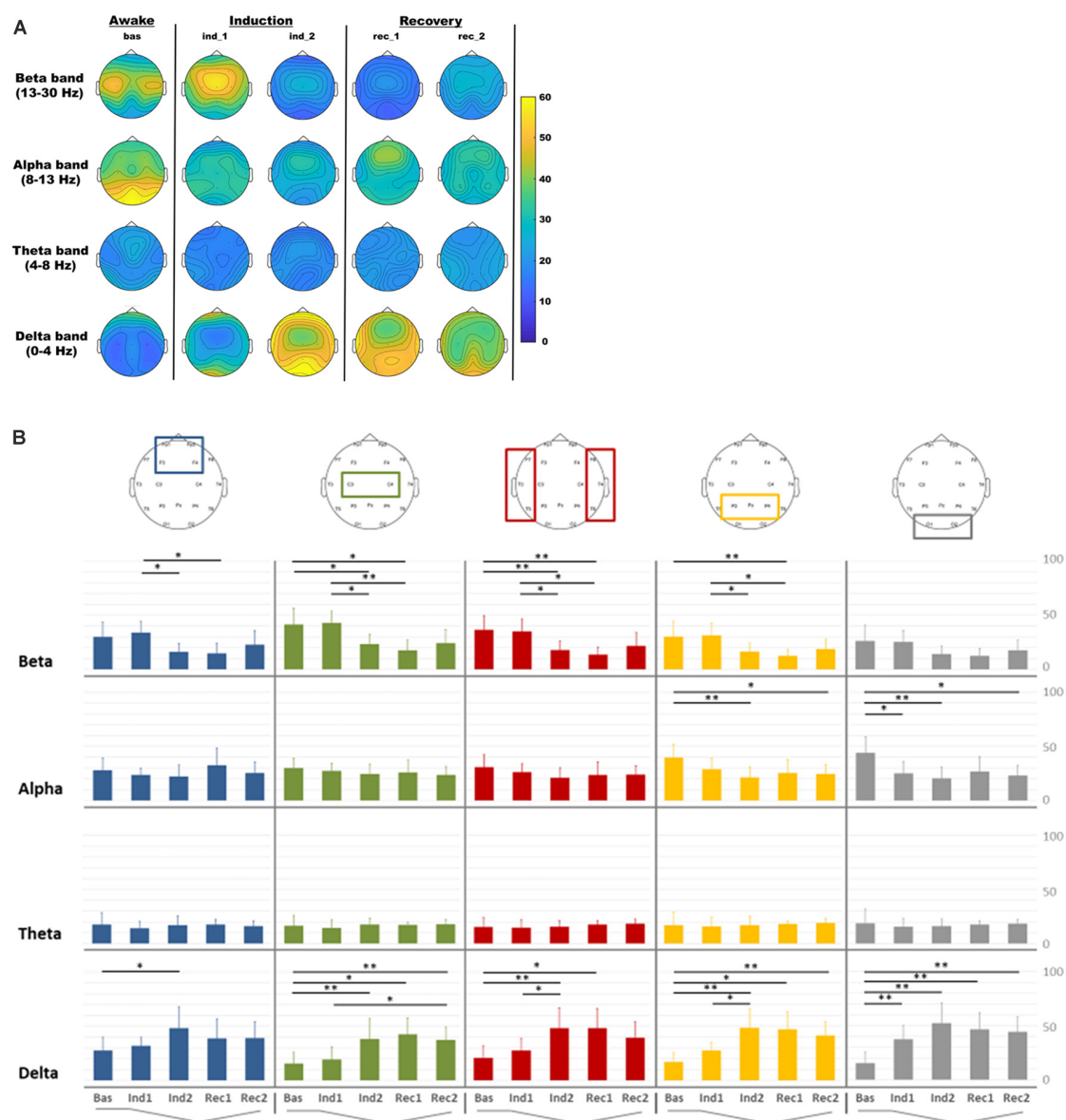
information transfer within regions during the induction phase. During the recovery, the LE did not return to the baseline value in the delta band ( $p = 0.005$ ) showing an increasing trend.

Characteristic path length confirmed an increase of network connections in lower bands, showing decreasing trend values from Bas to Ind2 in the alpha band, maintaining the lower values during all recovery from anesthesia, and not returning to the initial values in the pre-ROC interval (Bas vs. Rec2,  $p = 0.03$ ). Cluster Coefficient did not show any significant statistical differences comparing different stages and bands.

## Regional Graph Properties

RM-ANOVA showed significant results for most of the considered network indexes (in/out/total-degree and in/out/total-strength) in theta, alpha, and beta bands. In the beta band, post-hoc ANOVA for the effect of the interval demonstrated that, in particular, parietal ROI was involved in the change, covering all the period of sedation and recovery ( $p < 0.001$ ); indeed, the pairwise comparison showed a significant increase of total-degree (Bas-Rec2:  $p = 0.03$ ). This increment was confirmed observing also in the parietal out-degree index ( $p < 0.001$ ) during induction, with *post-hoc* significant changes found between Bas-Ind2 ( $p = 0.02$ ), Bas-Rec1 ( $p = 0.04$ ), and Ind1-Rec1 ( $p = 0.02$ ). In/out-Degree fraction ( $p < 0.001$ ) highlighted that, during emergence from anesthesia, the network properties of the awake state are not recovered (Bas- Rec1,  $p < 0.001$  and Bas-Rec2,  $p = 0.004$ ). On the other side, in the frontal region, a decreasing number of connections between the light sedation and the awakening (Ind1-Rec2,  $p = 0.041$ ) was observed.

In the alpha band, frontal and parietal ROIs displayed an opposite pattern, according to the in-degree value. Frontal regions increased the number and strength of the incoming connections during induction, with a highly significant change of in-strength index (Bas-Ind1:  $p < 0.001$ ; Bas-Ind2:  $p = 0.037$ ; Bas-Rec1:  $p < 0.001$ ; Ind1-Rec1:  $p = 0.01$ ), and of in-degree values (Bas-Ind1:  $p = 0.01$ ; Bas-Rec1:  $p = 0.003$ ). On the other side, parietal ROI revealed a decrease of in-degree from Bas to Rec1 interval ( $p = 0.04$ ). Finally, in the theta band, a decrease in the number of outgoing connections ( $p = 0.03$ ) and an increase in the strength of the incoming connections ( $p = 0.02$ ) between Ind1 and Rec1 was observed in frontal ROI.



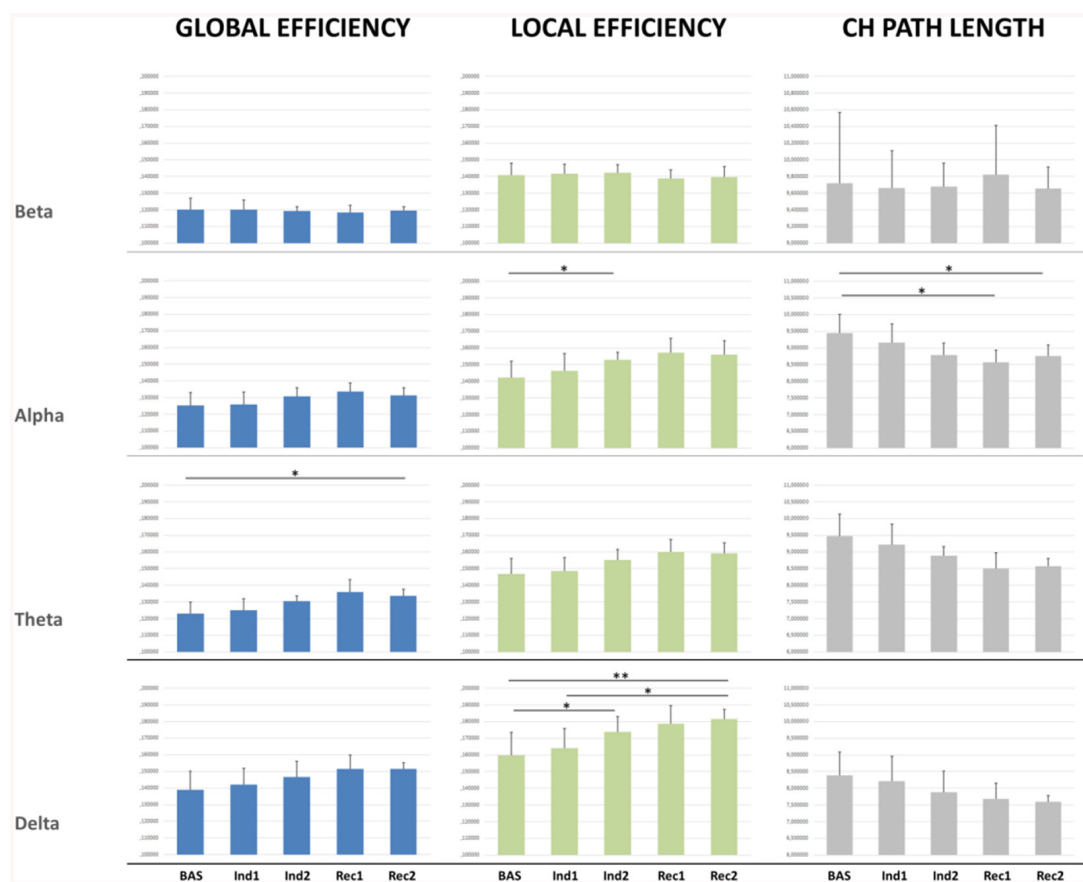
**FIGURE 2 |** Electroencephalography (EEG) relative power distribution for delta, theta, alpha, and beta bands in each condition. **(A)** EEG relative power maps. **(B)** EEG relative spectral power comparisons among regions of interest (ROIs) for each band analyzed in each condition: baseline (Bas), beginning phase of administration of propofol (Ind1), loss of consciousness (Ind2), propofol infusion stopped (Rec1) and pre-recovery of consciousness (Rec2) period for frontal (blue), central (green), temporal (red), parietal (orange) and occipital (gray) ROIs. \* $p < 0.05$ ; \*\* $p < 0.01$ .

In temporal ROI, in-strength increased (Bas-Ind2:  $p = 0.009$ ), while in central ROI, the number of incoming connections decreased (Ind1-Rec1,  $p = 0.01$ ). All ANOVAs and comparisons that resulted in statistically significant values were reported in **Figure 4**.

## ECG Analysis

Regarding the analysis of the ECG data, we observed a progressive decrease in the mean Heart Rate starting from the second induction period with significant bradycardia during the two phases of recovery (**Figure 5A**;  $p < 0.001$ ). Concerning

the temporal course of cardiovascular autonomic control, the symbolic analysis revealed an increase of pattern with no variation (0 V%; **Figure 5B**), a marker of cardiac sympathetic modulation, in both induction phases reaching the significance in Ind2 ( $p = 0.045$ ), while the first recovery period was characterized by values of 0 V% significantly lower than Ind2 ( $p = 0.002$ ). Conversely, the parasympathetic indices (2 LV% and 2 UV%) presented an opposite trend with a progressive reduction during the induction phases, followed by an increase in the first recovery period and a final return to baseline values (**Figures 5C,D**). Specifically, the frequency of 2 LV% in Ind2 was



**FIGURE 3** | Connectivity indexes measured for the whole brain network for each EEG band in each condition.\* $p < 0.05$ ; \*\* $p < 0.01$ .

significantly lower compared to 2 LV% Bas values ( $p < 0.05$ ) and 2 LV% values recorded in Rec1. Similarly, 2 UV% in Ind2 differed significantly from those obtained in the Rec1 period ( $p < 0.05$ ).

As to complexity measures, the index of regularity (Ro) showed an increase during the last period of induction, dropped below the baseline value in the first recovery period, and, at last, we observed a return to baseline values (**Figure 5E**). Complementarily, CCE was decreased in the second phase of induction and overshoot the baseline values during Rec1 (**Figure 5F**). We found a significant difference between Ind2 and Rec1 for both Ro and CCE mean values ( $p < 0.001$  and  $p = 0.006$ , respectively).

As to the temporal differences inside each condition, despite a lower mean Heart Rate, Ind2 was characterized by a higher sympathetic and lower parasympathetic modulation ( $p = 0.012$  and  $p < 0.001$ , respectively; **Figures 5B,C**) compared to Ind1. The same trend was present in the two recovery phases: Rec2 showed higher 0 V% values (**Figure 5B**,  $p = 0.013$ ) and lower 2 UV% values (**Figure 5D**,  $p = 0.004$ ) compared to Rec1. Finally, both the second phases of induction and recovery period showed greater regularity (**Figure 5E**,  $p = 0.007$  and  $p < 0.001$ ) and less

complexity (**Figure 5F**,  $p = 0.021$  and  $p = 0.026$ ) than the first phases.

## DISCUSSION

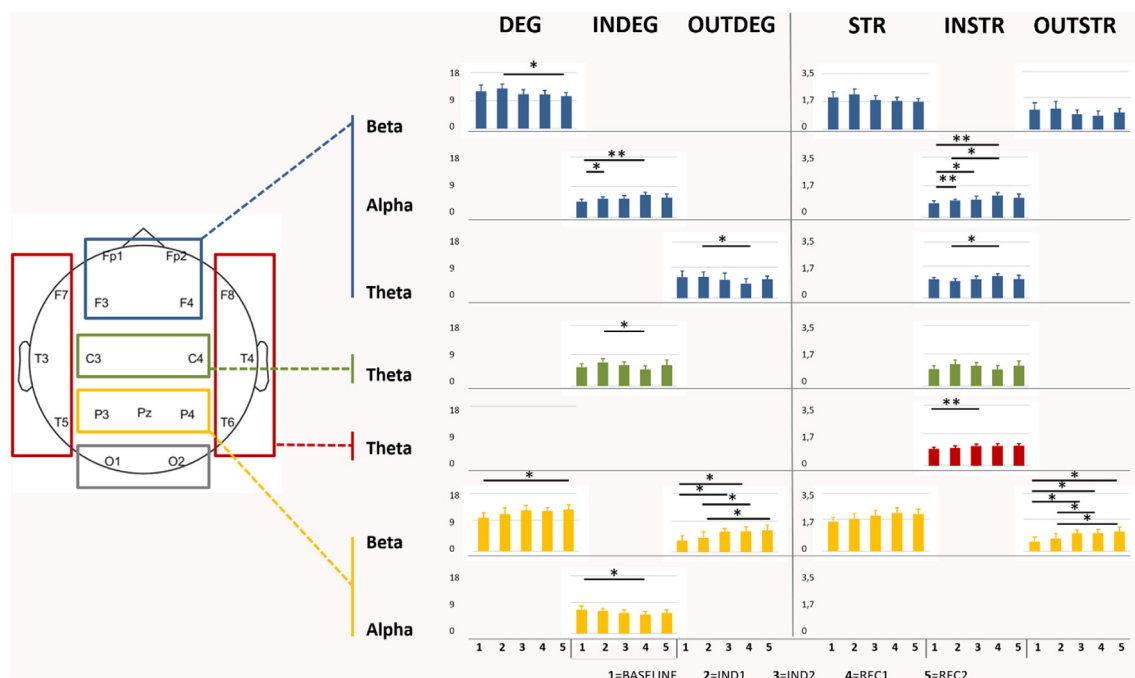
In this work, we studied the changes of EEG and HRV features during induction and recovery from general anesthesia as a theoretical model for the understanding of consciousness.

In summary, the main results are:

In the *EEG spectral analysis*: (i) a decrease in posterior alpha and beta power in all areas except for the occipital ones and an increase in delta power mainly during the induction phase; (ii) no statistically significant variation in the theta band across periods and areas.

In the *EEG connectivity analysis*: (iii) a significant increase of local efficiency index in alpha and delta bands between baseline and LOC as well as between baseline and ROC in delta band only, whereas the global efficiency index did not reveal any statistical differences except between baseline and ROC in the theta band; (iv) a significant reduction of the characteristic path length in alpha band between the baseline and ROC. Moreover, our data showed that; (v) in the alpha band there was mainly a progressive increase in the number and the strength





**FIGURE 4 |** Graphical representation of the significant differences of degree (DEG), Strength (STR), and the directional in/out-degree (INDEG/OUTDEG) and strength (INSTR/OUTSTR) values in each ROIs for each anesthesia period and EEG bands. Note: frontal area showed an increase in the in indexes in the alpha bands during the induction revealing different statistical differences between some periods whereas the statistically significant comparisons in the out indexes were highlighted in the beta band in the parietal area. \* $p < 0.05$ ; \*\* $p < 0.01$ .

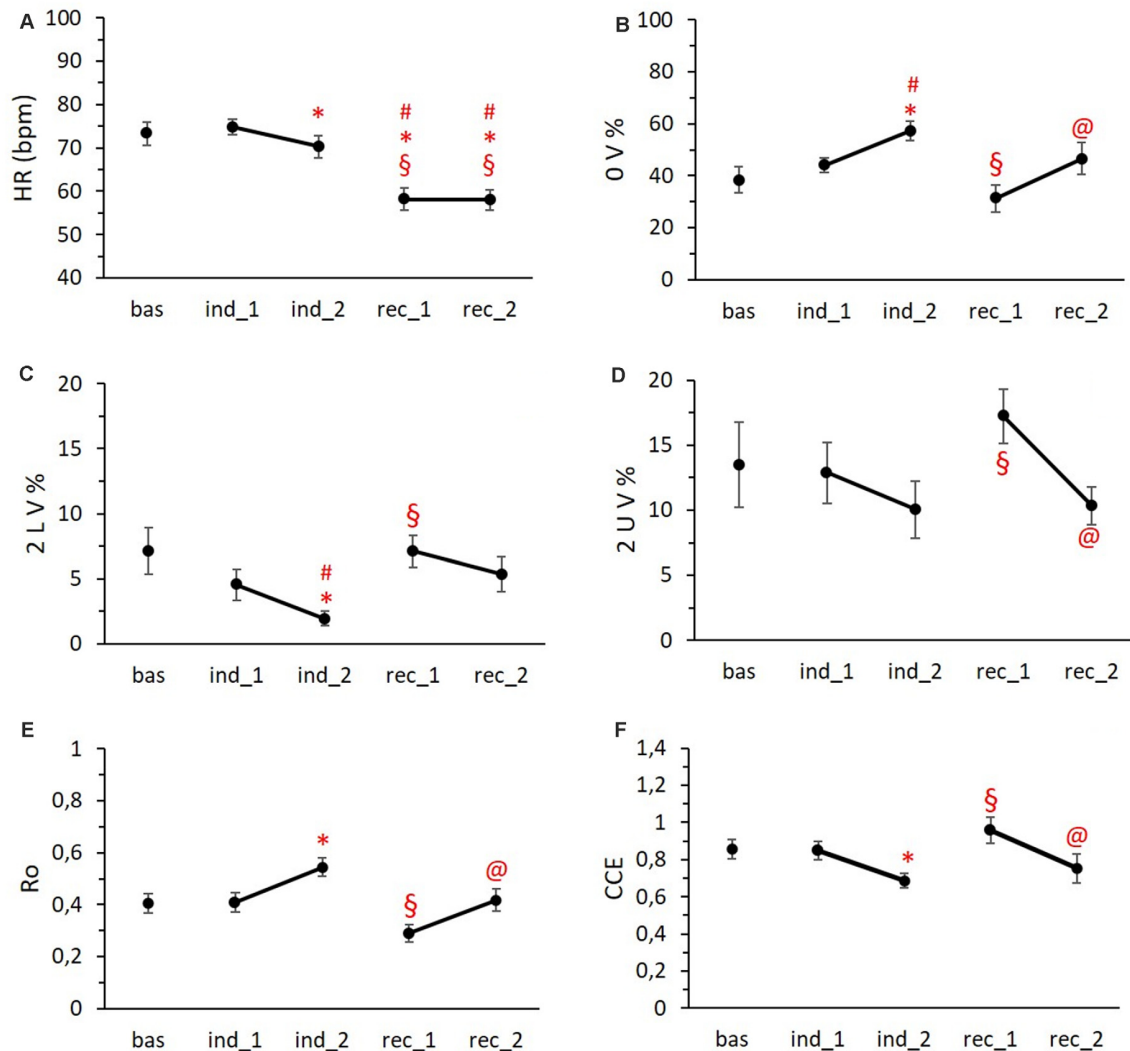
of incoming connections in the frontal ROI, while in the beta band the parietal ROI showed mainly a significant increase in the number and the strength of outgoing connections values with a persistent trend in the recovery period; and (vi) the theta band pattern resulted statistically different in different areas, with a statistically significant increase of the strength of incoming connections during the induction phase in the temporal area.

The *HRV analysis* showed that: (vii) the induction of anesthesia with propofol was associated with a progressive decrease in complexity (CCE) and a consequent increase in the regularity index Ro of the RR series; (viii) the anesthetic procedure determined bradycardia which was accompanied by an increase in cardiac sympathetic modulation (0 V%) and a decrease in cardiac parasympathetic modulation (2 LV%) during the induction; (ix) all indexes returned to basal values during the recovery period except for heart rate in contrast to the most analyzed EEG indexes. In the next paragraphs, all these results are thoroughly discussed.

## EEG Spectral Analysis

It is well-known that arousal changes following anesthesia with propofol are related to significant EEG modifications, although a direct compound among studies is difficult due to the different EEG epochs analyzed and the external variables influencing electrical activities present in a complex environment like recovery/surgical rooms.

The decrease in posterior alpha power is found also by other studies confirming that our result is in line with other researches (Gugino et al., 2001; Yeom et al., 2017). In contrast, beta power tends to increase initially, after administration of propofol, and then decrease significantly during the induction phase showing an effect that is debated in the literature (Gugino et al., 2001; John et al., 2001), and that seemed to be correlated to the direct action of propofol. Regarding the increase in slow (i.e., theta and delta) and very low (under 1 Hz) frequencies associated with the appearance of frontal alpha frequencies in deep sedation (CIT), our data confirmed these results although the very low frequencies were not analyzed in our study. These dynamics reverse during the recovery, although with an asymmetric timeframe as already described in the literature (Gugino et al., 2001), confirming that changes in the EEG bands could follow a different trend in the recovery period with respect to the induction one probably due to pharmacokinetic of anesthetic drugs. Indeed, during ROC Opioids and Remifentanil have a variable effect on the EEG because they predominantly affect no cortical structures. Larger doses can produce EEG slowing (until delta waves). A study shows a significant linear correlation between increasing remifentanil concentration (0.5–10 ng/ml) and decreasing of BIS (Koitabashi et al., 2002) and another study found that the effect of remifentanil on the spectrum and quantitative parameters of the electroencephalogram is significant and strongly dependent on the level of anesthesia: (1) decrease of activity in high frequencies (>14 Hz) during



**FIGURE 5 |** HR trend (A), symbolic metrics trends (B,C,D), and complexity metrics trends (E,F) during the different phases of the protocol. Data are reported as means  $\pm$  standard errors. #vs. Bas, \*vs. Ind1 §vs. Ind1, @vs. Rec1. Abbreviations: HR = heart rate; 0 V % = pattern with no variations; 2 LV % = pattern with two like variations; 2 UV % = pattern with two unlike variations; Ro = index of regularity; CCE = corrected conditional entropy.

light anesthesia; and (2) increase of activity in the extended alpha band (7–14 Hz) and decrease of activity in the delta band (0.5–4 Hz) during deep anesthesia (Bennett et al., 2009). Both studies cannot evaluate the effect of remifentanyl alone, since conducted during general anesthesia with propofol and interaction (Propofol /Remi) cannot be excluded. We think that minimal Remifentanyl dosages (as at ROC in our study) should not affect the EEG signals but this is a limitation of our study that we reported in the limit section.

John et al. (2001) demonstrated a significant change in hemispheric interactions during the loss of consciousness and sedation, with a “close coupling” of the anterior regions, paralleled by a clear anteriorization of power spectra. The same authors described the “uncoupling” of other regions (anterior and posterior regions on each hemisphere, as well as homologous regions between the two hemispheres). Finally, Soplata et al.

(2017) described the transition between consciousness and unconsciousness, highlighting the spectral properties of the EEG, including the alpha background activity (8–12 Hz), the slow-wave oscillation (0.1–1.5 Hz), and the dose-dependent phase-amplitude between them. In this model, the thalamus could have a pivotal role in propofol-induced anesthesia, generating a novel hyperpolarizing very-low-frequency rhythm probably due to a GABAergic effect on the cortex. Our results replicate these previous works, even with slight differences: We did not compare the dynamics of the two hemispheres because we grouped both left and right frontal, central, temporal, parietal, and occipital channels in the same ROIs. Moreover, our result is in line with previous works (Gugino et al., 2001; Yeom et al., 2017) as functional changes in brain dynamics are dissimilar during forward and reverse transition from one conscious state to the other. This complex phenomenon is

known as hysteresis, it has been described for both the spectral and the connectivity measures (Kim et al., 2018), and it could be accounted to the pharmacokinetic properties of anesthetic drugs (Kuizenga et al., 2018), although some aspects remain poorly understood (Bonhomme et al., 2019). Also, patients who underwent anesthesia showed a huge heterogeneity (greater than voluntary subjects) while recovering consciousness after drug discontinuation (Chander et al., 2014), potentially increasing the hysteretic effect.

## EEG Connectivity, Global and Local Indexes

Boussen et al. (2018) studied the effect of the re-emergence of consciousness from anesthesia directly on the cortex (through electrocorticography), demonstrating that the emergence states correspond to a simple signal transition in the frequency space, which is possible to be reproduced by a time rescaling of cortical activity during the unconscious state. The authors enhanced the role of the intracortical interactions in the ROC, limiting the thalamus in controlling or modulating the inhibition. Moreover, Lee and colleagues demonstrated that different anesthetics, including propofol, have the main effect to disrupt frontal-parietal networks (Lee et al., 2010, 2013a,b); at the same time, brain networks showed a significant reduction in the number of network connections during anesthesia, although maintaining a scale-free organization across consciousness, anesthesia, and recovery (Lee et al., 2010). On the contrary, a recent study (Vlissides et al., 2019) stated that the alpha band frontal-parietal connectivity was not able to distinguish levels of consciousness or anesthetic states in surgical patients as a single measure, implicitly reducing the importance of the parieto-frontal connections as a “marker of consciousness.”

Since a single measure of functional connectivity is probably not a reliable correlate of the effect induced by the anesthetic (Vlissides et al., 2019), in our work we considered network measures and connectivity indices for each ROI in the classical EEG band. Global efficiency, an index related to the efficiency of communication among all nodes in a network, showed, in the theta band, a significant difference between baseline and pre-ROC. Theta connectivity showed a complex pattern characterized by a decrease of the outgoing connections and an increase of the strength of the incoming connections from induction to initial recovery in the frontal region, with a concurrent decrease in incoming connections in the central region, while an increase of the strength of the incoming connections was observed in temporal ROI. This result appears noticeable because the theta band is the only one among the EEG bands that did not show significant differences in the spectral analysis, revealing that the connectivity and spectral analyses can provide complementary information. In the delta band, local efficiency showed a significant increase between baseline and induction, between baseline and beginning of the recovery, and between the first period of sedation and initial recovery. These results are in line with previous literature reporting that the appearance of theta and the decrease of delta activity characterize the transition from deep to light sedation (Sheeba et al., 2008). Furthermore, our data showed that in the alpha band there

was an increase in the number and the strength of incoming connections in the frontal ROI from baseline to induction and initial recovery with a concurrent decrease of the number of incoming connections in the parietal ROI. In the same band, local efficiency showed a significant increase between baseline and deeper sedation state and a significant reduction of the characteristic path length in the alpha band between the baseline and ROC.

Taken together, our results highlighted that: (a) brain networks exhibit a complex, non-linear (i.e., the recovery process is a different and slower period than the loss of consciousness process) pattern of modifications during anesthesia; (b) these modifications affect several frequency bands of the EEG signal; (c) leading to a “rebalance” between long and short-range cortical connections, and finally, that (d) different ROIs, namely frontal and parietal ones, show opposite behavior. This endorses the hypothesis of disrupted “long-range” networking from frontal to temporal and parietal areas (Supp et al., 2011) with the resulting prevalence of short-range connections during the loss of consciousness and recovery during the emergence period, albeit in a different range of time, with a longer ROC period.

## HRV Analysis

The analysis of the ANS showed that: (i) the induction of anesthesia with propofol was associated with a decrease in complexity, with a consequent increase in the regularity index, and in cardiac parasympathetic modulation; (ii) the anesthetic procedure determined bradycardia which was accompanied by an increase in cardiac sympathetic modulation during the induction; and (iii) the return to basal values was rather complete of all indexes during the recovery period except for heart rate.

General anesthesia determines a global depression of CNS functions, as well as the cardiovascular autonomic functions. As it has been previously reported in the literature, the nature and characteristics of the cardiovascular autonomic control alterations directly depend on the type and dose of anesthetic drugs (Ebert, 2005; Riznyk et al., 2005; Mäenpää et al., 2011). Despite different effects of the anesthetics on the two branches of ANS, to the best of our knowledge, the majority of studies confirms a general decrease in HRV during the induction phases due to the progressive reduction in the complexity of cardiovascular control systems (Scheffer et al., 1993; Kanaya et al., 2003; Tarvainen et al., 2012; Guzzetti et al., 2015a,b). These observations are in line with our data suggesting a reduction in the complexity of the physiological mechanisms regulating heart rate, as demonstrated by the reduction of CCE and the increase of the regularity index  $R_o$ .

The propofol administration resulted in a slowing of the HR during the induction phase, accompanied by an increase in cardiac sympathetic and reduction in cardiac vagal modulation. General anesthetic doses of propofol are associated with bradycardia, which is associated with low catecholamine plasma levels (Mustola et al., 1995) and with hypotension, which is the consequence of a direct effect of propofol on venous smooth muscle tone and presumably venous return (Cullen et al., 1987; Muzi et al., 1992; Koch et al., 2008). Moreover, propofol seems

to act at the cardiac level with a negative inotropic effect by depressing myocardial contractility (Riou et al., 1992; Cook and Housmans, 1994; Bilotta et al., 2001). Therefore, as per previous studies (Scheffer et al., 1993; Guzzetti et al., 2015a,b), our observations seem to confirm a direct effect of the propofol on the heart, resulting in a cardiac sympathetic activation and a reduction in vagal modulation in response to hemodynamic alterations, i.e., the effects on peripheral vascular control as a calcium antagonist.

As for the changes found in the transition from the second phase of induction to the first phase of recovery, it cannot be excluded that they are the result of the surgical intervention (Hirata et al., 2012). Further insights and the use of continuous ECG recordings are required to better understand the observed differences between the induction and awakening phases.

Finally, it is important to underline that there was a return of all the HRV parameters to the basal values, except for the HR. Despite a general recovery of autonomic functions, our data indicate that propofol-induced bradycardia may persist after discontinuing the infusion. Other studies have observed lasting effects of propofol on cardiovascular control with a resolution of the microcirculatory alterations 3 h after discontinuation of infusion (Kamijo et al., 1992; Koch et al., 2008) and that Remifentanyl effect on HRV is possible since Remifentanyl stimulates the parasympathetic nervous system but the effects of opioids on HRV remain poorly characterized. So future studies will take into consideration these effects also, considering that in the absence of nociception, HRV total power is reduced during GA with an increase of the proportional part of HF and that with Remifentanyl, HF decreases after nociceptive stimulation (Jeanne et al., 2009).

## Overview and Future Perspectives

Considering both EEG and HRV results in a wider framework, our data seem to support different theories on general anesthesia as well as on consciousness itself.

The so-called theory of the “consciousness switch” (Alkire et al., 2000) proposes that the mechanism supporting the loss of consciousness is a hyperpolarization block of thalamocortical neurons, due to a GABA neurotransmission enhancement and glutamate and cholinergic neurotransmission inhibition in the thalamocortical loops. A recent interesting perspective affirmed that anesthetic action results in an effective decrease in broadband neural activity with the neural structures involved become more regular and coherent. Indeed, Hutt et al. (2018) observed that weak fluctuations in the brain circuit caused a nonlinear oscillatory instability, yielding local intra-area synchrony in the  $\alpha$  frequency band. This intra-area synchrony disrupts synchrony between cortex and thalamus, with a relay role of the lateral geniculate nucleus, and contributes to the loss of consciousness. Moreover, it has been shown that a disruption of this system can drive thalamic nuclei into an attractor state of low-frequency bursting and further entrainment of thalamocortical circuits, with a characteristic downward shift of dominant  $\alpha$ -peaks in the EEG power spectra, together with increased power over the lower frequencies (van Wijngaarden et al., 2016). This is in line with the modification in EEG we

obtained, i.e., the decrease in posterior alpha and, after an initial increase in fronto-central beta power due to propofol and then a decrease of it, followed by an increase in slow frequencies typically observed in deep sedation and in thalamocortical disconnection.

The theory affirmed also that the precuneus and the adjacent areas within the posteromedial parietal cortex may play a central role as neural correlates of consciousness and self-reflection processes (Cavanna, 2007). Our results did not analyze the thalamocortical disconnection directly, but our findings revealed that the outgoing connectivity of the parietal region increased in the beta band in terms of both degree and strength, giving the impression that these nodes increased their information outflow processing during the induction phase of the anesthesia. Another system that should be taken into account for the consciousness and arousal modulation is the thalamic reticular activating system comprises the ascending reticular activating system (ARAS) and its thalamic targets, mainly intralaminar nuclei. The mesencephalic ARAS nuclei cover an array of structures, and experiments in mice have shown that the central lateral nucleus and ventrobasal nucleus co-operate to modulate the thalamocortical loops involved in temporal binding synchronizing cortical neuronal discharges as reported above (Llinás et al., 2002).

Considering the role of ANS, different results found in studies analyzing the changes in the HRV values during attentional tasks seemed to reveal that an increase in cardiac sympathetic modulation is counterbalanced by an increased vagal tone (Blase and van Waning, 2019). Our results are different as we highlighted a reduction in cardiac vagal modulation. However, the debate on the role of ANS during a performance that required focal or sustained attention is still controversial: Lower HRV is associated with worse attentional performance (Duschek et al., 2009; Williams et al., 2016), although these results were not confirmed by Frewen et al. (2013). The Neurovisceral Integration Hypothesis (Thayer and Lane, 2000, 2009), another interesting theory that tries to understand the mechanisms that allow consciousness generation, suggested that the brain areas involved in self-regulation and executive functions are also involved in cardiac autonomic activity through the vagus nerve (Thayer and Sternberg, 2006; Thayer et al., 2010), although the real mechanisms behind this association are far to be clear. An example of this activity is the so-called baroreceptor reflex, which coordinates cardiac output with blood pressure. A primary function of the cardiovascular system is to maintain optimal arterial blood pressure and to provide adequate blood flow to the brain, and other organs. The baroreceptor reflex involves lower brainstem nuclei that detect organ-specific visceral activity *via* afferent sympathetic and parasympathetic nerve pathways, using this information in homeostatic, reflex-like regulation of autonomic tone. The Neurovisceral Integration Hypothesis supports the idea that new types of information are hierarchically integrated between the vagal system, brainstem, and some cortical areas, suggesting that each level is more flexibly recruited to modify vagal tone than the level below (Thayer and Lane, 2000, 2009). The inverse relationship between HRV and blood pressure variability is an example of an important



physiological regulation strategy known as heterostasis (Davis, 1958), where variability in HRV is associated with stability in the blood pressure system. The hierarchical, prolonged and coordinated information transmission of blood pressure system with information derived from brainstem nuclei (e.g., nucleus ambiguus, the dorsal motor nucleus of the vagus, nucleus of the solitary tract) and other brain structures (e.g., amygdala and the basal forebrain) seems to be involved in the regulations of EEG bands (Sanchez-Alavez et al., 2014), as well as seems to be fundamental for downward predictions and for minimizing error in motor tasks, for example (Smith et al., 2017). It is premature to infer specific causal relationships among these systems and our results but a deep analysis of the interactions among these systems could be useful in the future.

Another interesting result is that the processes involved in the loss of consciousness seem to be different from the ones related to the recovery phase. This creates a new perspective according to the two phases of general anesthesia that should be studied separately: the mechanisms linked to the “switch on” of consciousness/attention could be not the same as the “switch off” process (Uhrig et al., 2014). Indeed, in our results, the changes in outgoing connectivity values of beta bands in parietal ROI were not restored during the recovery phase confirming how, as reported also by other authors (Blain-Moraes et al., 2017), the CNS (mainly) and ANS (for heart rate) mechanisms could be different during the loss and recovery of consciousness. If this result will be confirmed, the monitoring systems of the anesthesia during the two phases could be different in the near future to increase the sensitivity, with (e.g.) a focus on the posterior electrical activity of the brain during the induction phase (when possible) and a deep monitor of the HRV, instead, during the recovery phase.

Finally, another interesting theory called Global Neuronal Workspace (GNW; Dehaene et al., 1998; Dehaene and Changeux, 2011), explaining the brain mechanisms behind disorders of consciousness, affirmed that cortico-cortical long-range axons in prefronto-parietotemporo-cingulate cortex constitute the brain network through which the primary sensory information became conscious. The anesthetic drug seems to cause a widespread cortical activity disruption in the neurons of these areas with long-distance connectivity, while the delay for thalamic activity disruption is related to the indirect consequences of cortical feedback on the thalamus (Velly et al., 2007). Our results seem to confirm this evidence considering that we found an increase in the local efficiency value during the administration of propofol in alpha e delta bands with no statistically significant differences in global efficiency. Nevertheless, an explanation of the mechanisms through which these activities were associated with the results from ANS data along with the progressive bradycardia due to anesthesia is still lacking and it needs to be analyzed in future studies focusing on Central and Autonomic nervous systems correlates of consciousness (Riganello et al., 2019; Sattin et al., 2020).

## Limits

This pilot study has some limitations that should be considered. First, the number of patients analyzed was small and we cannot

draw any general conclusions. Furthermore, several variables should be taken into account to limit their confounding role; for instance, the neurosurgical and reanimation settings required careful work to limit the confounding variables influencing data acquisition, and single-subject variability in response to anesthetic agents should be considered as well. However, the aim of this pilot study was only to explore the differences among epochs both in the CNS and in the ANS during the propofol-induced anesthesia through some measures that gave us interesting perspectives that should be tested in future studies with adequate sample size. Another limitation of our study was that the total amount of Remifentanyl administered during GA was not recorded and this information is lacking in the discussion about the recovery period, although we monitored the Ce range in according to our GA protocol. However, we think that minimal Remifentanyl dosages (as at ROC in our study) should not affect the EEG signals.

Regarding limitations in EEG analyses, PDC and DTF, were used as they did not take into account instantaneous connectivity. The choice of the reference and volume conduction influence is still debatable, while some claim that DTF and PDC can indeed disentangle the true connectivity (Kaminski and Blinowska, 2014), even if spurious connections are present (Kaminski and Blinowska, 2017), others (Brunner et al., 2016) proved otherwise. The volume conduction error could affect the analysis on sensors and connectivity should have been done on source data. Unfortunately, with only 19 channels, source reconstruction could not be performed. In future studies, an HD-EEG or MEG should be used instead, besides source reconstruction and volume correction methods.

Furthermore, a limitation was related to the fact that we analyzed the EEG and ECG signals in a parallel way. Future studies are needed to combine all these signals.

Among the other limitations that could be highlighted in this pilot study, we want finally to underline that we only considered propofol anesthetic in our research and, hence, every general conclusion about brain mechanisms or neural processes should be cautiously discussed. Indeed, some studies reported as other drugs act on brain networks in different ways concerning propofol (Vakkuri et al., 2004; Nicolaou and Georgiou, 2014) producing the same loss of consciousness fundamental for a surgery. In this sense, future studies are needed to verify the association between CNS and ANS changes during the loss and recovery of consciousness in a wider sample.

## Conclusion

In conclusion, the results of this pilot study showed as propofol-induced anesthesia administration caused modifications affecting several frequency bands of the EEG signal, leading to a “rebalance” between long and short-range cortical connections with an increase in local efficiency in alpha e delta bands during the induction phase. Moreover, different ROIs, namely frontal and parietal ones, showed opposite effects with an increase in the incoming values in the alpha bands during the induction of anesthesia and a statistically significant comparison in the outgoing values in the beta bands, respectively. Finally, the induction of anesthesia with

propofol determined a progressive decrease in the complexity of the HRV and bradycardia which was accompanied by an increase in cardiac sympathetic modulation and a decrease in cardiac parasympathetic modulation during the induction phase followed by a return to basal except for the heart rate index, suggesting interesting perspectives for the interactions between central and autonomic nervous system during anesthesia.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://zenodo.org/communities/besta/?page=1&size=20>.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Local Ethics Committee Fondazione IRCCS Istituto Neurologico Carlo Besta. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

DS and DD: conceptualization, data interpretation, original draft writing, and critical revision of the manuscript draft. DD, ES, FP, DRS, EV, SF, ET, AC, VC, and NM: data collection,

data analyses and interpretation, and critical revision of the manuscript draft. SV, CC, EO, and DC: data collection and interpretation, critical revision of the manuscript draft. FGM and MC: data interpretation and critical revision of the manuscript draft. MP and LM: conceptualization and critical revision of the manuscript draft.

## FUNDING

This work was supported by Ministero della Salute award id RC-2018.

## ACKNOWLEDGMENTS

The present publication was submitted on behalf of the Coma Research Centre (CRC) multidisciplinary team. We thank all patients involved in the study and Rosalind Hendricks and Alessandra Atterrato for their support in language revision and Document delivery, respectively.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnsys.2021.652080/full#supplementary-material>.

## REFERENCES

- Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., and Suri, J. S. (2006). Heart rate variability: a review. *Med. Biol. Eng. Comput.* 44, 1031–1051. doi: 10.1007/s11517-006-0119-0
- Alkire, M., Haier, R., and Fallon, J. (2000). Toward a unified theory of narcosis: brain imaging evidence for a thalamocortical switch as the neurophysiologic basis of anesthetic-induced unconsciousness. *Conscious. Cogn.* 9, 370–386. doi: 10.1006/ccog.1999.0423
- Baccalá, L. A., and Sameshima, K. (2001). Overcoming the limitations of correlation analysis for many simultaneously processed neural structures. *Prog. Brain Res.* 130, 33–47. doi: 10.1016/s0079-6123(01)30004-3
- Bennett, C., Voss, L. J., Barnard, J. P. M., and Sleight, J. W. (2009). Practical use of the raw electroencephalogram waveform during general anesthesia: the art and science. *Anesth. Analg.* 109, 539–550. doi: 10.1213/ane.0b013e3181a9fc38
- Bilotta, F., Fiorani, L., La Rosa, I., Spinelli, F., and Rosa, G. (2001). Cardiovascular effects of intravenous propofol administered at two infusion rates: a transthoracic echocardiographic study. *Anaesthesia* 56, 266–271. doi: 10.1046/j.1365-2044.2001.01717-5.x
- Blain-Moraes, S., Tarnal, V., Vanini, G., Bel-Behar, T., Janke, E., Picton, P., et al. (2017). Network efficiency and posterior alpha patterns are markers of recovery from general anesthesia: a high-density electroencephalography study in healthy volunteers. *Front. Hum. Neurosci.* 11:328. doi: 10.3389/fnhum.2017.00328
- Blase, K. L., and van Waning, A. (2019). Heart rate variability, cortisol and attention focus during shamatha quiescence meditation. *Appl. Psychophysiol. Biofeedback* 44, 331–342. doi: 10.1007/s10484-019-09448-w
- Bonhomme, V., Staquet, C., Montupil, J., Defresne, A., Kirsch, M., Martial, C., et al. (2019). General anesthesia: a probe to explore consciousness. *Front. Syst. Neurosci.* 13:36. doi: 10.3389/fnsys.2019.00036
- Bosse, T., Jonker, C. M., and Treur, J. (2008). Formalisation of Damasio's theory of emotion, feeling and core consciousness. *Conscious. Cogn.* 17, 94–113. doi: 10.1016/j.concog.2007.06.006
- Boussen, S., Spiegler, A., Benar, C., Carrère, M., Bartolomei, F., Metellus, P., et al. (2018). Time rescaling reproduces EEG behavior during transition from propofol anesthesia-induced unconsciousness to consciousness. *Sci. Rep.* 8:6015. doi: 10.1038/s41598-018-24405-z
- Brunner, C., Billinger, M., Seeber, M., Mullen, T. R., Makeig, S., Lansky, P., et al. (2016). Volume conduction influences scalp-based connectivity estimates. *Front. Comput. Neurosci.* 10:121. doi: 10.3389/fncom.2016.00121
- Cavanna, A. E. (2007). The precuneus and consciousness. In *CNS Spectr.* 12, 545–552. doi: 10.1017/s1092852900021295
- Chander, D., García, P. S., MacColl, J. N., Illing, S., and Sleight, J. W. (2014). Electroencephalographic variation during end maintenance and emergence from surgical anesthesia. *PLoS One* 9:e106291. doi: 10.1371/journal.pone.0106291
- Chennu, S., O'Connor, S., Adapa, R., Menon, D. K., and Bekinschtein, T. A. (2016). Brain connectivity dissociates responsiveness from drug exposure during propofol-induced transitions of consciousness. *PLoS Comput. Biol.* 12:e1004669. doi: 10.1371/journal.pcbi.1004669
- Chernik, D. A., Gillings, D., Laine, H., Hendler, J., Silver, J. M., Davidson, A. B., et al. (1990). Validity and reliability of the observer's assessment of alertness/sedation scale: study with intravenous midazolam. *J. Clin. Psychopharmacol.* 10, 244–251.
- Cook, D. J., and Housmans, P. R. (1994). Mechanism of the negative inotropic effect of propofol in isolated ferret ventricular myocardium. *Anesthesiology* 80, 859–871. doi: 10.1097/0000542-199404000-00020
- Cullen, P., Turtle, M., Prys-Roberts, C., Way, W., and Dye, J. (1987). Effect of propofol anesthesia on baroreflex activity in humans. *Anesth. Analg.* 66, 1115–1120.
- Damasio, A. (2012). *Self Comes to Mind Constructing the Conscious Brain*. (Pantheon, Rome: Random House).
- Davis, R. (1958). The domain of homeostasis. *Psychol. Rev.* 65, 8–13. doi: 10.1037/h0045358
- Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018

- Dehaene, S., Kerszberg, M., and Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. U S A* 95, 14529–14534. doi: 10.1073/pnas.95.24.14529
- Duschek, S., Muckenthaler, M., Werner, N., and Del Paso, G. (2009). Relationships between features of autonomic cardiovascular control and cognitive performance. *Biolo. Psychol.* 81, 110–117. doi: 10.1016/j.biopsycho.2009.03.003
- Ebert, T. J. (2005). Sympathetic and hemodynamic effects of moderate and deep sedation with propofol in humans. *Anesthesiology* 103, 20–24. doi: 10.1097/0000542-200507000-00007
- Frewen, J., Finucane, C., Savva, G. M., Boyle, G., Coen, R. F., and Kenny, R. A. (2013). Cognitive function is associated with impaired heart rate variability in ageing adults: the Irish longitudinal study on ageing wave one results. *Clin. Auton. Res.* 23, 313–323. doi: 10.1007/s10286-013-0214-x
- Garret-Bernardin, A., Cantile, T., D'Antò, V., Galanakis, A., Fauxpoint, G., Ferrazzano, G. F., et al. (2017). Pain experience and behavior management in pediatric dentistry: a comparison between traditional local anesthesia and the wand computerized delivery system. *Pain Res. Manag.* 2017:7941238. doi: 10.1155/2017/7941238
- Gugino, L. D., Chabot, R. J., Prichep, L. S., John, E. R., Formanek, V., and Aglio, L. S. (2001). Quantitative EEG changes associated with loss and return of consciousness in healthy adult volunteers anaesthetized with propofol or sevoflurane. *Br. J. Anaesth.* 87, 421–428. doi: 10.1093/bja/87.3.421
- Guzzetti, S., Bassani, T., Latini, R., Masson, S., Barlera, S., Citerio, G., et al. (2015a). Autonomic cardiovascular modulation with three different anesthetic strategies during neurosurgical procedures. *Minerva Anestesiol.* 81, 3–11.
- Guzzetti, S., Marchi, A., Bassani, T., Citerio, G., and Porta, A. (2015b). Univariate and bivariate symbolic analyses of cardiovascular variability differentiate general anesthesia procedures. *Physiol. Meas.* 36:715. doi: 10.1088/0967-3334/36/4/715
- Guzzetti, S., Borroni, E., Garbelli, P. E., Ceriani, E., Della Bella, P., Montano, N., et al. (2005). Symbolic dynamics of heart rate variability: a probe to investigate cardiac autonomic modulation. *Circulation* 112, 465–470. doi: 10.1161/CIRCULATIONAHA.104.518449
- Hirata, N., Miyashita, R., Maruyama, D., Kawaguchi, R., Shimizu, H., and Yamakage, M. (2012). Heart rate variability during abdominal surgical manipulation under general and epidural anesthesia. *J. Anesth.* 26, 900–904. doi: 10.1007/s00540-012-1431-5
- Hutt, A., Lefebvre, J., Hight, D., and Sleight, J. (2018). Suppression of underlying neuronal fluctuations mediates EEG slowing during general anaesthesia. *NeuroImage* 179, 414–428. doi: 10.1016/j.neuroimage.2018.06.043
- Jeanne, M., Logier, R., De Jonckheere, J., and Tavernier, B. (2009). Heart rate variability during total intravenous anesthesia: effects of nociception and analgesia. *Auton. Neurosci.* 147, 91–96. doi: 10.1016/j.autneu.2009.01.005
- John, E. R., Prichep, L. S., Kox, W., Valdés-Sosa, P., Bosch-Bayard, J., Aubert, E., et al. (2001). Invariant reversible QEEG effects of anesthetics. *Conscious. Cogn.* 10, 165–183. doi: 10.1006/ccog.2001.0507
- Kamijo, Y., Goto, H., Nakazawa, K., Benson, K. T., and Arakawa, K. (1992). Arterial baroreflex attenuation during and after continuous propofol infusion. *Can. J. Anaesth.* 39, 987–991. doi: 10.1007/BF03008351
- Kaminski, M., and Blinowska, K. J. (2014). Directed Transfer Function is not influenced by volume conduction-inexpedient pre-processing should be avoided. *Front. Comput. Neurosci.* 8:61. doi: 10.3389/fncom.2014.00061
- Kaminski, M., and Blinowska, K. J. (2017). The influence of volume conduction on DTF estimate and the problem of its mitigation. *Front. Comput. Neurosci.* 11:36. doi: 10.3389/fncom.2017.00036
- Kanaya, N., Hirata, N., Kurosawa, S., Nakayama, M., and Namiki, A. (2003). Differential effects of propofol and sevoflurane on heart rate variability. *Anesthesiology* 98, 34–40. doi: 10.1097/0000542-200301000-00009
- Kim, H., Moon, J. Y., Mashour, G. A., and Lee, U. C. (2018). Mechanisms of hysteresis in human brain networks during transitions of consciousness and unconsciousness: theoretical principles and empirical evidence. *PLoS Comput. Biol.* 14:e1006424. doi: 10.1371/journal.pcbi.1006424
- Kim, P. J., Kim, H. G., Noh, G. J., Koo, Y. S., and Shin, T. J. (2017). Disruption of frontal-parietal connectivity during conscious sedation by propofol administration. *NeuroReport* 28, 896–902. doi: 10.1097/WNR.0000000000000853
- Koch, M., De Backer, D., Vincent, J. L., Barvais, L., Hennart, D., and Schmartz, D. (2008). Effects of propofol on human microcirculation. *Br. J. Anaesth.* 101, 473–478. doi: 10.1093/bja/aen210
- Koitaibashi, T., Johansen, J. W., and Sebel, P. S. (2002). Remifentanyl dose/electroencephalogram bispectral response during combined propofol/regional anesthesia. *Anesth. Analg.* 94, 1530–1533. doi: 10.1097/0000539-200206000-00028
- Kreuzer, M., Kochs, E. F., Schneider, G., and Jordan, D. (2014). Non-stationarity of EEG during wakefulness and anaesthesia: advantages of EEG permutation entropy monitoring. *J. Clin. Monit. Comput.* 28, 573–580. doi: 10.1007/s10877-014-9553-y
- Kuizenga, M. H., Colin, P. J., Reyntjens, K. M. E. M., Touw, D. J., Nalbat, H., Knotnerus, F. H., et al. (2018). Test of neural inertia in humans during general anaesthesia. *Br. J. Anaesth.* 120, 525–536. doi: 10.1016/j.bja.2017.11.072
- Lee, H., Mashour, G. A., Noh, G. J., Kim, S., and Lee, U. (2013a). Reconfiguration of network hub structure after propofol-induced unconsciousness. *Anesthesiology* 119, 1347–1359. doi: 10.1097/ALN.0b013e3182a8ec8c
- Lee, U., Ku, S., Noh, G., Baek, S., Choi, B., and Mashour, G. A. (2013b). Disruption of frontal-parietal communication by ketamine, propofol and sevoflurane. *Anesthesiology* 118, 1264–1275. doi: 10.1097/ALN.0b013e31829103f5
- Lee, U., Oh, G., Kim, S., Noh, G., Choi, B., and Mashour, G. A. (2010). Brain networks maintain a scale-free organization across consciousness, anesthesia and recovery: evidence for adaptive reconfiguration. *Anesthesiology* 113, 1081–1091. doi: 10.1097/ALN.0b013e3181f229b5
- Llinás, R. R., Leznik, E., and Urbano, F. J. (2002). Temporal binding via cortical coincidence detection of specific and nonspecific thalamocortical inputs: a voltage-dependent dye-imaging study in mouse brain slices. *Proc. Natl. Acad. Sci. U S A* 99, 449–454. doi: 10.1073/pnas.012604899
- Logier, R., Jeanne, M., Tavernier, B., and De Jonckheere, J. (2006). Pain/analgesia evaluation using heart rate variability analysis. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2006, 4303–4306. doi: 10.1109/IEMBS.2006.260494
- Mäenpää, M., Laitio, T., Kuusela, T., Penttilä, J., Kaisti, K., Aalto, S., et al. (2011). Delta entropy of heart rate variability along with deepening anesthesia. *Anesth. Analg.* 112, 587–592. doi: 10.1097/QAD.00000000000002883
- McFadden, J. (2007). Conscious electromagnetic (CEMI) field theory. *NeuroQuantology* 5, 262–270. doi: 10.14704/nq.2007.5.3.135
- Minto, C. F., Schnider, T. W., and Shafer, S. L. (1997). Pharmacokinetics and Pharmacodynamics of Remifentanyl. *Anesthesiology* 86, 24–33. doi: 10.1097/0000542-199701000-00005
- Mustola, S. T., baer, G. A., metsä-ketelä, T., and laippala, P. (1995). Haemodynamic and plasma catecholamine responses during total intravenous anaesthesia for laryngomicroscopy: thiopentone compared with propofol. *Anaesthesia* 50, 108–113. doi: 10.1111/j.1365-2044.1995.tb15090.x
- Muzi, M., Berens, R. A., Kampine, J. P., and Ebert, T. J. (1992). Venodilatation contributes to propofol-mediated hypotension in humans. *Anesth. Analg.* 74, 877–883. doi: 10.1213/0000539-199206000-00017
- Nicolaou, N., and Georgiou, J. (2014). Neural network-based classification of anesthesia/awareness using Granger causality features. *Clin. EEG Neurosci.* 45, 77–88. doi: 10.1177/1550059413486271
- Pinsker, M. C. (2012). Unresponsiveness versus unconsciousness. *Anesthesiology* 117:1140. doi: 10.1097/ALN.0b013e31826f8b9a
- Porta, A., Guzzetti, S., Montano, N., Furlan, R., Pagani, M., Malliani, A., et al. (2001). Entropy, entropy rate and pattern classification as tools to typify complexity in short heart period variability series. *IEEE Trans. Biomed. Eng.* 48, 1282–1291. doi: 10.1109/10.959324
- Porta, A., Gnechchi-Ruscone, T., Tobaldini, E., Guzzetti, S., Furlan, R., Montano, N., et al. (2007a). Progressive decrease of heart period variability entropy-based complexity during graded head-up tilt. *J. Appl. Physiol.* 103, 1143–1149. doi: 10.1152/japplphysiol.00293.2007
- Porta, A., Tobaldini, E., Guzzetti, S., Furlan, R., Montano, N., Gnechchi-Ruscone, T., et al. (2007b). Assessment of cardiac autonomic modulation during graded head-up tilt by symbolic analysis of heart rate variability. *Am. J. Physiol. Heart Circ. Physiol.* 293, 702–708. doi: 10.1152/ajpheart.00006.2007
- Prakash, R., Prakash, O., Prakash, S., Abhishek, P., and Gandotra, S. (2008). Global workspace model of consciousness and its electromagnetic correlates. *Ann. Indian Acad. Neurol.* 11, 146–153. doi: 10.4103/0972-2327.42933
- Purdon, P. L., Pierce, E. T., Mukamel, E. A., Prerau, M. J., Walsh, J. L., Wong, K. F. K., et al. (2013). Electroencephalogram signatures of loss and

- recovery of consciousness from propofol. *Proc. Natl. Acad. Sci. U S A* 110, E1142–E1151. doi: 10.1073/pnas.1221180110
- Riganello, F., Larroque, S. K., Perri, C. D., Prada, V., Sannita, W. G., Laureys, S., et al. (2019). Measures of CNS-autonomic interaction and responsiveness in disorder of consciousness. *Front. Neurosci.* 13:530. doi: 10.3389/fnins.2019.00530
- Riou, B., Besse, S., Lecarpentier, Y., and Viars, P. (1992). *in vitro* effects of propofol on rat myocardium. *Anesthesiology* 76, 609–616. doi: 10.1097/00000542-199204000-00019
- Riznyk, L., Fijakowska, M., and Przesmycki, K. (2005). Effects of thiopental and propofol on heart rate variability during fentanyl-based induction of general anesthesia. *Pharmacol. Rep.* 57, 128–134.
- Ryu, J. H., Kim, P. J., Kim, H. G., Koo, Y. S., and Shin, T. J. (2017). Investigating the effects of nitrous oxide sedation on frontal-parietal interactions. *Neurosci. Lett.* 651, 9–15. doi: 10.1016/j.neulet.2017.04.036
- Sanchez-Alavez, M., Robledo, P., Wills, D. N., Havstad, J., and Ehlers, C. L. (2014). Cholinergic modulation of event-related oscillations (ERO). *Brain Res.* 559, 11–25. doi: 10.1016/j.brainres.2014.02.043
- Sanders, R. D., Tononi, G., Laureys, S., and Sleight, J. W. (2012). Unresponsiveness ≠ unconsciousness. *Anesthesiology* 116, 946–959. doi: 10.1097/ALN.0b013e318249d0a7
- Sattin, D., Leonardi, M., and Picozzi, M. (2020). The autonomic nervous system and the brainstem: a fundamental role or the background actors for consciousness generation? Hypothesis, evidence and future directions for rehabilitation and theoretical approaches. *Brain Behav.* 10:e01474. doi: 10.1002/brb3.1474
- Sawaguchi, K., Matsuura, N., and Ichinohe, T. (2016). Comparison of the effect of electrical stimulations on the chin skin on autonomic nervous activities during propofol sedation with or without midazolam. *J. Oral Maxillofac. Surg.* 74:1751. doi: 10.1016/j.joms.2016.03.040
- Scheffer, G. J., Ten Voorde, B. J., Karemaker, J. M., Ros, H. H., and De Lange, J. J. (1993). Effects of thiopentone, etomidate and propofol on beat-to-beat cardiovascular signals in man. *Anaesthesia* 48, 849–855. doi: 10.1111/j.1365-2044.1993.tb07412.x
- Schneider, G., Jordan, D., Schwarz, G., Bischoff, P., Kalkman, C. J., Kuppe, H., et al. (2014). Monitoring depth of anesthesia utilizing a combination of electroencephalographic and standard measures. *Anesthesiology* 120, 819–828. doi: 10.1097/ALN.0000000000000151
- Schnider, T. W., Minto, C. F., Shafer, S. L., Gambus, P. L., Andresen, C., Goodale, D. B., et al. (1999). The influence of age on propofol pharmacodynamics. *Anesthesiology* 90, 1502–1516. doi: 10.1097/00000542-199906000-00003
- Sebel, P. S., and Lowdon, J. D. (1989). Propofol: a new intravenous anesthetic. *Anesthesiology* 71, 260–277.
- Sheeba, J. H., Stefanovska, A., and McClintock, P. V. E. (2008). Neuronal synchrony during anesthesia: a thalamocortical model. *Biophys. J.* 95, 2722–2727. doi: 10.1529/biophysj.108.134635
- Smith, R., Thayer, J. F., Khalsa, S. S., and Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neurosci. Biobehav. Rev.* 75, 274–296. doi: 10.1016/j.neubiorev.2017.02.003
- Soplat, A. E., McCarthy, M. M., Sherfey, J., Lee, S., Purdon, P. L., Brown, E. N., et al. (2017). Thalamocortical control of propofol phase-amplitude coupling. *PLoS Comput. Biol.* 13:e1005879. doi: 10.1371/journal.pcbi.1005879
- Storella, R. J., Morrow, J. C., and Polansky, M. (1999). Differences among heart rate variability measures after anesthesia and cardiac surgery. *J. Cardiothorac. Vasc. Anesth.* 13, 451–453. doi: 10.1016/s1053-0770(99)90219-7
- Supp, G. G., Siegel, M., Hipp, J. F., and Engel, A. K. (2011). Cortical hypersynchrony predicts breakdown of sensory processing during loss of consciousness. *Curr. Biol.* 21, 1988–1993. doi: 10.1016/j.cub.2011.10.017
- Tarvainen, M. P., Georgiadis, S., Laitio, T., Lipponen, J. A., Karjalainen, P. A., Kaskinoro, K., et al. (2012). Heart rate variability dynamics during low-dose propofol and dexmedetomidine anesthesia. *Ann. Biomed. Eng.* 40, 1802–1813. doi: 10.1007/s10439-012-0544-1
- Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. *Circulation* 93, 1043–1065. doi: 10.1161/01.CIR.93.5.1043
- Thayer, J. F., and Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *J. Affect. Disord.* 61, 201–216. doi: 10.1016/s0165-0327(00)00338-4
- Thayer, J. F., and Lane, R. D. (2009). Claude Bernard and the heart-brain connection: further elaboration of a model of neurovisceral integration. In *Neurosci. Biobehav. Rev.* 33, 81–88. doi: 10.1016/j.neubiorev.2008.08.004
- Thayer, J. F., and Sternberg, E. (2006). Beyond heart rate variability: vagal regulation of allostatic systems. *Ann. N Y Acad. Sci.* 1088, 361–372. doi: 10.1196/annals.1366.014
- Thayer, J. F., Yamamoto, S. S., and Brosschot, J. F. (2010). The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. *Int. J. Cardiol.* 141, 122–131. doi: 10.1016/j.ijcard.2009.09.543
- Tononi, G., and Edelman, G. (1998). “Consciousness and the integration of information in the brain,” in *Consciousness: at the Frontiers of Neuroscience*, eds H. H. Jasper, L. Descarries, V. F. Castellucci, and S. Rossignol (Philadelphia, PA: Lippencott-Raven), 245–279.
- Uhrig, L., Dehaene, S., and Jarraya, B. (2014). Cerebral mechanisms of general anesthesia. *Ann. Fr. Anesth. Reanim.* 33, 72–82. doi: 10.1016/j.annfar.2013.11.005
- Vakkuri, A., Yli-Hankala, A., Talja, P., Mustola, S., Tolvanen-Laakso, H., Sampson, T., et al. (2004). Time-frequency balanced spectral entropy as a measure of anesthetic drug effect in central nervous system during sevoflurane, propofol and thiopental anesthesia. *Acta Anaesthesiol. Scand.* 48, 145–153. doi: 10.1111/j.0001-5172.2004.00323.x
- van Wijngaarden, J. B. G., Zucca, R., Finnigan, S., and Verschure, P. F. M. J. (2016). The impact of cortical lesions on thalamo-cortical network dynamics after acute ischaemic stroke: a combined experimental and theoretical study. *PLoS Comput. Biol.* 12:e1005048. doi: 10.1371/journal.pcbi.1005048
- Velly, L. J., Rey, M. F., Bruder, N. J., Gouvisos, F. A., Witjas, T., Regis, J. M., et al. (2007). Differential dynamic of action on cortical and subcortical structures of anesthetic agents during induction of anesthesia. *Anesthesiology* 107, 202–212. doi: 10.1097/01.anes.0000270734.99298.b4
- Vlissides, P. E., Li, D., Zierau, M., Lapointe, A. P., Ip, K. I., McKinney, A. M., et al. (2019). Dynamic cortical connectivity during general anesthesia in surgical patients. *Anesthesiology* 130, 885–897. doi: 10.1097/ALN.0000000000002677
- Williams, D. W. P., Thayer, J. F., and Koenig, J. (2016). Resting cardiac vagal tone predicts intraindividual reaction time variability during an attention task in a sample of young and healthy adults. *Psychophysiology* 53, 1843–1851. doi: 10.1111/psyp.12739
- Yeom, S. K., Won, D. O., Chi, S. I., Seo, K. S., Kim, H. J., Müller, K. R., et al. (2017). Spatio-temporal dynamics of multimodal EEG-fNIRS signals in the loss and recovery of consciousness under sedation using midazolam and propofol. *PLoS One* 12:e0187743. doi: 10.1371/journal.pone.0187743
- Yu, L., and Blumenfeld, H. (2009). Theories of impaired consciousness in epilepsy. *Ann. N Y Acad. Sci.* 1157, 48–60. doi: 10.1111/j.1749-6632.2009.04472.x

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Sattin, Duran, Visintini, Schiaffi, Panzica, Carozzi, Rossi Sebastiano, Visani, Tobaldini, Carandina, Citterio, Magnani, Cacciatore, Orena, Montano, Caldiroli, Franceschetti, Picozzi and Matilde. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Dynamic Patterns of Global Brain Communication Differentiate Conscious From Unconscious Patients After Severe Brain Injury

Daniel Golkowski<sup>1,2\*</sup>, Rebecca Willnecker<sup>1</sup>, Jennifer Rösler<sup>1</sup>, Andreas Ranft<sup>3</sup>, Gerhard Schneider<sup>3</sup>, Denis Jordan<sup>3†</sup> and Rüdiger Ilg<sup>1,4†</sup>

<sup>1</sup> Department of Neurology, School of Medicine, Technical University of Munich, Munich, Germany, <sup>2</sup> Neurologische Klinik und Poliklinik, University of Heidelberg, Heidelberg, Germany, <sup>3</sup> Department of Anesthesiology and Intensive Care, School of Medicine, Technical University of Munich, Munich, Germany, <sup>4</sup> Asklepios Clinic, Department of Neurology, Bad Tölz, Germany

## OPEN ACCESS

### Edited by:

Kathinka Evers,  
Uppsala University, Sweden

### Reviewed by:

Robert N. S. Sachdev,  
Humboldt University of Berlin,  
Germany  
Caroline Schnakers,  
Casa Colina Hospital and Centers  
for Healthcare, United States  
Zirui Huang,  
University of Michigan, United States

### \*Correspondence:

Daniel Golkowski  
daniel.golkowski@tum.de

<sup>†</sup> These authors have contributed  
equally to this work

**Received:** 04 November 2020

**Accepted:** 17 August 2021

**Published:** 09 September 2021

### Citation:

Golkowski D, Willnecker R,  
Rösler J, Ranft A, Schneider G,  
Jordan D and Ilg R (2021) Dynamic  
Patterns of Global Brain  
Communication Differentiate  
Conscious From Unconscious  
Patients After Severe Brain Injury.  
*Front. Syst. Neurosci.* 15:625919.  
doi: 10.3389/fnsys.2021.625919

The neurophysiology of the subjective sensation of being conscious is elusive; therefore, it remains controversial how consciousness can be recognized in patients who are not responsive but seemingly awake. During general anesthesia, a model for the transition between consciousness and unconsciousness, specific covariance matrices between the activity of brain regions that we call patterns of global brain communication reliably disappear when people lose consciousness. This functional magnetic imaging study investigates how patterns of global brain communication relate to consciousness and unconsciousness in a heterogeneous sample during general anesthesia and after brain injury. First, we describe specific patterns of global brain communication during wakefulness that disappear during propofol ( $n = 11$ ) and sevoflurane ( $n = 14$ ) general anesthesia. Second, we search for these patterns in a cohort of unresponsive wakeful patients ( $n = 18$ ) and unmatched healthy controls ( $n = 20$ ) in order to evaluate their potential use in clinical practice. We found that patterns of global brain communication characterized by high covariance in sensory and motor areas or low overall covariance and their dynamic change were strictly associated with intact consciousness in this cohort. In addition, we show that the occurrence of these two patterns is significantly related to activity within the frontoparietal network of the brain, a network known to play a crucial role in conscious perception. We propose that this approach potentially recognizes consciousness in the clinical routine setting.

**Keywords:** consciousness, brain injury, coma, unresponsive wakefulness syndrome, fMRI, anesthesia, propofol, sevoflurane

## INTRODUCTION

From a basic research perspective, the subjective experience of being aware of oneself and the environment most likely emerges as an epiphenomenon of cerebral information processing (Dehaene et al., 2003, 2017) and is thus encoded in the activity of the brain and its neurons. Attempts to identify brain activity specific for state of being consciously focused on the question of where conscious perception takes place (Dehaene and Changeux, 2011) and how brain regions

are affected by models for unconsciousness, namely, general anesthesia, sleep, or unconsciousness after brain injury (Brown et al., 2010). These approaches were able to link conscious brain function with widespread brain networks: the frontoparietal network, default mode network, and ascending reticular activation system as modulators of these (Boveroux et al., 2010; Demertzi et al., 2015; Bonhomme et al., 2016; Ranft et al., 2016). The activity within these networks could be associated with conscious perception [frontoparietal network, Dehaene et al. (2003) and He et al. (2007)], mind wandering [default mode network, Vanhaudenhuyse et al. (2011)], or the level of consciousness (Gili et al., 2013). In addition, the activity within these networks was specifically diminished during general anesthesia irrespective of the anesthetic agent used (Boveroux et al., 2010; Jordan et al., 2013; Bonhomme et al., 2016; Ranft et al., 2016). However, measuring the activity only in the named networks in patients after brain injury only unreliably detected intact consciousness after brain injury (Vanhaudenhuyse et al., 2010). We hypothesized that conscious processing in the brain requires not only specific networks as a common endpoint but also involves various levels of information processing.

Based on this idea, our data analysis aimed to integrate different levels of hierarchy and considered information processing in the human brain to be divided into (Dehaene et al., 2017) local information processing within specialized brain areas, for example in the primary visual cortex, and (Dehaene et al., 2003) global communication reflected by information exchange between such areas, for instance when we react to complex external stimuli. Both types of information processing potentially play an important role in the generation of consciousness (Barttfeld et al., 2015; Ranft et al., 2016; Demertzi et al., 2019; Golkowski et al., 2019). In addition, we assumed that the dynamics of this local and global information processing is a key feature of consciousness.

Local information processing in unconscious humans is known to be significantly diminished in the medial prefrontal cortex, precuneus, posterior cingulate cortex, superior parietal lobe, and in the dorsolateral prefrontal and inferior parietal cortices during general anesthesia when analyzed by functional MRI (fMRI) (Ranft et al., 2016; Golkowski et al., 2019). In awake humans, these brain areas frequently communicate and are known as the default mode network and the frontoparietal network. This data analysis identified this local information processing through independent maps of a spatial independent component analysis' (ICA) resulting in 57 brain regions covering the whole brain and their time course of activity.

Global communication has been mainly investigated between specific brain areas using fMRI and EEG. Like local information processing, differences between consciousness and unconsciousness were observed mainly in the frontoparietal network and the default mode network. Specifically, information exchange between the areas encompassing these networks was significantly reduced in various studies on unconsciousness (He et al., 2007; Jordan et al., 2013; Bonhomme et al., 2016; Ranft et al., 2016; van Vugt et al., 2018). Later, it was demonstrated that patterns of global brain communication are significantly altered during unconsciousness and that the transition between

patterns was reduced (Barttfeld et al., 2015; Hudetz et al., 2015; Ma et al., 2017; Golkowski et al., 2019). Similar results were obtained in patients suffering from disorders of consciousness after brain injury, thus showing that the concept of global brain communication can be generalized to patients with different reasons for unconsciousness (Luppi et al., 2019). We modeled this global brain communication by calculating the correlation between brain regions in a sliding window approach, resulting in a series of correlation matrices.

We reasoned that specific patterns of global brain communication and their dynamic change are an absolute requirement for the emergence of consciousness. In order to identify specific patterns of global brain communication, we employed a k-means algorithm on covariance matrices. The change of covariance matrices over time was regarded as the dynamics of global brain communication (**Figure 1**).

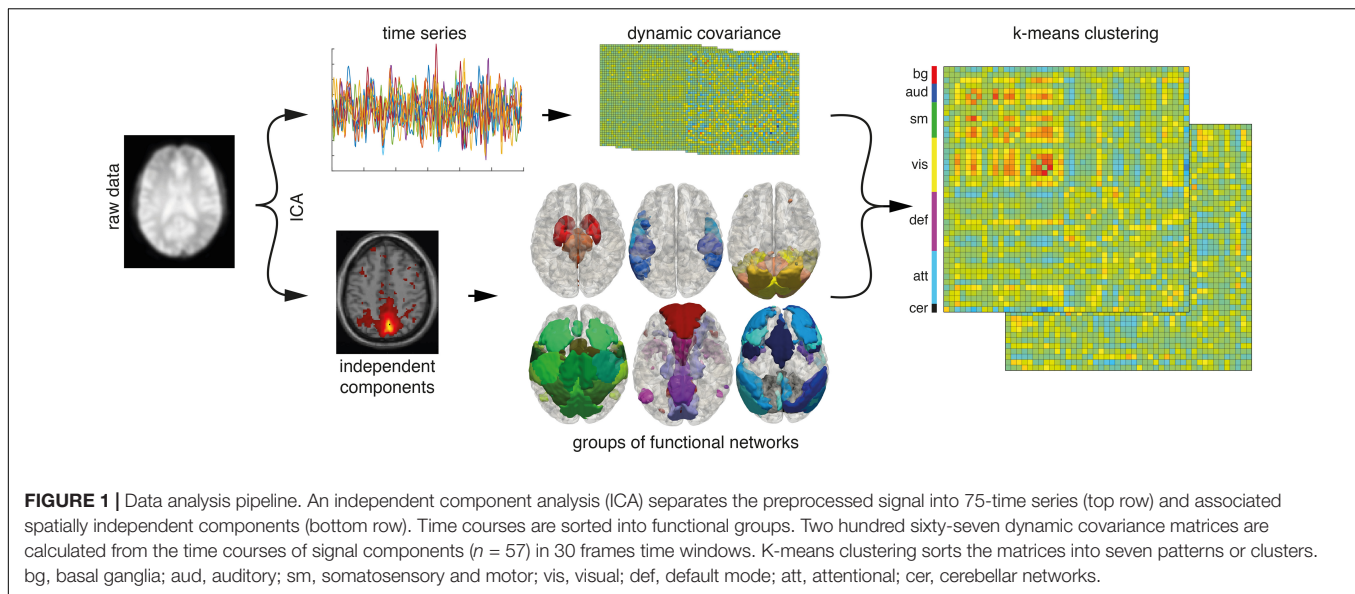
## MATERIALS AND METHODS

### Ethics Statement

The study was conducted in accordance with the Declaration of Helsinki, and the protocol of the study was approved by the ethics committee of the medical school of the Technische Universität München (München, Germany). The volunteers or the legal surrogates of the enrolled patients were given detailed information about the procedures and potential risks prior to written informed consent. Medical history was taken from every applicant, especially inquiring contraindications for an examination in a magnetic resonance scanner. Data were acquired between April 2013 and December 2019.

### Participants

Eighteen patients (eight male and 10 female, aged 36–84 years) were acquired through intensive care units. Inclusion criteria were patients with unresponsive wakefulness syndrome (UWS) through any brain damage, i.e., traumatic ( $n = 3$ ), stroke ( $n = 13$ ), anoxia ( $n = 1$ ), and toxic/metabolic ( $n = 1$ ). Under stroke, we summarized ischemic stroke as well as subarachnoid hemorrhage and intracerebral hemorrhage with no traumatic cause. We included all types of brain injuries for two main reasons: (1) we wanted to generalize the approach, and (2) we aimed to define the AWAKE state by specific brain activity and the UWS state by the absence of this activity. Based on this, we regarded the exact brain pathology as secondary. Any preexisting medication was left unchanged for the measurement accepting oral sedatives (e.g., low-dose benzodiazepines or neuroleptics). Intravenous sedatives were not applied. The subjects fasted for 6 h prior to the measurement. The patients with UWS were rated using the Coma Recovery Scale Revisited (CRS-R) at least three times (the day before the measurement, on the day of measurement, and in one follow-up) by an experienced medical doctor. All the patients had UWS on the day of measurement. All the other volunteers [ $n(\text{propofol}) = 11$ ,  $n(\text{sevoflurane}) = 14$ ,  $n(\text{controls}) = 20$ ] were healthy, male volunteers (aged 20–36 years) and had no contraindication for general anesthesia using propofol or sevoflurane (American Society of Anesthesiology I). All the



subjects with contraindications for MRI or positron emission tomography (PET) were excluded. We included unmatched healthy controls for the patients with UWS recorded on the same PET scanner that was also used for the UWS experiments. Because of the complex nature of the experiment (e.g., application of a PET contrast agent), we were unable to record a group of age-matched patients with cerebral pathology during the AWAKE state. This approach might introduce a bias since group differences might be affected by structural brain alterations or age differences.

## Experimental Protocol

### Controls and UWS Patients

For AWAKE and UWS resting-state measurement, the participants lay supine in the scanner and were told not to fall asleep while keeping their eyes closed.

### Propofol Anesthesia

The first resting-state recording was acquired identical to the control subjects. In the following, this condition is referred to as AWAKE. The measurement during propofol-induced loss of responsiveness (PROP) was carried out using a target-controlled infusion pump (Open TCI; Space infusion system; Braun Medical, Melsungen, Germany). Propofol concentration was increased by 0.4  $\mu\text{g/ml}$  steps beginning at 1.2  $\mu\text{g/ml}$  until volunteers stopped responding to the verbal command “squeeze my hand” (equivalent to a Ramsay sedation scale score of 5–6). The concentration was then kept stable for the remaining fMRI measurement. This point was reached at plasma concentrations of  $2.97 \pm 0.47 \mu\text{g/ml}$  (mean  $\pm$  SD). Ten minutes of equilibration time were waited before the actual measurement took place. Throughout this article, this state is referred to as PROP. Details of the propofol anesthesia protocol can be found in Jordan et al. (2013).

### Sevoflurane Anesthesia

The resting-state was acquired identical to the propofol setting. Image acquisition during the sevoflurane-induced loss of responsiveness (in this article: SEVO) was carried out after intubation with a magnetic resonance tomograph-compatible laryngeal mask and during artificial ventilation using an anesthesia machine (Fabius Tiro, Dräger, Germany). Sevoflurane was kept stable at 2 volume percent end-tidal concentration during this condition. The subjects were unresponsive to the command “squeeze my hand” during this condition, tolerated the laryngeal mask well, and showed reduced movements when compared with the wakeful state. It is also noteworthy that clinically, this sedation was deeper when compared with PROP (corresponding to a Ramsay scale score of 6). All the participants were asked for any memories of the unresponsive state, and all reported amnesia for the procedure. Details of the sevoflurane anesthesia protocol can be found in Ranft et al. (2016).

### Data Acquisition

Data from the patients with UWS ( $n = 18$ ) and healthy controls ( $n = 20$ ) were acquired using an integrated Siemens Biograph mMR scanner (Siemens Medical Solutions, Erlangen, Germany) capable of registering concurrent positron emission tomography (PET) and MRI from 3 T data using the vendor-supplied 12-channel phase-array head coil. PET volume, two T2\*-weighted echo-planar imaging (EPI) MRI, and magnetization-prepared rapid acquisition gradient echo (MPRAGE) T1-weighted anatomic volume data were recorded. Scanning parameters of the EPI included repetition time (TR)/echo time (TE)/flip angle, 2,000 ms/30 ms/90°; 35 slices (gap 0.6 mm) aligned to the anterior-posterior commissure covering the whole brain; field of view (FoV), 192 mm; matrix size,  $64 \times 64$ ; and voxel size, 3 mm  $\times$  3 mm  $\times$  3.0 mm. Each of the two measurements consisted of 300 acquisitions in interleaved mode, with a total scan time of 10 min and 8 s. Scanning

parameters for the MPRAGE sequence included TR/TE/flip angle, 2,300 ms/2.98 ms/9°; 160 slices (gap, 0.5 mm) covering the whole brain; FoV, 256 mm; matrix size, 256 × 256; and voxel size, 1 mm × 1 mm × 1 mm. The total scan time was 5 min and 3 s. The PET data were not further included in this study.

Both the propofol and sevoflurane image acquisitions were carried out using a 3-T whole-body magnetic resonance tomographic scanner (Achieva Quasar Dual 3.0T 16CH; Philips, Medical Systems International Inc., Best, Netherlands) employing an eight-channel, phased-array head coil. fMRI was performed with a gradient EPI sequence (echo time = 30 ms, repetition time = 1.838 ms, flip angle = 75°, field of view = 220 mm × 220 mm, matrix = 72 × 72, 32 slices, slice thickness = 3 mm, and 1 mm interslice gap; 300 volumes were acquired in the propofol cohort and 350 volumes were acquired in the sevoflurane cohort. Of these 350 frames, the last 50 were discarded). Anatomy was acquired before the functional scan using a single T1-weighted sequence and 1 mm × 1 mm × 1 mm voxel size per subject. Only data sets that not exceeded 2 mm of translation in either z-, y-, or z-direction were included in the subsequent data analysis.

## fMRI Data Analysis

DPARSF 4.0 (Chao-Gan and Yu-Feng, 2010) and SPM12 (Friston, 2007) were used for preprocessing. Functional and anatomical images were realigned manually along the AC-PC plane. The first three time points were removed, and slice timing was corrected. Images were segmented and normalized using DARTEL to a voxel size of 2 mm × 2 mm × 2 mm. Functional images were smoothed with a 4 mm full-width at half-maximum Gaussian kernel. A 2 mm × 2 mm × 2 mm-voxel-size template was created from a standard epi template in SPM12 ("EPI.nii"). Both the anatomical and functional images were co-registered to this template. Six movement parameters (x-, y-, and z-translation and the corresponding rotations) and their first derivatives were regressed out using a general linear model. The time series were de-spiked after the regression and band-pass filtered between 0.01 and 1 Hz. No other global signal removal or scrubbing of the frame was performed.

In the first step, the preprocessed fMRI data were analyzed using a group-level spatial iICA (applied on the whole data set, namely AWAKE, PROP, SEVO, and UWS) as implemented in the Gift toolbox (version 4.0b, Group ICA/IVA of fMRI Toolbox; Georgia State University, Atlanta, GA, United States) and GICA3 back reconstruction. The INFOMAX algorithm was employed together with an ICASSO and 20 repetitions to decompose the data into 75 spatially independent components. We used the GICA3 back reconstruction to generate individual spatial maps and associated (potentially correlated) time courses for each individual map (Figure 1). We chose ICA because it was able to identify networks in altered brain anatomy robustly, and it represents the standard technique of dimensionality reduction/brain parcellation in altered states of consciousness [see, e.g., Demertzi et al. (2014)].

This resulted in two types of data: (1) spatially independent components for each subject that were z-standardized. We refer

to these maps as  $S_i$ ; and (2) one time course for each component and each subject. The time courses were also standardized to be standard normally distributed to account for different scanners and amplitudes of the raw blood oxygenation level-dependent signal. We refer to these time courses as  $R_i$ .

A signal from neuronal origin was assumed if the component projected on gray matter showed no similarity to venous vessels and if it showed a characteristic frequency spectrum with a clear peak below 0.1 Hz. Noise components usually showed a flat frequency spectrum similar to Gaussian noise. Fifty-seven functional networks were identified and included in the subsequent analysis. These networks were close to identical to the ones previously published (Golkowski et al., 2019).

The  $R_i$  time courses were despiked and low-pass filtered < 0.15 Hz (Allen et al., 2014; Golkowski et al., 2019). Correlation matrices were calculated in a sliding window manner as implemented in the Temporal dFNC toolbox of GIFT 4.0b using 30-time points from each measurement, resulting in 267 correlation matrices per measurement (which is 300 recorded volumes minus 3 discarded during preprocessing minus sliding window length). The size of the resulting matrices was 57 × 57 due to the symmetry of the correlation matrices they contained 57 × 56/2 independent data points. To make the correlation matrices more accessible for visual inspection, we sorted the functional networks into groups, namely, basal ganglia networks, auditory networks, somatosensory and motor networks, visual networks, default mode networks, attentional networks, and cerebellar networks. It is noteworthy that this sorting into functional groups has no effect on the later k-means clustering algorithm since the algorithms only use the L1 distance of the correlation matrices (i.e., the sum of the absolute differences) as a measure of distance and is, therefore, independent of the order of correlation values.

We used the k-means clustering algorithm also implemented in GIFT 4.0b with the L1-("Manhattan")-distance with 20 repetitions to assign the between-network connectivity matrices to between network-connectivity patterns. These patterns are not predefined and are generated by the k-means algorithm as clusters or patterns of the between-network connectivity matrices. The generally arbitrary number of patterns was defined to be 7 to make the results comparable to Allen et al. (2014) and Golkowski et al. (2019).

After variable ranking using the Fisher score, one to seven variables were chosen for a support vector machine. The aim was to test whether the information from the anesthesia experiments can be generalized. For training of the support vector machine algorithm, the 50 data sets from the anesthesia experiments were employed. Eighteen data sets from the patients with UWS and 20 controls were used for the test data set. We also conducted the support vector machine algorithm with a random permutation of groups of the test data set to illustrate chance level. These steps were carried out with FSLib (Version 5.1). All analysis steps were carried out in Matlab (R2016a; Mathworks, Natick, MA, United States).



## Statistical Analysis

All the group comparisons were conducted with the two-sided Mann–Whitney/rank sum test. *P*-values were corrected for multiple comparisons using Bonferroni's method (multiplication factor 6). Differences in pattern distribution between the groups (on the nominal scale of pattern numbers) were tested by using a chi-square test with Pearson's chi-square distance. Significance was assumed if  $p < 0.05$ .

Logistic regression was carried out in Matlab using the “mnrfilt” and logit transfer functions. The binary categorical responses of the appearance of patterns specific for consciousness (patterns 1 and 3) and all the other patterns were modeled on the time series of each individual network ( $n = 57$ ) after smoothing using the mean of 30 frames sliding window. This was the same slide window length from which the covariance matrices were calculated. The time series were transformed to be standard normally distributed in each recording to remove variance or amplitude effects. Significance was assumed if  $p < 0.05$ .

## RESULTS

### Baseline Characteristics

Aiming to transfer findings from the model of anesthesia-induced unconsciousness to disorders of consciousness, we divided the data into a training data set, namely, subjects during wakefulness and general anesthesia, and a test data set, namely, patients with UWS and healthy controls. The training data set undergoing general anesthesia comprised 11 healthy volunteers during wakefulness and propofol-induced general anesthesia (PROP) and of 14 healthy volunteers during wakefulness and sevoflurane-induced general anesthesia (SEVO). The test data set comprised 20 healthy volunteers and 18 patients with UWS (see **Table 1** for baseline characteristics). The one patient with 9 points in the CRS-R fulfilled no criteria for being minimally conscious on the day of measurement but showed eye fixation the day before. We decided to include this patient because the measurement had a sufficient data quality, and we saw a higher chance to find patterns specific for intact consciousness in this patient. We do not think that this introduced a bias to data analysis. Four patients were more than 100 days after brain injury, and 14 patients were included in the study during the acute hospitalization with less than 42 days after brain injury.

### Pattern Separation

To test if patterns of global brain communication are specific for consciousness, we conducted an ICA with 75 components using the complete data set and separated functional brain networks ( $n = 57$ ) from noise sources ( $n = 18$ ). For each network, we then calculated the associated covariance matrices in 30-frame time window each (**Figure 1**). This resulted in 267 covariance matrices per recording. The k-means clustering of all the resulting matrices into seven clusters showed that the various groups had a significantly different distribution across distinct patterns of global communication ( $p < 0.001$  for AWAKE vs. PROP, SEVO and UWS, chi2test, uncorrected, **Figure 2A**). Patterns 1 and 3 were almost exclusively assumed by the AWAKE

**TABLE 1** | Clinical features of the unresponsive, wakeful patients.

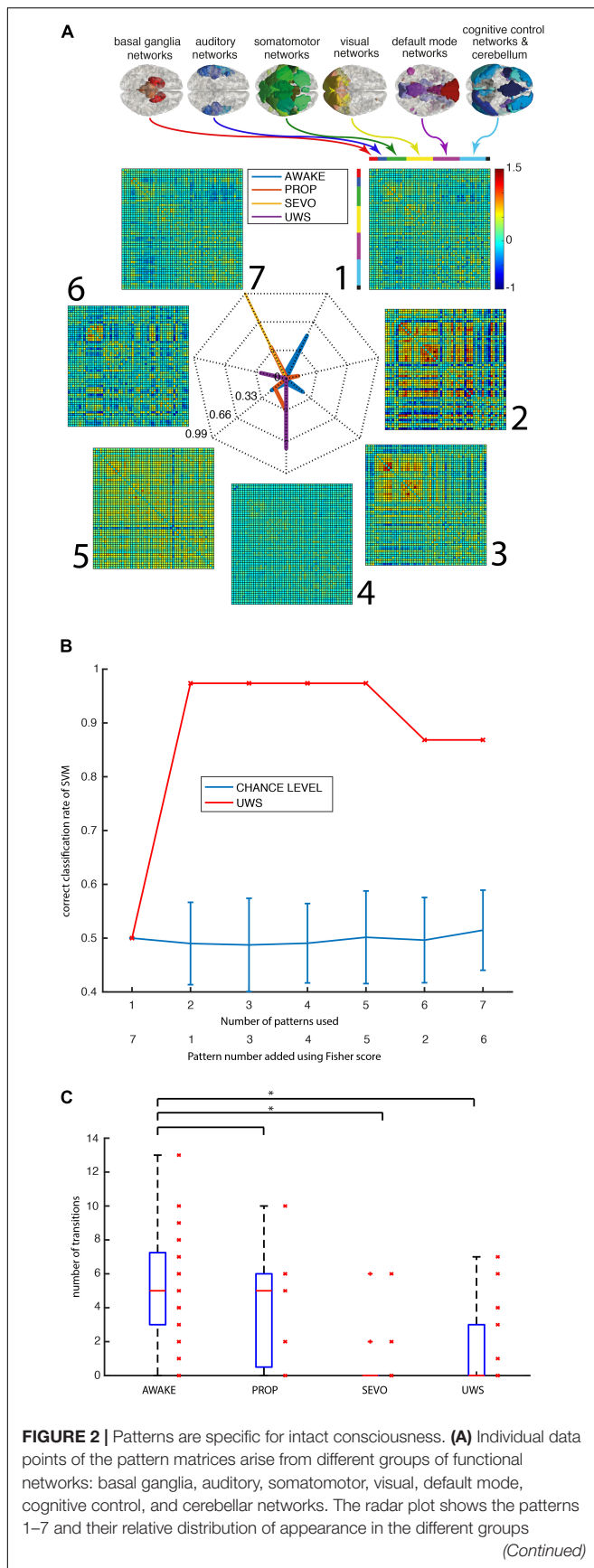
	Mean $\pm$ SD	Range
Age, y	62 $\pm$ 14	36–84
Time since injury, d	62 $\pm$ 60	18–206
CRS-R on day of measurement	4.1 $\pm$ 1.9	2–9
Sex	10 female	8 male
Etiology, number		
Stroke	13	
TBI	3	
Anoxia	1	
Metabolic	1	

subjects (probability: pattern 1:0.5, pattern 3:0.22) and rarely during PROP (probability: pattern 1:0.015, pattern 3:0.001). Patterns 1 and 3 did not occur in UWS and SEVO subjects. The distribution of patterns showed an overlap between the SEVO group, which was almost exclusive in pattern 7 (probability 0.99), and the PROP group (probability for pattern 7:0.37). The latter group also showed an overlap with AWAKE subjects (mainly pattern 5) and subjects with UWS (pattern 4). Pattern 2 was only observed during PROP (probability 0.13), and pattern 6 exclusively appeared in the UWS (probability 0.26) group. In summary, this clustering showed that patterns 1 and 3 are specific for the AWAKE group.

To test if the appearance of specific connectivity patterns is sufficient to distinguish consciousness from unconsciousness, we separated the data into one training data set containing the subjects from PROP ( $n = 11$ ), SEVO ( $n = 14$ ), and AWAKE ( $n = 25$ ) and one test data set containing the UWS ( $n = 18$ ) and AWAKE subjects from the same MRI scanner ( $n = 20$ , these subjects were not included in the training data set). We classified the test data set employing a support vector machine and prior Fisher scoring for feature selection (**Figure 2B**). We composed the groups this way because general anesthesia, unlike many types of disorders of consciousness, is a well-controlled, reversible form of unconsciousness with intact brain anatomy. As expected from the k-means clustering, the inclusion of only one pattern (pattern 7) resulted in a chance level result, because this pattern was only encountered in the SEVO and PROP groups. Combining patterns 1 and 7 resulted in a correct classification rate of 0.97. The one misclassified data set was from the AWAKE group. Precision remained constant upon further addition of features and then declined above 6 features. We concluded that knowledge about the subjects being either in the PROP, SEVO, or AWAKE group allowed a precise prediction of whether patients were in the UWS or the AWAKE group.

### Pattern Dynamics

Next, we tested if decreased dynamics of connectivity patterns are a feature of diminished consciousness (**Figure 2C**) by analyzing the absolute number of transitions between the different patterns of global brain communication as the parameter. The difference was significant for SEVO and UWS vs. controls ( $p < 0.05$ , Mann–Whitney test, uncorrected) but not for PROP vs. controls. We also observed a relevant overlap between

**FIGURE 2 |** (Continued)

(AWAKE in blue, PROP in red, SEVO in yellow, and UWS in purple).

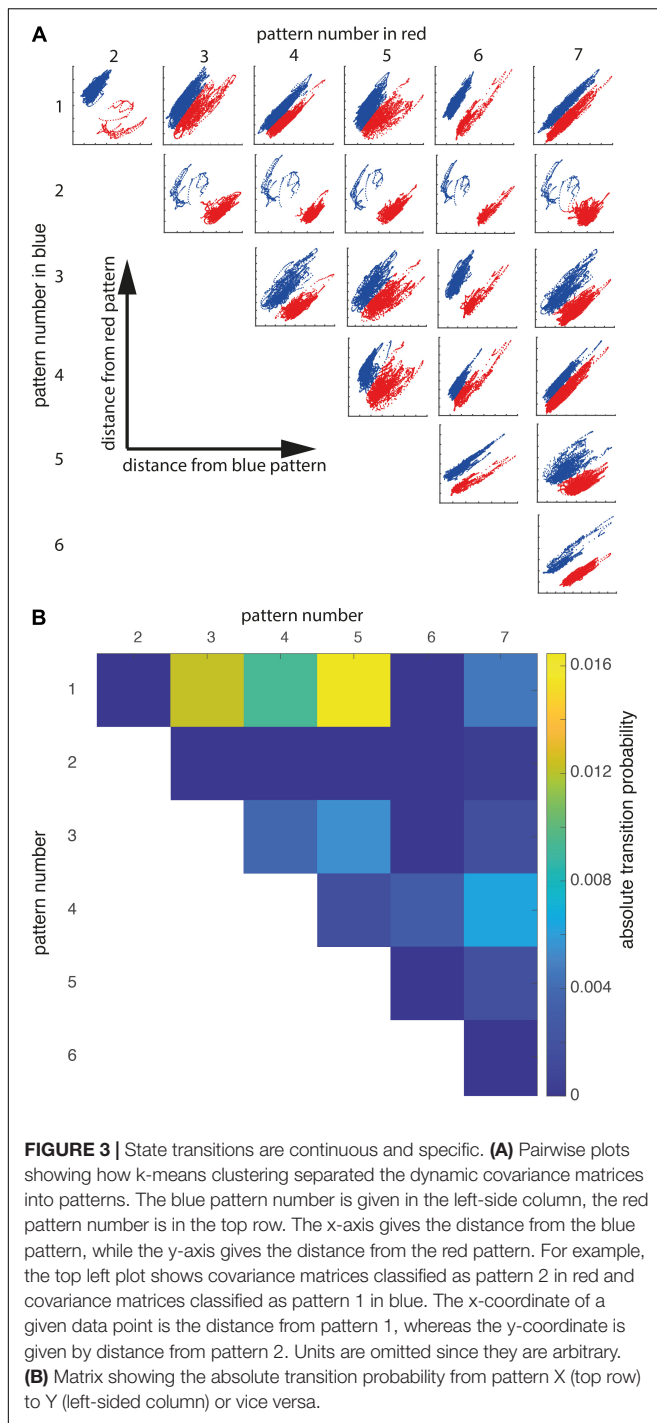
**(B)** Correct classification rate of a support vector machine trained in anesthesia data and tested on the subjects with UWS ( $n = 19$ ) and 19 controls (red) vs. chance level in a random assignment of groups and 100 repetitions (blue with standard errors). **(C)** An absolute number of transitions between patterns in the four groups. (\* $p < 0.05$ , Mann-Whitney test, uncorrected).

the groups and the span of values in the AWAKE group was (0, 13), thus covering the whole range of encountered values in all the groups. Consequently, the dynamics of global brain communication could not be used as a specific predictor of consciousness.

The distribution of patterns of global brain communication across groups and their dynamics had two further implications: (1) the AWAKE subjects can be both in patterns specific for AWAKE and patterns that can also be seen in the other groups, and (2) the absolute number of transitions is not specific for AWAKE, while the ability to transition to certain patterns is. To better understand when these group-specific transitions happen, we considered L1 distances (the sum of the absolute difference) of single covariance matrices with respect to a given pattern. Since the resulting space had seven dimensions (one dimension for each distance of covariance matrix to one of the seven clusters), we visualized the data in a pairwise manner (Figure 3A). This resulted in 21 plots, giving the L1 distances of covariance matrices from a pair of patterns and the pattern they were classified as. This visualization demonstrated that there are different types of separations into patterns: patterns separated visually by a gap and patterns with a continuous transition. For example, the AWAKE-specific pattern 1 shows a continuous transition with patterns 3, 4, and 5. In contrast, pattern 2 only exhibits this type of transition with pattern 7, two patterns that were only encountered during general anesthesia. In summary, the plots show how subjects within a group move rather continuously from one pattern to another than in a binary manner.

This observation was further supported by the transition probability matrix (Figure 3B). The probability to change between patterns of global brain communication that show a gap in Figure 3A was zero or close to zero, while pattern separation without gap showed frequent transitions. Notably, patterns observed in the AWAKE group showed the highest probability of transitions. Hence, the absolute number of transitions was not a characteristic for consciousness but the ability to transition into specific patterns of global brain communication.

Aiming to identify networks driving these patterns associated with consciousness, we conducted a logistic regression with patterns associated with consciousness, namely patterns 1 and 3, as the response variable and the time courses of all networks as predictors. This regression showed that 6 out of 56 networks had a significant positive slope with respect to the appearance of conscious-related patterns. These networks encompassed sensory networks: the bilateral primary auditory cortex ( $p < 0.05$ ,  $B = 0.31$ ), the bilateral calcarine gyrus ( $p < 0.05$ ,  $B = 0.28$ ), and the bilateral fusiform gyrus ( $p < 0.05$ ,



$B = 0.32$ ), as well as networks of the frontal lobe: the bilateral dorsomedial prefrontal cortex ( $p < 0.05$ ,  $B = 0.37$ ), the bilateral dorsolateral prefrontal cortex 1 ( $p < 0.001$ ,  $B = 0.49$ ), and the bilateral dorsolateral prefrontal cortex 2 ( $p < 0.05$ ,  $B = 0.33$ , see **Figure 4**). Thus, primary auditory and visual cortices as well as dorsolateral and dorsomedial prefrontal areas' activity is positively related to the appearance of patterns associated with consciousness.

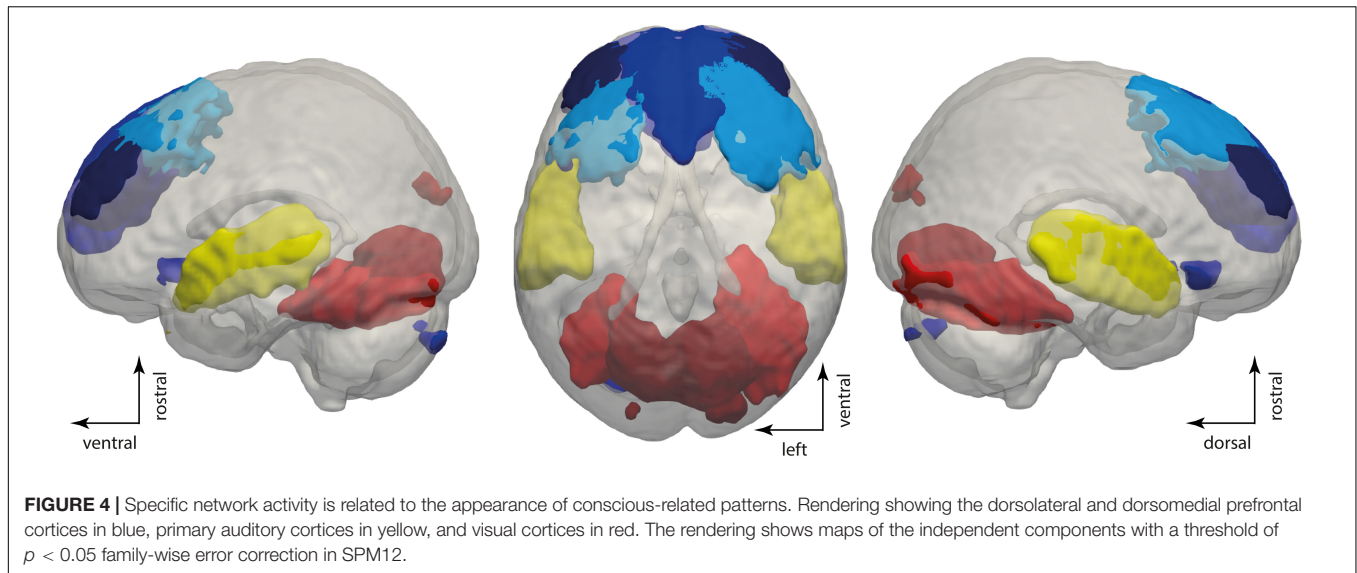
## DISCUSSION

In summary, our data analysis demonstrates that patterns of global brain communication specific for consciousness exist in the model of general anesthesia, irrespective of anesthetic agent. The patterns specific for AWAKE were characterized by either a high correlation across various hierarchical levels of sensory and motor systems (pattern 3) or by a state of low overall correlation (pattern 1). This finding could be translated to patients during unresponsive wakefulness and controls, where the same patterns reliably identified intact consciousness and were absent in UWS. One main difference between the model of general anesthesia and patients with UWS is certainly the altered brain anatomy. The UWS cohort, however, also included subjects with intact brain anatomy and only subcortical pathology. The findings were also applicable to these individuals. In addition, the ICA is largely independent of the brain anatomy and was able to identify brain networks outside of lesioned brain tissue. We regard this relative independence of the data analysis pipeline from anatomical alterations as an advantage over atlas-based brain parcellations, allowing the future automatization of this approach. In summary, we hope that the direct identification of conscious information processing might be superior when compared with current state-of-the-art approaches using signal complexity (Casarotto et al., 2016) or machine learning of large data sets of various states of consciousness (Demertzi et al., 2015) in the future. Theoretically, the direct measurement of conscious-specific brain activity might be more sensitive toward intact consciousness than surrogate markers that are statistically linked to the clinical state.

Both the models of general anesthesia and UWS showed reduced dynamics between patterns when compared with the AWAKE state. However, we encountered a relevant overlap between the different groups. This result demonstrates that the AWAKE state is much more characterized by continuous transitions between patterns specific and not specific for AWAKE than by the mere number of transitions. How such patterns arise in the brain has been investigated by computational models showing that specific patterns and their dynamics arise from the connectome of the human brain only if the brain works at a "critical" level (Haimovici et al., 2013; Deco et al., 2017), a tight balance between excitation and inhibition. In addition, both focal lesioning (Alstott et al., 2009; Gratton et al., 2012) and pharmacological manipulation (Tagliazucchi et al., 2016; Lee et al., 2019) were predicted to lead the activity of the brain away from this level. These theoretical considerations also gained experimental support in animals (Ma et al., 2017) and humans during general anesthesia with different anesthetic agents (Zhang et al., 2018; Golkowski et al., 2019). Our data illustrate that both specific patterns and their dynamic change are abolished during unconsciousness irrespective of the exact cause. Thus, this study is in line with these theoretical properties of brain organization and previous experiments further bridging the gap between theory and the subjective experience of being conscious.

The activity within the dorsal prefrontal cortical areas, primary sensory networks, was also linked to the appearance conscious-related patterns of global brain communication. This observation is not only a mere reproduction of previous findings





on the role of the dorsal attention system in consciousness (Ranft et al., 2016; Golkowski et al., 2019) but also highlights that the activity in the dorsal attention system (Mashour et al., 2020) is temporally related to patterns of global brain communication specifically for the AWAKE state in both the model general anesthesia and in the clinical setting.

### Limitations of the Study

In our study, the control cohort for the test data set consisted of healthy volunteers. The separation of the two groups is presumably easier than between a group of wakeful patients with severely damaged brains and our UWS group, a setting in clinical practice where this approach has to prove whether it actually delivers additional information to the clinical evaluation, i.e., is able to identify intact consciousness in patients with UWS in a real-world setting.

In addition, the presented data analysis is very complex and not yet fully automatized. For an application in clinical routine, the analysis pipeline needs to be implemented in an automatized toolbox and made available to the general public.

The limited number of subjects in each group is certainly another limiting factor. For the anesthesia experiments, it was already difficult to obtain ethics committee approval for a limited number of subjects. For the patients with UWS, it took several years to obtain sufficient data sets of adequate quality in this single-center study.

These two major flaws should be addressed in a future study equipped with an easily available analysis tool and a comprehensive clinical study design.

### Conclusion

In summary, the findings of this study show that complex patterns of global brain communication are capable to separate consciousness from unconsciousness during both general anesthesia and disorders of consciousness. Thus, this approach might not only reveal intact consciousness in seemingly wakeful

but unresponsive patients but also give an insight into the neurophysiological basis of consciousness itself.

### DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, upon reasonable request.

### ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of the Technical University Munich. The patients/participants provided their written informed consent to participate in this study.

### AUTHOR CONTRIBUTIONS

DG designed and conducted the experiments, conducted the data analysis, and wrote the manuscript. RW and JR conducted the experiments. AR, DJ, and RI designed and conducted the experiments and helped in writing the manuscript. GS helped in writing the manuscript. All authors contributed to the article and approved the submitted version.

### FUNDING

This study was supported by the ZNS, Hannelore Kohl Stiftung, Bonn and by the Klinikum rechts der Isar of the Technical University of Munich (Grant No. 2017021).

### ACKNOWLEDGMENTS

We would like to thank Martin Golkowski for his help in writing manuscript.



## REFERENCES

- Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb Cortex* 24, 663–676. doi: 10.1093/cercor/bhs352
- Alstott, J., Breakspear, M., Hagmann, P., Cammoun, L., and Sporns, O. (2009). Modeling the impact of lesions in the human brain. *PLoS Comput. Biol.* 5:e1000408. doi: 10.1371/journal.pcbi.1000408
- Barttfeld, P., Uhrig, L., Sitt, J. D., Sigman, M., Jarraya, B., and Dehaene, S. (2015). Signature of consciousness in the dynamics of resting-state brain activity. *Proc. Natl. Acad. Sci. U.S.A.* 112, 887–892. doi: 10.1073/pnas.1418031112
- Bonhomme, V., Vanhaudenhuyse, A., Demertzi, A., Bruno, M. A., Jaquet, O., Bahri, M. A., et al. (2016). Resting-state network-specific breakdown of functional connectivity during ketamine alteration of consciousness in volunteers. *Anesthesiology* 125, 873–888. doi: 10.1097/ALN.0000000000001275
- Boveroux, P., Vanhaudenhuyse, A., Bruno, M. A., Noirhomme, Q., Lauwick, S., Luxen, A., et al. (2010). Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. *Anesthesiology* 113, 1038–1053. doi: 10.1097/ALN.0b013e3181f697f5
- Brown, E. N., Lydic, R., and Schiff, N. D. (2010). General anesthesia, sleep, and coma. *N. Engl. J. Med.* 363, 2638–2650. doi: 10.1056/NEJMra0808281
- Casarotto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fecchio, M., Napolitani, M., et al. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Ann. Neurol.* 80, 718–729. doi: 10.1002/ana.24779
- Chao-Gan, Y., and Yu-Feng, Z. (2010). DPARSF: A MATLAB Toolbox for "Pipeline" Data Analysis of Resting-State fMRI. *Front. Syst. Neurosci.* 4:13. doi: 10.3389/fnsys.2010.00013
- Deco, G., Van Hartevelt, T. J., Fernandes, H. M., Stevner, A., and Kringelbach, M. L. (2017). The most relevant human brain regions for functional connectivity: evidence for a dynamical workspace of binding nodes from whole-brain computational modelling. *NeuroImage* 146, 197–210.
- Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018
- Dehaene, S., Lau, H., and Kouider, S. (2017). What is consciousness, and could machines have it? *Science* 358, 486–492. doi: 10.1126/science.aan8871
- Dehaene, S., Sergent, C., and Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8520–8525. doi: 10.1073/pnas.1332574100
- Demertzi, A., Antonopoulos, G., Heine, L., Voss, H. U., Crone, J. S., de Los Angeles, C., et al. (2015). Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. *Brain* 138, 2619–2631. doi: 10.1093/brain/awv169
- Demertzi, A., Gomez, F., Crone, J. S., Vanhaudenhuyse, A., Tshibanda, L., Noirhomme, Q., et al. (2014). Multiple fMRI system-level baseline connectivity is disrupted in patients with consciousness alterations. *Cortex* 52, 35–46. doi: 10.1016/j.cortex.2013.11.005
- Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Barttfeld, P., Raimondo, F., et al. (2019). Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Sci. Adv.* 5:eaat7603. doi: 10.1126/sciadv.aat7603
- Friston, K. J. (2007). *Statistical Parametric Mapping: the Analysis of Functional Brain Images*. Boston: Elsevier/Academic Press.
- Gili, T., Saxena, N., Diukova, A., Murphy, K., Hall, J. E., and Wise, R. G. (2013). The thalamus and brainstem act as key hubs in alterations of human brain network connectivity induced by mild propofol sedation. *J. Neurosci.* 33, 4024–4031. doi: 10.1523/JNEUROSCI.3480-12.2013
- Golkowski, D., Larroque, S. K., Vanhaudenhuyse, A., Plenevaux, A., Boly, M., Di Perri, C., et al. (2019). Changes in whole brain dynamics and connectivity patterns during sevoflurane- and propofol-induced unconsciousness identified by functional magnetic resonance imaging. *Anesthesiology* 130, 898–911. doi: 10.1097/ALN.0000000000002704
- Gratton, C., Nomura, E. M., Perez, F., and D'Esposito, M. (2012). Focal brain lesions to critical locations cause widespread disruption of the modular organization of the brain. *J. Cogn. Neurosci.* 24, 1275–1285. doi: 10.1162/jocn\_a\_00222
- Haimovici, A., Tagliazucchi, E., Balenzuela, P., and Chialvo, D. R. (2013). Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Phys. Rev. Lett.* 110:17. doi: 10.1103/PhysRevLett.110.178101
- He, B. J., Snyder, A. Z., Vincent, J. L., Epstein, A., Shulman, G. L., and Corbetta, M. (2007). Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. *Neuron* 53, 905–918.
- Hudetz, A. G., Liu, X., and Pillay, S. (2015). Dynamic repertoire of intrinsic brain states is reduced in propofol-induced unconsciousness. *Brain Connect* 5, 10–22. doi: 10.1089/brain.2014.0230
- Jordan, D., Ilg, R., Riedl, V., Schorer, A., Grimberg, S., Neufang, S., et al. (2013). Simultaneous electroencephalographic and functional magnetic resonance imaging indicate impaired cortical top-down processing in association with anesthetic-induced unconsciousness. *Anesthesiology* 119, 1031–1042. doi: 10.1097/ALN.0b013e3182a7ca92
- Lee, H., Golkowski, D., Jordan, D., Berger, S., Ilg, R., Lee, J., et al. (2019). Relationship of critical dynamics, functional connectivity, and states of consciousness in large-scale human brain networks. *Neuroimage* 188, 228–238. doi: 10.1016/j.neuroimage.2018.12.011
- Luppi, A. I., Craig, M. M., Pappas, I., Finoia, P., Williams, G. B., Allanson, J., et al. (2019). Consciousness-specific dynamic interactions of brain integration and functional diversity. *Nat. Commun.* 10:4616. doi: 10.1038/s41467-019-12658-9
- Ma, Y., Hamilton, C., and Zhang, N. (2017). Dynamic Connectivity Patterns in Conscious and Unconscious Brain. *Brain Connect* 7, 1–12.
- Mashour, G. A., Roelfsema, P., Changeux, J. P., and Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron* 105, 776–798. doi: 10.1016/j.neuron.2020.01.026
- Ranft, A., Golkowski, D., Kiel, T., Riedl, V., Kohl, P., Rohrer, G., et al. (2016). Neural correlates of sevoflurane-induced unconsciousness identified by simultaneous functional magnetic resonance imaging and electroencephalography. *Anesthesiology* 125, 861–872. doi: 10.1097/ALN.0000000000001322
- Tagliazucchi, E., Chialvo, D. R., Siniatchkin, M., Amico, E., Brichant, J. F., Bonhomme, V., et al. (2016). Large-scale signatures of unconsciousness are consistent with a departure from critical dynamics. *J. R. Soc. Interf.* 13:20151027. doi: 10.1098/rsif.2015.1027
- van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., et al. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science* 360, 537–542. doi: 10.1126/science.aar7186
- Vanhaudenhuyse, A., Demertzi, A., Schabus, M., Noirhomme, Q., Bredart, S., Boly, M., et al. (2011). Two distinct neuronal networks mediate the awareness of environment and of self. *J. Cogn. Neurosci.* 23, 570–578. doi: 10.1162/jocn.2010.21488
- Vanhaudenhuyse, A., Noirhomme, Q., Tshibanda, L. J., Bruno, M. A., Boveroux, P., Schnakers, C., et al. (2010). Default network connectivity reflects the level of consciousness in non-communicative brain-damaged patients. *Brain* 133, 161–171. doi: 10.1093/brain/awp313
- Zhang, J., Huang, Z., Chen, Y., Zhang, J., Ghinda, D., Nikolova, Y., et al. (2018). Breakdown in the temporal and spatial organization of spontaneous brain activity during general anesthesia. *Hum. Brain Mapp.* 39, 2035–2046. doi: 10.1002/hbm.23984

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Golkowski, Willnecker, Rösler, Ranft, Schneider, Jordan and Ilg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Criticality Creates a Functional Platform for Network Transitions Between Internal and External Processing Modes in the Human Brain

Minkyung Kim<sup>1,2†</sup>, Hyoungkyu Kim<sup>1,2†</sup>, Zirui Huang<sup>1,2</sup>, George A. Mashour<sup>1,2,3</sup>, Denis Jordan<sup>4,5,6</sup>, Rüdiger Ilg<sup>4,5,6</sup> and UnCheol Lee<sup>1,2\*</sup>

<sup>1</sup> Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, MI, United States, <sup>2</sup> Center for Consciousness Science, University of Michigan Medical School, Ann Arbor, MI, United States, <sup>3</sup> Neuroscience Graduate Program, University of Michigan, Ann Arbor, MI, United States, <sup>4</sup> Applied Mathematics and Statistics, University of Applied Sciences Northwestern Switzerland, Muttenz, Switzerland, <sup>5</sup> Department of Anesthesiology, Klinikum rechts der Isar, Technische Universität München, Munich, Germany, <sup>6</sup> Department of Neurology, Klinikum rechts der Isar, Technische Universität München, Munich, Germany

## OPEN ACCESS

### Edited by:

Maurizio Mattia,  
National Institute of Health (ISS), Italy

### Reviewed by:

Benjamin R. Pittman-Polletta,  
Boston University, United States  
Andrew Rich McKinstry-Wu,  
University of Pennsylvania,  
United States

### \*Correspondence:

UnCheol Lee  
uclee@med.umich.edu

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

**Received:** 24 January 2021

**Accepted:** 29 October 2021

**Published:** 25 November 2021

### Citation:

Kim M, Kim H, Huang Z,  
Mashour GA, Jordan D, Ilg R and  
Lee U (2021) Criticality Creates  
a Functional Platform for Network  
Transitions Between Internal  
and External Processing Modes  
in the Human Brain.  
*Front. Syst. Neurosci.* 15:657809.  
doi: 10.3389/fnsys.2021.657809

Continuous switching between internal and external modes in the brain appears important for generating models of the self and the world. However, how the brain transitions between these two modes remains unknown. We propose that a large synchronization fluctuation of brain networks, emerging only near criticality (i.e., a balanced state between order and disorder), spontaneously creates temporal windows with distinct preferences for integrating the network's internal information or for processing external stimuli. Using a computational model, electroencephalography (EEG) analysis, and functional magnetic resonance imaging (fMRI) analysis during alterations of consciousness in humans, we report that synchronized and incoherent networks, respectively, bias toward internal and external information with specific network configurations. In the brain network model and EEG-based network, the network preferences are the most prominent at criticality and in conscious states associated with the bandwidth 4–12 Hz, with alternating functional network configurations. However, these network configurations are selectively disrupted in different states of consciousness such as general anesthesia, psychedelic states, minimally conscious states, and unresponsive wakefulness syndrome. The network preference for internal information integration is only significant in conscious states and psychedelic states, but not in other unconscious states, suggesting the importance of internal information integration in maintaining consciousness. The fMRI co-activation pattern analysis shows that functional networks that are sensitive to external stimuli—such as default mode, dorsal attentional, and frontoparietal networks—are activated in incoherent states, while insensitive networks, such as global activation and deactivation networks, are dominated in highly synchronized states. We suggest that criticality produces a functional platform for the brain's capability for continuous switching between two modes, which is crucial for the emergence of consciousness.

**Keywords:** criticality, consciousness, oscillator model, EEG, fMRI, brain network transition, information processing

## INTRODUCTION

Continuous switching between internal and external modes allows neural circuits to identify the contrast between different sources of information and reduce the mismatch between them (Hinton and McClelland, 1987; O'Reilly, 1996). Continuous switching between two modes has been considered a functional basis at a system level for constructing inner models in the brain that support perception, prediction, and action in the external world (Hasselmo, 1995; Honey et al., 2017). However, the origin of such modes in the brain, the mechanism by which they transition, and whether these transitions represent a necessary process for supporting higher-order brain functions are unknown. In our previous computational model study, we demonstrated that the brain's responsiveness to external stimuli depends on the level of global brain network synchronization, and this dependence only emerges near a critical state (Kim and Lee, 2020). In this study, we expanded the previous computational model study, suggesting that a large synchronization fluctuation emerging near a critical state may produce a functional platform upon which functional brain networks fluctuate between two distinct modes, one of which is conducive to the integration of internal information in the network and the other of which is highly susceptible to external stimuli. Such distinct preferences for internal or external information originate from the general property of the network's responsiveness to the synchronization fluctuation. We also analyzed both high-density electroencephalogram (EEG) and functional magnetic resonance imaging (fMRI) data of various states of consciousness (conscious, anesthetized, psychedelic, and pathological) to investigate how the network's preference for internal or external information in the time domain is associated with different states of consciousness.

Recent computational modeling and empirical studies suggest that consciousness occurs when brain dynamics are near criticality (i.e., poised at the border of multiple states) and that losing criticality (i.e., after a transition to one of the possible states) is related to various altered states of consciousness (Kitzbichler et al., 2009; Tagliazucchi et al., 2012; Haimovici et al., 2013; Muñoz, 2018; Kim and Lee, 2019). Critical dynamics confer biological advantages that may establish a functional foundation for the emergence of consciousness: an optimal balance between stability and instability, optimal computational capability, flexibility to adapt to a changing environment, and wide repertoires of brain states (Beggs, 2008; Cocchi et al., 2017). In both biological and non-biological systems, a large global fluctuation is one of the most representative signal characteristics of criticality, with an increase in autocorrelation (Scheffer et al., 2009; Van De Leemput et al., 2014). In our previous brain network modeling study, we found that a large synchronization fluctuation near a critical state is associated with a highly variable brain sensitivity to external stimuli (Kim and Lee, 2020). Specifically, low and high levels of synchronization in the brain network, respectively, provide susceptible and refractory time windows to the stimuli. In addition, it has been suggested that the level of neural synchronization reflects the brain's capability for information transmission and integration

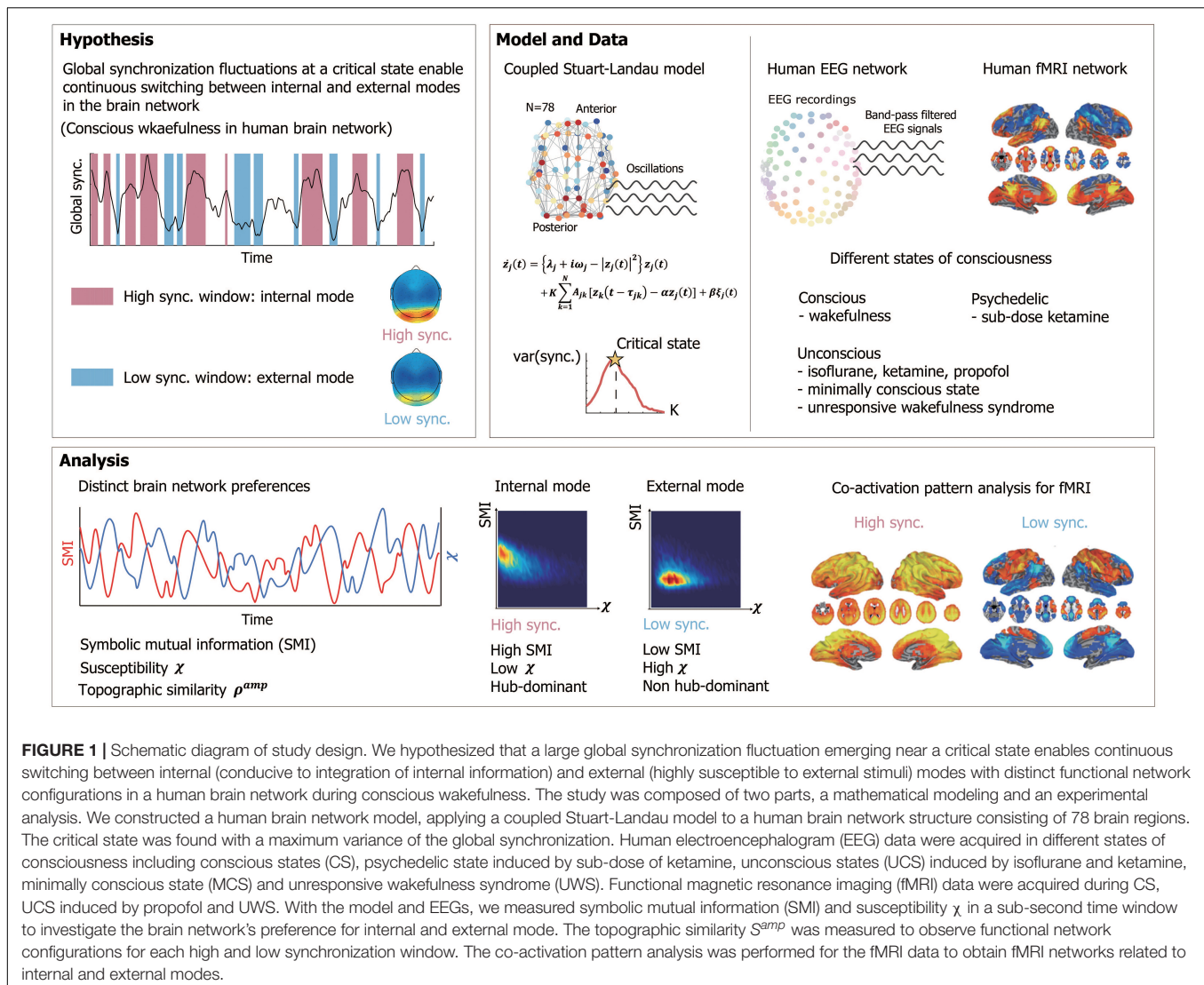
across the cortex (Fries, 2005; Bastos et al., 2015). Based on these findings, we hypothesize that the functional brain network may be highly susceptible to external stimuli but less internally integrative at low levels of synchronization. Conversely, at high levels of synchronization, the brain network favors information integration within the network but is less susceptible to external stimuli. Thus, we hypothesized that the brain network at high and low levels of network synchronization possesses distinct preferences for, respectively, internal and external information, which may induce the internal and external modes of the brain in the time domain. Such distinct preferences may be the most significant in conscious states (i.e., near criticality) with the maximal difference between high and low levels of synchronization. In contrast, these preferences may be mitigated in different states of consciousness (i.e., sub- or supercritical states) with reduced synchronization fluctuation.

To test this hypothesis, we used a large-scale human brain network model, modulating criticality and examining whether brain networks at high and low levels of synchronization near criticality have distinct preferences for the internal information of the network and external stimuli, respectively. Symbolic mutual information (SMI) (King et al., 2013) and network susceptibility (Yoon et al., 2015) were measured for each sub-second temporal window of simulated brain signals to evaluate quantitatively the brain network's preference. The modeling results were tested in humans with high-density EEG during various states of consciousness: conscious wakefulness, anesthetized (with isoflurane and ketamine), psychedelic (with a subanesthetic-dose of ketamine), and pathological conditions such as minimally conscious states and unresponsive wakefulness syndrome. We also examined functional brain network configurations at high and low levels of synchronization that systematically contribute to those preferences. In addition, we analyzed the fMRI data during wakefulness, propofol-induced unconscious states, and unresponsiveness wakefulness syndrome to examine whether the results observed in the model and EEG data were also consistent with what we observed in the fMRI functional networks. We investigated whether the fMRI functional networks that predominate at high or low levels of synchronization are relevant to the well-known networks presumably associated with internal or external information processing. The schematic diagram of the study is presented in **Figure 1**. The figures from the fMRI data analysis were adapted from Huang et al. (2020).

## MATERIALS AND METHODS

### Simulation of Spontaneous Neural Oscillations

A coupled Stuart-Landau model with human brain network structure has been widely used to simulate the oscillatory dynamics from various types of imaging modalities including EEG (Kim and Lee, 2019), magnetoencephalography (MEG), and fMRI (Deco et al., 2017, 2018). Here we also used the coupled Stuart-Landau model to simulate the oscillatory dynamics of brain network and investigated relationships between network synchronization and variables associated brain network's



**FIGURE 1 |** Schematic diagram of study design. We hypothesized that a large global synchronization fluctuation emerging near a critical state enables continuous switching between internal (conducive to integration of internal information) and external (highly susceptible to external stimuli) modes with distinct functional network configurations in a human brain network during conscious wakefulness. The study was composed of two parts, a mathematical modeling and an experimental analysis. We constructed a human brain network model, applying a coupled Stuart-Landau model to a human brain network structure consisting of 78 brain regions. The critical state was found with a maximum variance of the global synchronization. Human electroencephalogram (EEG) data were acquired in different states of consciousness including conscious states (CS), psychedelic state induced by sub-dose of ketamine, unconscious states (UCS) induced by isoflurane and ketamine, minimally conscious state (MCS) and unresponsive wakefulness syndrome (UWS). Functional magnetic resonance imaging (fMRI) data were acquired during CS, UCS induced by propofol and UWS. With the model and EEGs, we measured symbolic mutual information (SMI) and susceptibility  $\chi$  in a sub-second time window to investigate the brain network's preference for internal and external mode. The topographic similarity  $S^{amp}$  was measured to observe functional network configurations for each high and low synchronization window. The co-activation pattern analysis was performed for the fMRI data to obtain fMRI networks related to internal and external modes.

preferences for integrating information internally in the brain network and receiving stimuli from the outside of the brain.

Spontaneous networked oscillations were generated using a coupled Stuart-Landau model in a group-averaged anatomical human brain network constructed from diffusion tensor imaging (DTI) of 78 nodes (Gong et al., 2009). The coupled Stuart-Landau model, composed of the  $N$  number of oscillators, is defined as the following:

$$\dot{z}_j(t) = \left\{ \lambda_j + i\omega_j - |z_j(t)|^2 \right\} z_j(t) + \sum_{k=1}^N A_{jk} K_{jk} \{ z_k(t - \tau_{jk}) - \alpha z_j(t) \} + \beta \xi_j(t)$$

where a complex variable  $z_j(t)$  represents the oscillatory dynamics of brain region  $j$  at time  $t$ .  $\omega_j$  is an initial angular natural frequency of oscillator  $j$ . We used Gaussian distribution for the natural frequency with a mean frequency of 10 Hz and a standard deviation of 0.5 Hz to simulate the peak frequency

bandwidth of human EEG activity (Kim et al., 2018; Kim and Lee, 2020).  $A_{jk} = 1$  if oscillators  $j$  and  $k$  are connected, and  $A_{jk} = 0$  if they are not, based on the structural brain network.  $\tau_{jk}$  is a time delay between oscillators  $j$  and  $k$ , defined as  $D_{jk}/s$ , with the distance between brain regions  $j$  and  $k$ ,  $D_{jk}$ , and the average speed of axons in brain regions,  $s$ . Here we used  $s = 7ms$ . The node  $j$  receives input from connected node  $k$  after the time delay  $\tau_{jk}$ .  $\lambda_j$  and  $K_{jk}$  are a bifurcation parameter of oscillator  $j$  and a coupling strength between oscillators  $j$  and  $k$ , respectively. Modulating these parameters induce competition of independent behavior of oscillator and the coupling among the oscillators so that complex oscillatory dynamics differently emerge in different parameter regions. Each node shows supercritical Hopf bifurcation, and the dynamics of the oscillator settle on a limit cycle if  $\lambda_j > 0$ , and on a stable focus if  $\lambda_j < 0$ . We used a homogeneous bifurcation parameter  $\lambda_{jk} = \lambda$  from  $-3.2$  to  $3.2$  with  $\delta\lambda = 0.2$ , and a coupling strength  $K_{jk} = K$  from  $0$  to  $1$  with  $\delta K = 0.02$ . We additionally modulated a diffusive coupling parameter  $\alpha$ . The  $\alpha$  controls the degree of outgoing flow of node  $j$ . In neural



networks, two extreme values of  $\alpha$ , 0 and 1 indicate two types of synapses, chemical synapses, and gap junctions, respectively. We tested 0, 0.5, and 1 for the  $\alpha$ , and set  $\alpha = 0.5$  as the empirically well-fitted parameter. Further investigation for this parameter will be explored in the future study.  $\xi_j(t)$  is a Gaussian white noise for each node  $j$  and added to the dynamics with the standard deviation  $\beta = 0.05$ . We numerically solved differential equations of the Stuart-Landau model using the Stratonovich-Heun method with 1,000 discretization steps, then we resampled the data with 500 Hz. The last 60 s were used for the analysis of each simulation after 15 s of saturation periods. Therefore, spontaneous oscillatory dynamics were generated for each brain region at each  $\lambda$  and each  $K$ . We repeated this simulation with one hundred different initial frequency distributions to obtain statistical stability.

## Experimental Protocol and Electroencephalography Acquisition Isoflurane Anesthesia

The isoflurane data were collected from a cohort of 20 healthy volunteers (20–40 years) recorded at the University of Michigan, Ann Arbor, United States (Protocol #HUM0071578). The study has been reviewed by the Institutional Review Boards specializing in human subject research at University of Michigan. Written informed consent in accordance with the Declaration of Helsinki to participate in the study was obtained from all participants. Ten participants underwent general anesthesia. The participants in the anesthesia group initially received propofol at increasing infusion rates over three consecutive 5-min blocks (block 1: 100  $\mu\text{g/kg/min}$ , block 2: 200  $\mu\text{g/kg/min}$ , block 3: 300  $\mu\text{g/kg/min}$ ). Responsiveness was quantified every 30 s by the response to the verbal command “Squeeze your left/right hand twice” with random allocation to left/right. Isoflurane was then administrated with air and 40% oxygen at 1.3 age-adjusted minimum alveolar concentration. The isoflurane was administrated for 3 h and discontinued.

We analyzed EEG data of 9 out of 10 subjects who underwent general anesthesia due to the availability of high-density EEG data (128-channel HydroCel nets, Net Amps 400 amplifiers; Electrical Geodesic, Inc., United States). All EEG channels were referenced to a vertex with 500 Hz sampling frequency. EEG data of 4-min of eye-closed resting state before isoflurane administration (baseline) and 4-min of periods during general anesthesia (ISO) without burst-suppression were used in the current study. The data have been published with different analyses and hypotheses (Lee et al., 2019).

## Ketamine Anesthesia

The ketamine data were collected from 15 healthy volunteers (20–40 years) recorded at University of Michigan, Ann Arbor, with written informed consent to participate in the study. This study was approved by the University of Michigan Medical School Institutional Review Board, Ann Arbor, MI, United States (HUM00061087). The data have been published with different analyses (Lee et al., 2019). EEG data were acquired during a 5-min eyes-closed resting state before ketamine administration, subanesthetic ketamine infusion (0.5 mg/kg) over 40 min,

followed by 8 mg ondansetron, break for completion of questionnaire, anesthetic (1.5 mg/kg) bolus dose, and recovery period. EEG data of 4-min eyes-closed resting state (baseline), 4-min subanesthetic ketamine-induced state (PSY), and 4-min ketamine-induced unconsciousness (KET) after bolus anesthetic administration were used in the current study. The EEG data were acquired with 128-channel HydroCel nets, Net Amps 400 amplifiers (Electrical Geodesic Inc., United States). All channels were referenced to a vertex with 500 Hz sampling frequency.

## Disorders of Consciousness

Electroencephalography data were originally collected from a cohort of 80 patients with disorders of consciousness caused by ischemic stroke, intracerebral hemorrhage, subarachnoid hemorrhage, subdural hematoma, traumatic brain injury, meningitis, or hyperglycemic brain injury. Patients were diagnosed as minimally conscious or vegetative states/unresponsiveness wakefulness syndrome (UWS) using the Coma Recovery Scale-Revised (CRS-R). The CRS-R status was acquired again after 6 months of investigating the follow-up changes. The data from 17 subjects (the Munich cohort) were recorded on two different systems; 15 subjects were recorded with 64-channel, ring-type sintered, and nonmagnetic Ag/AgCl electrodes (Easycap, Herrsching, Germany); 2 subjects were recorded with 32-channel, nonmagnetic, and battery-operated electroencephalographic amplifiers (BrainAmp MR, Brain Products, Gilching, Germany). Both EEG data were recorded at 5 kHz sampling rates (BrainVision Recorder, Brain Products). The data from 63 patients (the Burgau cohort) were recorded with a 256-channel high-density Geodesic sensor net, a Net Amps 300 amplifier, and Net Station 4.5 software (Electrical Geodesic Inc., Eugene, OR, United States). The sampling rates were 1 kHz. All data were preprocessed to have 63-channel. In the current study, we analyzed the 4-min of the data from 9 subjects who were diagnosed as UWS without the evolution of CRS-R status and not showing suppression patterns and 16 subjects who were diagnosed as minimally conscious state (MCS).

In this study, we defined PSY as an altered state of consciousness (i.e., pharmacologically perturbed with consciousness maintained) and ISO, KET, MCS, and UWS as depressed states of consciousness (i.e., pharmacologically or pathologically perturbed with no conscious response).

## Electroencephalography Data Preprocessing

With three datasets, we analyzed 6 different states of consciousness, such as conscious state, (CS,  $n = 24$ ), ketamine-induced psychedelic state (PSY,  $n = 15$ ), isoflurane-induced unconsciousness (ISO,  $n = 9$ ), ketamine-induced unconsciousness (KET,  $n = 15$ ), MCS ( $n = 16$ ), and UWS ( $n = 9$ ). We selected 96-channel for the isoflurane and ketamine data and selected 56-channel for the MCS and UWS data that cover the scalp for the analysis. The MCS and UWS data were down-sampled to 250 Hz. After selecting the EEG channels, we high-pass filtered the data with 0.5 Hz cutoff frequency using MATLAB function “pop\_eegfiltnew.m” in the EEGLAB toolbox. We then removed noisy epochs using MATLAB

function “trimOutlier.m” with a standard deviation of channel amplitude rejection thresholds 100  $\mu V$ , amplitude rejection thresholds 300  $\mu V$ , and the range for rejection period 200-ms. Rejected channels were reconstructed using MATLAB function “pop\_interp.m” in the EEGLAB with a spherical interpolation method. All EEG data from isoflurane, ketamine-induced unconsciousness, and disorders of consciousness were re-referenced to the average of all channels. We band-pass filtered the data with a frequency range from 4 to 12 Hz to capture the properties of globally coupled oscillators. We also analyzed the data with other frequency bands (1–4 Hz, 1–20 Hz, 8–12 Hz, 13–30 Hz, and 30–42 Hz).

## Level of Network Synchronization and Pair Correlation Function

In this study, the level of network synchronization is an important factor to determine the brain network's preference for processing internal information of the network or external information from the outside of the brain network. The instantaneous level of network synchronization  $r(t)$  at time  $t$  was measured by the order parameter,

$$r(t) e^{i\psi(t)} = \frac{1}{N} \sum_{j=1}^N e^{i\theta_j(t)}$$

where  $\theta_j(t)$  is a phase of oscillator  $j$  and  $\psi(t)$  is the average global phase at time  $t$ . Here  $r(t)$  equals to 0 if phases of oscillators are uniformly distributed and 1 if all oscillators have the same phase. The  $r(t)$  was calculated for all parameter combinations in the model, all states from EEG, and fMRI data.

We then measured a pair correlation function  $PCF \equiv N \left[ \langle r^2(t) \rangle - \langle r(t) \rangle^2 \right]$ , which is a surrogate measure of criticality, to define the critical state (Yoon et al., 2015) in the model, under the assumption that the largest synchronization fluctuations are associated with the largest number of metastable states of brain network and should be shown at criticality. We measured PCF for all parameter combinations in the model, all states of EEG data.

## Temporal Window Classification Based on the Network Synchronization

To investigate the network's information processing preference with reliable time resolution, we classified the network transient states into high and low  $R$  temporal windows. We calculated an average network synchronization  $R$  as  $\langle r(t) \rangle_T$  for each temporal window, where  $T$  is the size of the temporal window. In the model, we set  $T = 250$  msec with 50 msec overlaps and classified the temporal windows into two different windows: one of which is the incoherent state and the other is the highly synchronized state (low and high  $R$  windows). In the model, we set thresholds as  $R = 0.3$  and  $R = 0.5$  for low and high  $R$  windows, respectively. We tested various threshold values such as 0.25 and 0.55, and 0.2 and 0.6, and could not find qualitative differences among them. We avoided using thresholds that were too small or too large because it eliminates low and high  $R$  time windows. For the EEG data, we set  $T = 250$  msec with

50 msec overlaps for CS, PSY, ISO, and KET, and  $T = 300$  msec with 50 msec overlaps for MCS and UWS due to the sampling frequency of the data. We concatenated  $R$  values across all temporal windows, all subjects, and all six different states of consciousness to find the total mean and standard deviation of  $R$ . With thresholds  $\langle R \rangle_{total} \pm 0.5 \sqrt{\langle R^2(t) \rangle - \langle R(t) \rangle^2}$ , we classified all temporal windows across all states into windows with low and high  $R$ .

## Symbolic Mutual Information and Susceptibility

To calculate the information processing capabilities for each temporal window, we measured symbolic mutual information (SMI) and network susceptibility in the model and EEG signals. The SMI measures the amount of information sharing, quantifying the extent of non-random joint fluctuations between two signals  $X$  and  $Y$ . To calculate the SMI, the signals  $X$  and  $Y$  are first transformed into sequences of discrete symbols  $(\hat{X}, \hat{Y})$  with a fixed symbol size  $m$  with a temporal separation  $\tau$ . It calculates a joint probability of co-occurring symbolic patterns between two signals.

$$SMI(\hat{X}, \hat{Y}) = \frac{1}{\log(m!)} \sum_{\hat{x} \in \hat{X}} \sum_{\hat{y} \in \hat{X}} p(\hat{x}, \hat{y}) \log \frac{p(\hat{x}, \hat{y})}{p(\hat{x})p(\hat{y})}$$

where  $p(\hat{x})$  is the probability occurring symbol  $\hat{x}$  in the time series  $\hat{X}$ . We set  $m = 3$ , leading to  $3! = 6$  of different symbol pairs  $(\hat{x}, \hat{y})$  that can potentially exist in the transformed symbolic time-series. In the model, we used  $\tau = 14$  (28 ms) to get the maximum resolved frequency  $f_m = \frac{f_s}{k \times \tau} = \frac{500 \text{ Hz}}{3 \times 14} = 11.9 \text{ Hz}$ , which is suitable for our interest of frequency range ( $\sim 12 \text{ Hz}$ ) (Imperator et al., 2019). The SMI was calculated between all node pairs in the model (all channel pairs of the EEG signals) for all temporal windows we defined above. We took an average across SMI values between pairs and defined the average SMI as the amount of total information sharing in each temporal window. For the EEG signals, we compared pairwise SMI values of real data to the pairwise SMI values of twenty surrogate data set with phase randomization technique and used only statistically significant SMI values ( $p < 0.01$ , Wilcoxon rank-sum test). The average SMI value of EEG channel pairs was used as an amount of total information sharing in each temporal window.

The network susceptibility was also measured in both model and EEG signals. The network susceptibility  $\chi$  was defined as following (Yoon et al., 2015):

$$\chi = N \frac{\langle r^2(t) \rangle - \langle r(t) \rangle^2}{\langle r(t) \rangle}$$

We defined the stationary dynamical susceptibility  $\chi$  for each temporal window to see how susceptible the brain state would be to the external information from the outside of the brain in each temporal window.

## Topographic Similarity

To understand distinct functional network configurations engaged in low and high  $R$  windows, we measured topographic

similarity  $S^{amp}$ , which is defined as Spearman correlation between node degrees and node amplitudes.

In the model, the degree of brain regions was defined as the number of structural connections between one region and the other regions. The instantaneous amplitude  $Z(t)$  was measured by the absolute value of the complex variable  $z(t)$ . The instantaneous Spearman correlation between degree and amplitude,  $S^{amp}(t)$ , was calculated for each time point for 60 s for all parameter combinations of the model. Then we took average of  $S^{amp}(t)$  for each temporal window. For the EEG data, the degree of a node (channel) was inferred from the average functional connectivity strength over time measured by weighted phase lag index (wPLI) within a frequency range from 4 to 12 Hz for each subject. The wPLI is a measure that captures phase locking between two signals  $X$  and  $Y$ , relatively robust to volume conduction of EEG (Vinck et al., 2011).

$$wPLI_{XY} = \frac{|E\{I(C_{XY})\}|}{E\{|I(C_{XY})|\}} = \frac{|E\{|I(C_{XY})| \operatorname{sgn}(I(C_{XY}))\}|}{E\{|I(C_{XY})|\}}$$

where  $I(C_{XY})$  is an imaginary part of cross-spectrum  $C_{XY}$  between two signals  $X$  and  $Y$ . Here we used Hilbert-transformed complex signals for the calculation of cross-spectrum  $C_{xy}$ . The  $wPLI_{XY}$  equals to 1 if the phases of one signal  $X$  always lead or lag the phases of the signal  $Y$ , and equals to 0 if the phase lead/lag relationship between two signals is perfectly random. We constructed the wPLI matrix across all channels for each 30-s epoch with a 5-s overlap and binarized the wPLI matrix for each epoch with the top 30 % wPLI values. We calculated the degree of channels for each epoch and took the average over all epochs so that we can capture the statistically intuitive structural degree that reflects the strength of neural communication across brain regions associated with each EEG electrode. Note that the degree is extracted from the CS (baseline) for each subject and applied to the analysis for the disrupted states of consciousness induced by isoflurane and ketamine. Since MCS and UWS patients have no baseline, we excluded those data set for this analysis. The instantaneous amplitude was calculated by the absolute value of the Hilbert transformed EEG signals from a frequency range of 4–12 Hz. The instantaneous Spearman correlation between degree and amplitude,  $S^{amp}(t)$ , was calculated for each subject and CS, ISO, and KET.

## Correlation Between Network Synchronization, Symbolic Mutual Information, Network Susceptibility, and Topographic Similarity

According to our hypothesis, levels of network synchronization are associated with the network's preference for internal and external modes in the brain on a sub-second time scale. Therefore, we calculated Spearman correlation between the level of network synchronization and the information processing metrics (SMI and  $\chi$ ). We also calculated Spearman correlation between the level of network synchronization and the topographic similarity. The correlation between  $R$ , SMI,  $\chi$ , and  $S^{amp}$  were calculated across all temporal windows.

For the model, we calculated Spearman correlations for each parameter combinations with one hundred different simulations (Figure 2). For the EEG data, we calculated Spearman correlations between  $R$ , SMI,  $\chi$ , and  $S^{amp}$  for each subject and each state to investigate whether the relationships we found from the model hold for the empirical EEG data.

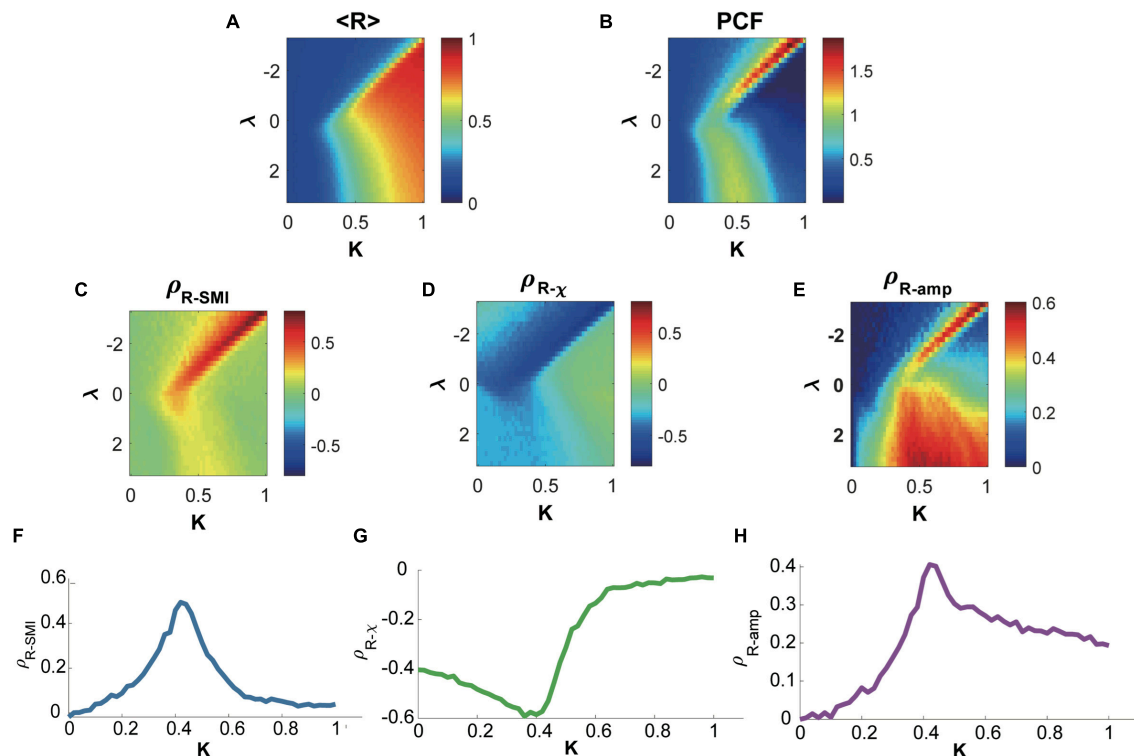
## Joint Histogram Between Symbolic Mutual Information and $\chi$

To visualize the distinct network preferences for internal and external processing modes, a joint histogram of SMI and  $\chi$  of low and high  $R$  windows was calculated. The joint histogram for the model was calculated near and far from the critical state across all temporal windows of one hundred frequency distributions with the bin size 0.02 for SMI and 0.1 for  $\chi$ . For the EEG data, we calculated a joint histogram of SMI and  $\chi$  for each temporal window with the bin size 0.02 for SMI and 0.1 for  $\chi$  during CS, ISO, KET, PSY, MCS, and UWS. We can easily figure out the temporal window's preference for internal or external information by calculating the joint histogram.

## Power-Law Analysis for Dwell-Time of the Positive Correlation Between Degree and Amplitude

It has been known that one of the characteristic features of criticality is the power-law distributions of dynamics. The dynamics following probability distribution  $p(x) \propto x^{-\beta}$  imply that all values  $x$  can occur without a characteristic size or scale (Alstott et al., 2014). It has been also known that the frequency density of phase-locking intervals and the change in the number of phase-locked pairs between successive time points display power-law distributions at criticality in the Ising model and Kuramoto model (Kitzbichler et al., 2009). Since the network synchronization is correlated with the topographic similarity  $S^{amp}$  (relationship between functional network configuration constrained by network structure in the brain), we estimated the probability distribution of dwell-time of  $S^{amp}$  to check whether the relationship between functional network configuration and network structure follows power-law during conscious wakefulness. For the EEG data, we measured dwell-time of positive correlation periods (hub-dominant configuration periods) across all subjects for each state. We used a python package “power-law”<sup>1</sup> to estimate the probability distributions of positive correlation periods. We fitted dwell-time distributions to power-law and exponential distributions and compared their loglikelihood values to determine which distribution is well-fitted to the data. Using the loglikelihood ratio  $R_L$  and p-value, we obtained the estimated probability distribution with the exponent  $\beta$ .  $R_L$  has a positive value if the power-law fit is more appropriate, while it has a negative value if the exponential fit  $p(x) \propto e^{-\beta}$  is more appropriate. We set the minimal value  $x_{min}$  as 200-ms for the EEG data to match the speed of

<sup>1</sup><https://pypi.org/project/powerlaw/>



**FIGURE 2 |** Relationships between network synchronization, symbolic mutual information, network susceptibility, and topographic similarity in the brain network model near and far from criticality. **(A)** The average network synchronization  $\langle R \rangle$  across all parameter sets  $K$  and  $\lambda$ . **(B)** Pair correlation function (PCF), a surrogate indicator of criticality in parameter space  $K$  and  $\lambda$ . Accordingly, a critical state is defined with the maximum PCF. Spearman correlations between network synchronization  $R$  and **(C)** symbolic mutual information SMI, **(D)** network susceptibility  $\chi$ , and **(E)** topographic similarity  $S^{amp}$ . The brain network model shows a maximum positive (negative) correlation between  $R$  and SMI ( $R$  and  $\chi$ ) near the critical states (states with maximum PCF along with  $K$  for each  $\lambda$ ). The  $S^{amp}$  also shows a maximum positive correlation with  $R$  near the critical states. A representative example for the correlation between **(F)**  $R$  and SMI, **(G)**  $R$  and  $\chi$ , and **(H)**  $R$  and  $S^{amp}$  relationship was investigated at  $\lambda = -0.6$ . The critical state appears  $K \cong 0.42$ .

the brain network dynamics and focused on a “heavy-tailed” characteristic of power-law.

## Functional Magnetic Resonance Imaging Experimental Protocol, Data Acquisition, and Preprocessing

The fMRI data were collected at two different research sites, Wisconsin and Shanghai. The experimental protocol for the first data set recorded from Wisconsin was approved by the Institutional Review Board of Medical College of Wisconsin (MCW). Propofol was administrated to 15 healthy volunteers (male/female 9/6; 19–35 years) and the OAAS (observer’s assessment of alertness/sedation) was quantified to measure the level of behavioral responsiveness. This dataset has been previously analyzed with a different hypothesis and published (Huang et al., 2020). In the current study, we used conscious states (baseline) that subjects responded readily to verbal commands (OAAS score = 5) and deep sedation that subjects have no response to verbal commands (OAAS score = 2–1). At the deep sedation level, the plasma concentration of propofol was maintained at equilibrium by continuously adjusting the infusion rate. The corresponding target plasma concentrations

of propofol vary across subjects ( $1.88 \pm 0.24 \mu\text{g/ml}$ ) due to the individual variability of anesthetic sensitivity. Total 14 subjects were analyzed in the current study because one subject had to be excluded due to excessive movements. Rs-fMRI data of the conscious state and deep sedation both consisted of a 15-min scan. Gradient-echo EPI images of the whole brain were acquired on a 3T Signa GE 750 scanner (GE Healthcare, Waukesha, Wisconsin, United States) with a standard 32-channel transmit/receive head coil (41 slices, TR/TE = 2000/25 ms, slice thickness = 3.5 mm, field of view = 224 mm, flip angle =  $77^\circ$ , image matrix:  $64 \times 64$ ). Rs-fMRI was co-registered by high-resolution anatomical images. See (Huang et al., 2020) for a more detailed experimental protocol.

The second dataset includes 28 healthy participants (male/female 14/14) and 21 patients with disorders of consciousness (male/female 18/3). The study was approved by the Institutional Review Board of Huashan Hospital, Fudan University. The healthy controls had no history of neurological or psychiatric disorders or were taking any kind of medication. The patients were diagnosed as either minimally conscious or in the vegetative state/unresponsive wakefulness syndrome (UWS) according to the Coma Recovery Scale-Revised (CRS-R) on the day of fMRI scanning. We analyzed the data of 13 patients who



were diagnosed as UWS in the current study. This dataset also has been published using different a hypothesis (Huang et al., 2020). Gradient echo EPI images of the whole brain for the second dataset were acquired on a Siemens 3T scanner (Siemens MAGNETOM, Germany) with a standard 8-channel head coil (33 slices, TR/TE = 2000/35 ms, slice thickness = 4 mm, field of view = 256 mm, flip angle = 90, image matrix: 64 × 64). Total two hundred EPI volumes (6 min and 40 s) were acquired with high-resolution anatomical images.

Preprocessing steps were implemented in AFNI<sup>2</sup>, which included slice timing correction, head motion correction/realignment, frame-wise displacement estimation, coregistration with high-resolution anatomical images, spatial normalization into Talairach stereotactic space, high-pass filtering (>0.008 Hz), regressing out undesired components (e.g., physiological estimates, motion parameters), spatial smoothing (6 mm full-width at half-maximum isotropic Gaussian kernel), temporal normalization (zero mean and unit variance). Global signal regression (GSR) was not applied to preserve whole-brain patterns of co-activation or co-deactivation. More detailed preprocessing steps were described in the paper (Huang et al., 2020).

## Co-activation Pattern Analysis for Functional Magnetic Resonance Imaging Data

An unsupervised machine-learning approach named k-means clustering algorithm was used to define co-activation patterns (CAPs). The algorithm operates as a classifier of a set of objects (e.g., fMRI volumes) to minimize within category (e.g., patterns) differences and maximize across category differences. Specifically, the fMRI volumes were pooled together across states (baseline, propofol-induced states, ketamine-induced states, and UWS) and subjects, and classified into k number of clusters (patterns) based on their spatial similarity, yielding a set of CAPs or brain states (Huang et al., 2020). As such, we obtained time-series of discrete CAP labels from the original fMRI time-series (voxels × volumes). The volumes containing motion artifact tagged by motion censoring procedure were not included in the CAP analysis. We evaluated the validity of the number of clusters by measuring the inter-dataset similarity (Euclidean distance) of the average CAP occurrence rate distributions of conscious and unresponsive conditions. We searched the number of k from 2 to 30 using an index,  $(CC + UU)/(2 \times CU)$ , as the ratio of inter-dataset similarity among conscious conditions (CC) and among unresponsive conditions (UU) versus the inter-dataset similarity among conscious and unresponsive conditions (CU) across all datasets and found the optimized number of k as 8. Therefore, each fMRI volume from all datasets corresponded to one of the patterns from 8 CAPs. In the current study, we used pre-defined CAPs and analyzed a subset ( $n = 69$ ) of originally published datasets including subjects under wakefulness baseline condition ( $n = 42$ ), deep sedation induced by propofol ( $n = 14$ ), and UWS patients ( $n = 13$ ).

<sup>2</sup><http://afni.nimh.nih.gov/>

## Dominant Co-activation Patterns for Incoherent and Highly Synchronized States

In the EEG data, we defined the incoherent and highly synchronized state that presumably corresponds to externally susceptible/internally integrated state using certain thresholds of the level of synchronization. To define the incoherent/highly synchronized state for the fMRI data, we also measured the level of synchronization  $r(t)$  across voxels for each fMRI volume. Then we classified fMRI transient state to incoherent or highly synchronized state using the thresholds  $\langle r(t) \rangle_{total} \pm 0.5\sqrt{\langle r^2(t) \rangle - \langle r(t) \rangle^2}$ . The  $\langle r(t) \rangle_{total}$  were calculated from all concatenated fMRI volumes of three different states of consciousness (baseline, propofol-induced deep sedation, and UWS). Then we investigated occurrence rate distributions of CAPs for the incoherent or highly synchronized state with a permutation test. For the permutation test, we generated null CAP time-series by 20000 permutations, randomly and uniformly exchanging CAPs in time. Then the original fMRI time points (volumes) that levels of synchronization have below (above) threshold  $\langle r(t) \rangle_{total} - 0.5\sqrt{\langle r^2(t) \rangle - \langle r(t) \rangle^2}$  ( $\langle r(t) \rangle_{total} + 0.5\sqrt{\langle r^2(t) \rangle - \langle r(t) \rangle^2}$ ) for each state of consciousness were selected and patterns corresponding to those time points were used to examine whether the occurrence probabilities significantly deviate from uniformly random sequences. The significantly dominant CAPs were determined at the significance level of  $p < 0.005$  (0.5% percentile of the null distributions: one-sided).

## RESULTS

### Criticality Produces Distinct Preferences for Internal Information and External Stimuli in the Brain Network: A Computational Model Study

We propose that criticality in the brain network produces distinct preferences for internal information of the network and external stimuli, which may result from the general property of the network responsiveness with high and low levels of global network synchronization. By contrast, the preferences may vanish when the brain network is positioned far from criticality. To test this hypothesis, we used a large-scale human brain network model, simulating various brain network behaviors near and far from criticality.

The coupled Stuart-Landau model, which has successfully described the characteristics of EEG, MEG, and fMRI in different states of consciousness, was used in this study. Spontaneous oscillations of the brain network were simulated with different sets of model parameters including the bifurcation parameter  $\lambda$ , coupling strength  $K$ , and diffusive coupling  $\alpha$  (see “Materials and Methods” for details), which produces different brain states near and far from criticality. The level of network synchronization  $R$  was measured with a temporal average of

the order parameter  $r(t)$ . **Figure 2A** presents  $\langle R \rangle$ , which is an average  $r(t)$  over 60 s, in the parameter space  $(\lambda, K)$ . **Figure 2B** presents a pair correlation function (PCF) in the parameter space  $(\lambda, K)$ . The PCF measures a temporal variance of network synchronization as a surrogate indicator of criticality (Yoon et al., 2015). The critical state was defined with the maximal PCF, indicating the largest synchronization fluctuation at criticality. For instance, we defined the brain state at  $K \cong 0.42$  as one of the critical states in the model (**Figure 2B**). The brain network's preferences for integrating internal information of the network and external stimuli were respectively evaluated with symbolic mutual information SMI (King et al., 2013) and network susceptibility  $\chi$  (Yoon et al., 2015). SMI quantifies the amount of shared information and  $\chi$  quantifies network susceptibility to external stimuli. **Figures 2C,D** show that both measures have maximal (minimal) Spearman correlation coefficients with  $R$  near the critical states. In other words, a brain network with a larger  $R$  has a larger SMI and smaller  $\chi$  (preference for integrating internal information of the network), whereas a brain network with a smaller  $R$  has a smaller SMI and larger  $\chi$  (preference for external stimuli). Here,  $R$ , SMI, and  $\chi$  were calculated within each 250-ms temporal window with 50-ms overlap, and the correlations among them were calculated across the temporal windows. Note that the sub-second temporal scale is a suitable time resolution for the information integration processing in a large-scale network level (Bressler and Tognoli, 2006).

We also examined the functional network configurations of high and low  $R$  windows to investigate any structural difference in functional brain networks for two different preferences. We first calculated topographic similarity  $S^{amp}$ , which is defined as a Spearman correlation between node degrees of the anatomical brain network and average amplitudes of the simulated brain activities. In this analysis, a larger  $S^{amp}$  implies that hub nodes are more dominant in network dynamics along with higher amplitudes. Then we calculated the Spearman correlation coefficient between  $S^{amp}$  and  $R$ . **Figure 2E** shows that the correlation between  $R$  and  $S^{amp}$  is also maximal near the critical state, suggesting that a highly synchronized brain network is more likely to have a hub-dominant functional network configuration that is optimal for integrating internal information of the brain network. We present a representative example of the relationships at  $\lambda = -0.6$  with the diffusive coupling of  $\alpha = 0.5$ . **Figures 2F–H** show that those relationships are the largest near the critical state ( $\rho^{R-SMI} = 0.50$  at  $K = 0.42$ ,  $\rho^{R-\chi} = -0.59$  at  $K = 0.38$ , and  $S^{amp} = 0.41$ , respectively). The results are consistent with different diffusive coupling parameters,  $\alpha = 0$  and  $\alpha = 1$  (**Supplementary Figures 1, 2**).

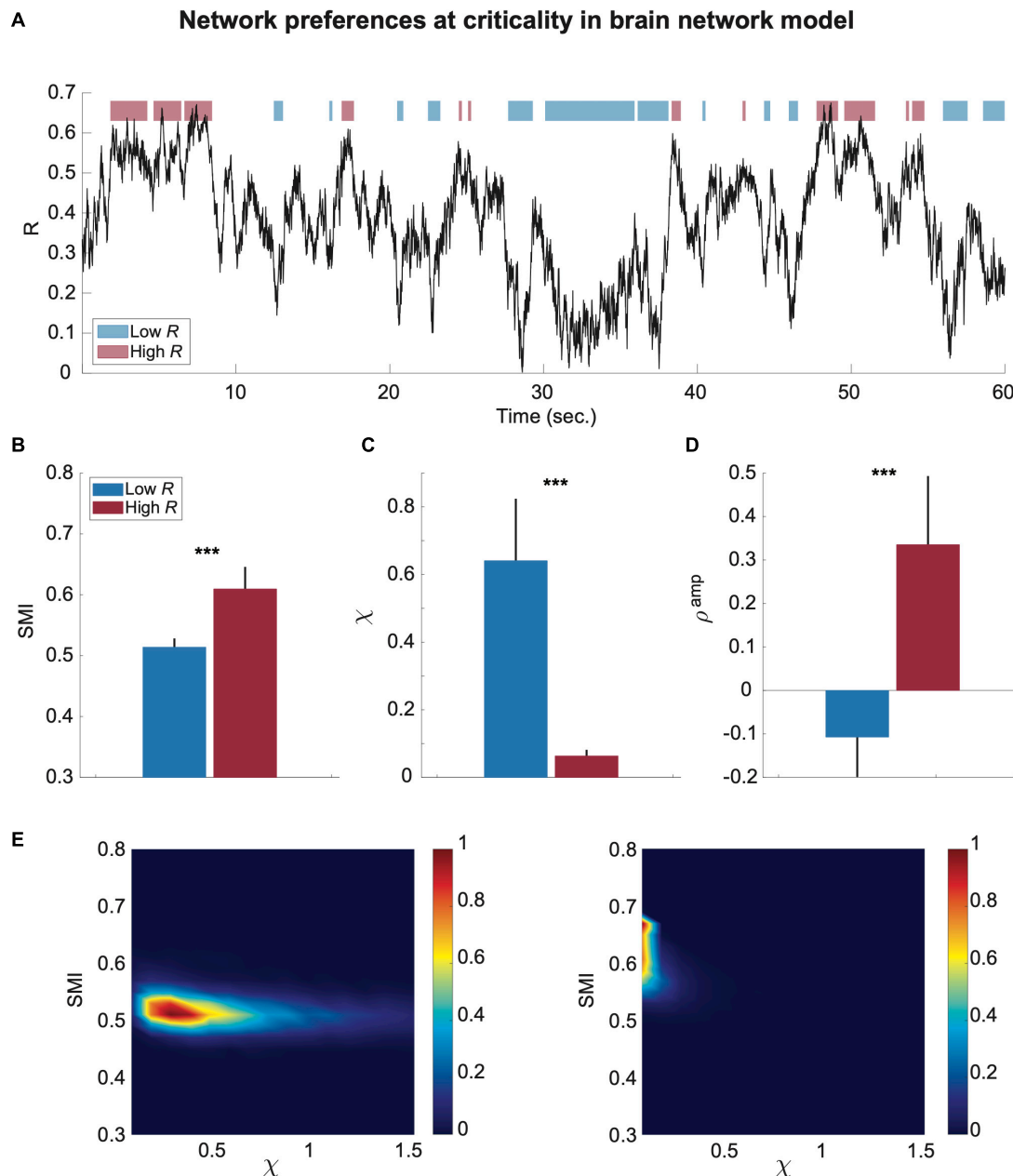
## The Preferences for Internal Information and External Stimuli of the Brain Network Are Prominent Near Criticality in a Sub-Second Time Window: A Computational Model Study

To test whether the preferences of the brain network are significantly different in the time domain, we classified the

simulated brain signals into low and high  $R$  windows and investigated SMI and  $\chi$  values for each temporal window (250-ms). **Figure 3A** shows a temporal evolution of  $R$  during 60 s in the brain network model near a critical state at  $\lambda = 0.6$ . The low  $R$  and high  $R$  windows were defined with thresholds  $R < 0.3$  and  $R > 0.5$ , respectively (blue: low  $R$  and red: high  $R$  in **Figure 3A**). The high  $R$  windows are characterized by large SMI and small  $\chi$ , whereas the low  $R$  windows have relatively small SMI and large  $\chi$  ( $***p < 0.001$  for SMI;  $***p < 0.001$  for  $\chi$ , Wilcoxon rank-sum test, in **Figures 3B,C**). The results show that the brain states of these high and low  $R$  windows define preferences for internal information vs. external responsiveness in the network. We also showed that high  $R$  windows have a large positive  $S^{amp}$ ; by contrast, low  $R$  windows have negative  $S^{amp}$  ( $***p < 0.001$ , Wilcoxon rank-sum test, **Figure 3D**). The  $S^{amp}$  value itself was variable across different parameter sets, but the positive correlation between  $R$  and  $S^{amp}$  near criticality was consistent (**Supplementary Figures 1, 2**). Our results suggest that the fluctuation of network synchronization near criticality naturally produces a continuous switching between internal and external modes in the brain network with different functional network configurations. In **Figure 3E**, the joint histograms of SMI and  $\chi$  (see “Materials and Methods” for more details) clearly show the distinct preferences of high (right) and low (left)  $R$  windows for internal information of the network and external stimuli (i.e., larger SMI and smaller  $\chi$  for high  $R$  windows; smaller SMI and larger  $\chi$  for low  $R$  windows), respectively. Notably, such typical preferences of high and low  $R$  windows only appear near criticality. We also compared the mean values of  $R$ , SMI, and  $\chi$  of low and high  $R$  windows for three different states in the model (critical state  $Cr$ , below critical state  $Cr_b$ , and above critical state  $Cr_a$ ); the results were presented in **Supplementary Figure 3**. The results show that the mean value of  $R$  is not linearly correlated with SMI and  $\chi$ , rather it shows obvious state-dependence. For instance, for  $Cr_a$ , the SMI of low  $R$  windows are larger than those of high  $R$  windows for  $Cr$  and  $Cr_b$ , even though their  $R$  values are smaller. Such state dependence also holds true for  $\chi$ .

## The Preferences of the Brain Network for Internal Information and External Stimuli Are Prominent in the Conscious State: An Electroencephalography Study

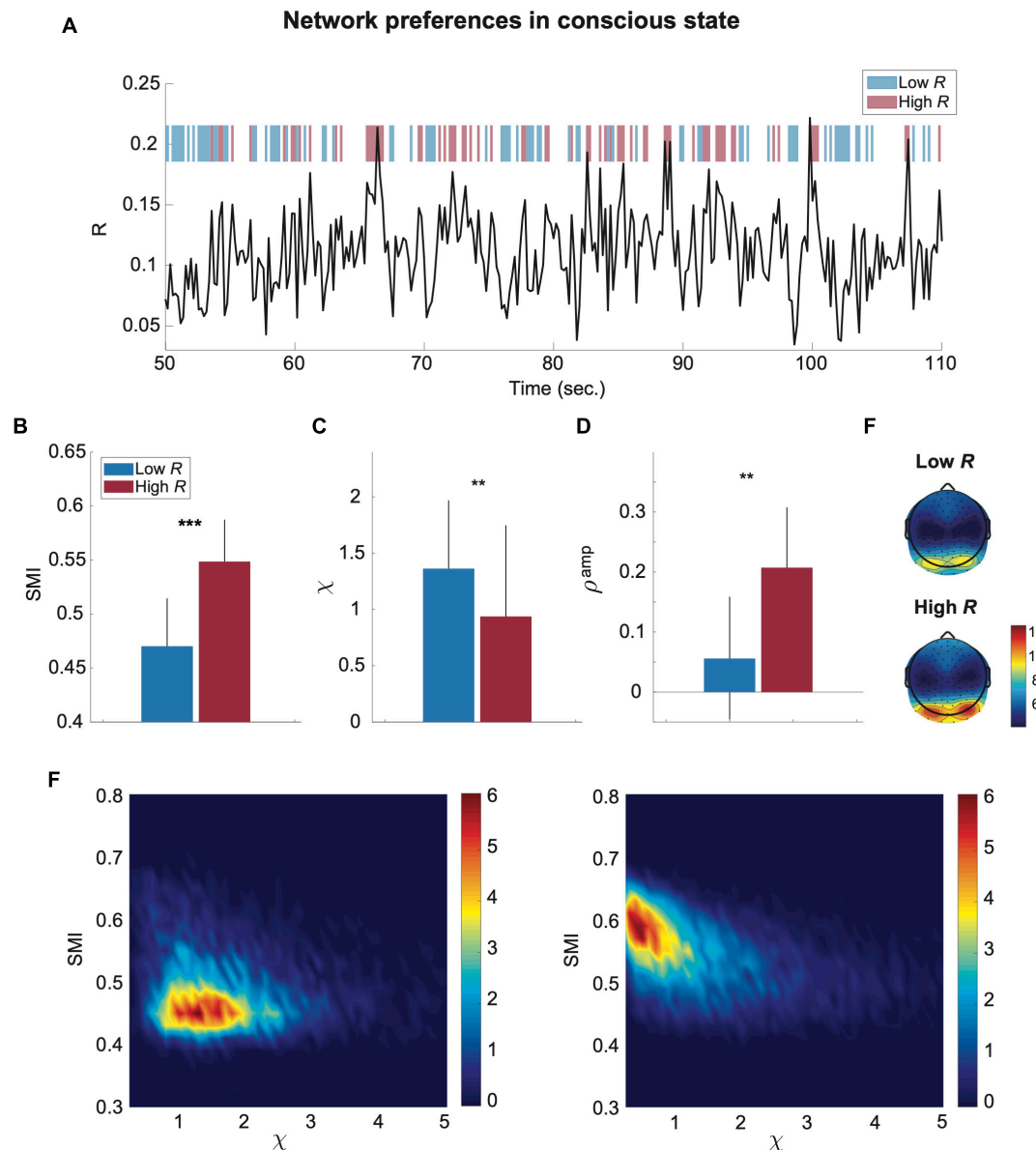
To test the modeling results empirically, we analyzed EEG signals of conscious eye closed-resting states in humans ( $n = 24$ ). Computational model and empirical studies suggested that theta and alpha-band (4–12 Hz) oscillations are globally networked in the cortex and spatiotemporally organize neural processing with traveling waves across the human cortex (Zhang et al., 2018; Roberts et al., 2019). Thus, here we focused on the EEG signals of 4–12 Hz that is suitable to test the characteristic behavior of the brain network near criticality. We also observed that the results of this frequency band (4–12 Hz) are most consistent with the model predictions (**Supplementary Figure 4**). As in the model data analysis, we calculated  $R$ , SMI,  $\chi$ , and  $S^{amp}$  for each temporal window of 250-ms. **Figure 4A** presents a temporal



**FIGURE 3 |** The preferences of the brain network for internal information integration and external responsiveness are prominent in sub-second time windows (250-ms) in the model. **(A)** An example of a network synchronization fluctuation near a critical state. Red and blue bars indicate temporal windows classified into low and high  $R$ . The SMI,  $\chi$ , and  $S^{amp}$  were calculated for each window. Comparisons of **(B)** SMI, **(C)**  $\chi$ , and **(D)**  $S^{amp}$  between temporal windows of low  $R$  (blue) and high  $R$  (red) are presented. The black lines of the colored bars indicate from 25% to 75% quantiles. A Wilcoxon rank-sum test was performed (\*\* $p < 0.001$ ). **(E)** Joint histograms of the low (left) and high (right)  $R$  windows. The low  $R$  window is characterized by relatively low SMI and high  $\chi$ , while the high  $R$  window is characterized by relatively high SMI and low  $\chi$ .

evolution of  $R$  for 60 s. We used thresholds  $R = \langle R \rangle_{total} \pm 0.5\sqrt{\langle R^2(t) \rangle - \langle R(t) \rangle^2}$ , where  $\langle R \rangle_{total}$  is an average of  $R$  across all subjects and all states of EEGs (see “Materials and Methods” for more details), to classify the temporal windows into high  $R$  and low  $R$  windows (blue and red blocks in **Figure 4A**). Note that we used different thresholds for the source signals (model) and sensor signals (EEG) due to the differences in their

signal variability. The results of the EEG analysis were consistent with the results of the computational model and show large SMI and small  $\chi$  for high  $R$  windows and small SMI and large  $\chi$  for low  $R$  windows (\*\* $p < 0.001$  for SMI; \*\* $p < 0.01$  for  $\chi$ , Wilcoxon rank-sum test, **Figures 4B,C**). The high  $R$  windows also presented a large positive  $S^{amp}$  (\*\* $p < 0.01$ , Wilcoxon rank-sum test, in **Figure 4D**). For  $S^{amp}$ , we used average degrees



**FIGURE 4 |** EEG functional brain networks in conscious state present distinct preferences for internal information integration and external responsiveness in the time domain. **(A)** An example of a network synchronization fluctuation during a conscious state. Red and blue bars respectively indicate temporal windows classified into low and high  $R$ . The SMI,  $\chi$ , and  $S^{amp}$  were calculated for each 250-ms window. Comparisons of **(B)** SMI, **(C)**  $\chi$ , and **(D)**  $\rho^{amp}$  between low  $R$  (blue) and high  $R$  (red) windows are presented. The black lines of the colored bars indicate from 25% to 75% quantiles. A Wilcoxon rank-sum test was performed (\*\* $p < 0.001$  and \*\* $p < 0.01$ ). **(E)** Functional brain network configurations for low and high  $R$  windows. As expected in the model study, the amplitudes of the posterior area (hub regions) in high  $R$  windows are larger than those in low  $R$  windows. **(F)** Joint histograms of the temporal window of low (left) and high (right)  $R$ . The low  $R$  windows are characterized by relatively low SMI and high  $\chi$ , while the high  $R$  windows are characterized by relatively high SMI and low  $\chi$ . The empirical data are consistent with the model prediction.

of wPLI networks and average amplitudes of EEG signals of a temporal window, assuming that the wPLI network averaged over a long time may resemble its underlying structural network (Kim et al., 2018). Similar to the brain network model near criticality, the brain network during conscious states demonstrates the distinct preferences of high and low  $R$  windows for internal information and external stimuli as well as the posterior hub-dominant network configuration in high  $R$  windows (Figure 4E).

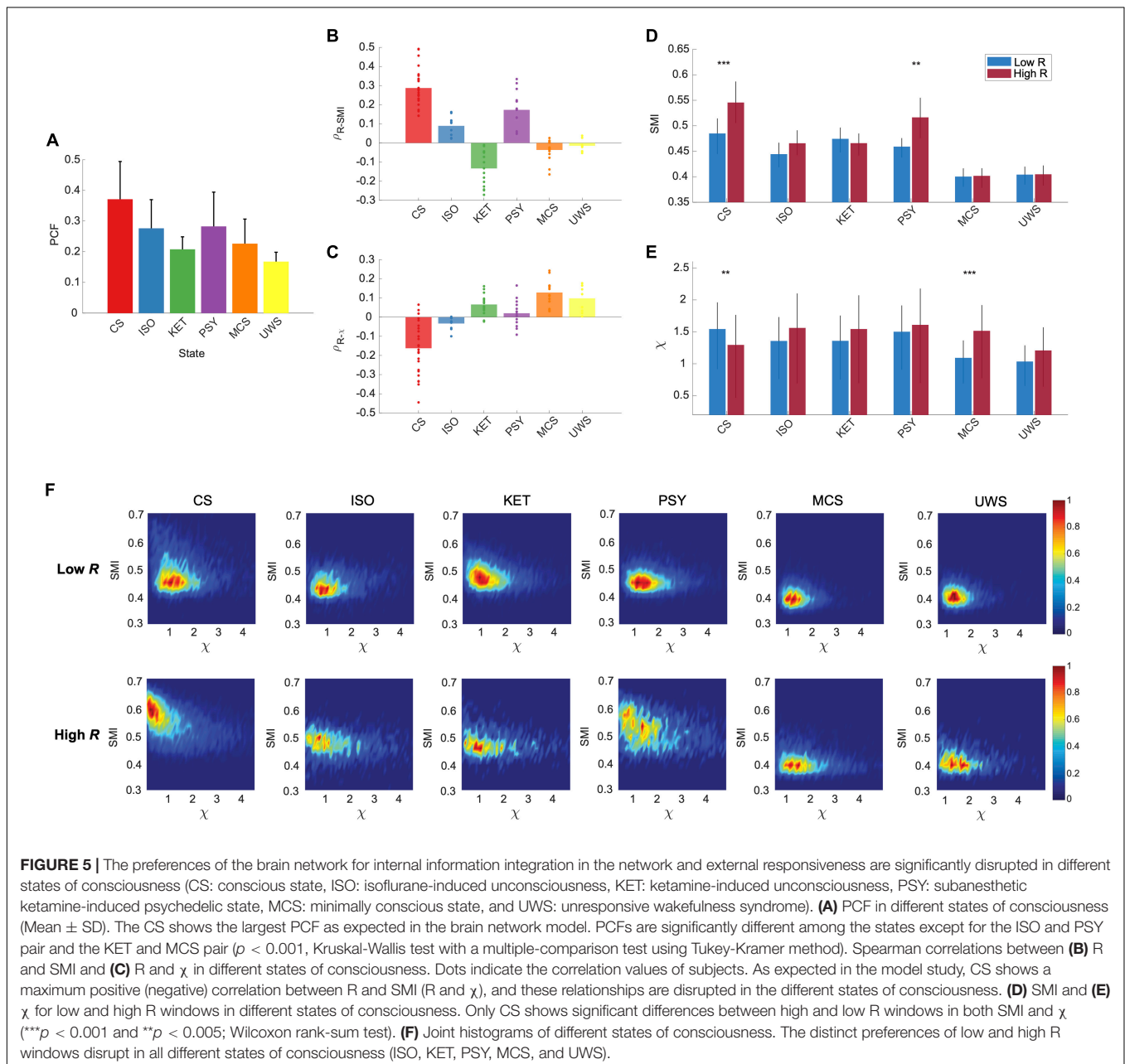
Furthermore, joint histograms of SMI and  $\chi$  clearly visualize the distinctive preferences of high (right) and low (left)  $R$  windows in conscious states (Figure 4F). In sum, the brain network constructed from EEGs during conscious states shows the typical network properties of high and low  $R$  windows in terms of SMI and  $\chi$ , which is predicted by the model study, and temporally yields continuous switching between preferences for internal information and external stimuli.



## The Preferences of the Brain Network for Internal Information and External Stimuli Vanishes in Different States of Consciousness: An Electroencephalography Study

Our modeling results predict that the distinct preferences of high and low  $R$  windows for internal information and external stimuli vanish when the brain network deviates from criticality, with reduced synchronization fluctuation (Figure 2). To test this model prediction, we analyzed EEGs from conscious eyes-closed resting states in humans (CS,  $n = 24$ ), isoflurane-induced unconsciousness (ISO,  $n = 9$ ),

ketamine-induced unconsciousness (KET,  $n = 15$ ), subanesthetic ketamine-induced psychedelic state (PSY,  $n = 15$ ), and disorders of consciousness such as minimally conscious states (MCS,  $n = 16$ ) and unresponsive wakefulness syndrome (UWS,  $n = 9$ ). We first calculated the PCF to evaluate the synchronization fluctuation and demonstrated that the PCF significantly decreases in all altered and disrupted states of consciousness ( $p < 0.001$ , Kruskal-Wallis test, Figure 5A). The PCFs of the five different states of consciousness are significantly different from one another ( $p < 0.001$ , multiple-comparison test with Tukey-Kramer method) except for the ISO and PSY ( $p = 0.59$ ) pair and the KET and MCS pair ( $p = 0.08$ ).



We also calculated the  $R$ ,  $SMI$ , and  $\chi$  for each temporal window and calculated the correlations between them. With reduced PCFs, as expected, all of the altered and disrupted states of consciousness lost the correlations between  $R$  and  $SMI$  and between  $R$  and  $\chi$ , except for  $R$  and  $SMI$  of PSY (Figures 5B,C). While investigating several other frequency bands (1–4 Hz, 1–20 Hz, 8–12 Hz, 13–30 Hz, and 30–42 Hz), we found that the frequency band of 8–12 Hz is also consistent with the results from 4 to 12 Hz (Supplementary Figure 3). Furthermore, we also classified the temporal windows of all states into high and low  $R$  windows with the same thresholds of CS. The distinct preferences between internal information and external stimuli observed in CS (large  $SMI$ /small  $\chi$  for high  $R$  and small  $SMI$ /large  $\chi$  for low  $R$  window) faded away in the different states of consciousness ( $SMI$ ; CS:  $***p < 0.001$ , ISO:  $p = 0.03$ , KET:  $p = 0.48$ , PSY:  $**p < 0.005$ , MCS:  $p = 0.90$ , and UWS:  $p = 1$ ,  $\chi$ ; CS:  $**p < 0.005$ , ISO:  $p = 1$ , KET:  $p = 0.06$ , PSY:  $p = 0.10$ , MCS:  $***p < 0.001$ , UWS:  $p = 0.14$ , Wilcoxon rank-sum test) (Figures 5D,E). We also found that the  $SMI$  of high  $R$  windows in CS is significantly reduced in ISO, KET, PSY, MCS, and UWS ( $p < 0.001$ ; CS vs. all other states, one-way ANOVA with multi-comparison test with Tukey-Kramer method, suggesting that disrupting the brain's capability for processing internal information may be a common mechanism for inducing both altered and depressed states of consciousness. The  $\chi$  of low  $R$  windows of ISO, KET, MCS, and UWS was significantly decreased or maintained compared to that of CS. In Figure 5F, joint histograms of  $SMI$  and  $\chi$  visualize that only CS shows distinctive preferences of low and high  $R$  windows for internal information of the network and external stimuli. Furthermore, we investigated association of the posterior hub-dominant structure in the various states of consciousness (Supplementary Figure 5). Different from CS, ISO shows anterior dominant network configuration for the high  $R$  windows and KET shows no difference between low and high  $R$  windows. However, similar to CS, PSY presents the posterior hub-dominant structure in high  $R$  windows. Notably, the significant network preference for internal integration with the posterior hub-dominant configuration only occurred in CS and the altered state of consciousness (PSY) associated with consciousness (Supplementary Figure 5), implying that the brain's capability for internal information integration with the specific network configuration might play a more prominent role in maintaining consciousness than network susceptibility.

## Dominant Functional Magnetic Resonance Imaging Co-activation Patterns at High and Low Levels of Synchronization in Conscious and Unconscious States

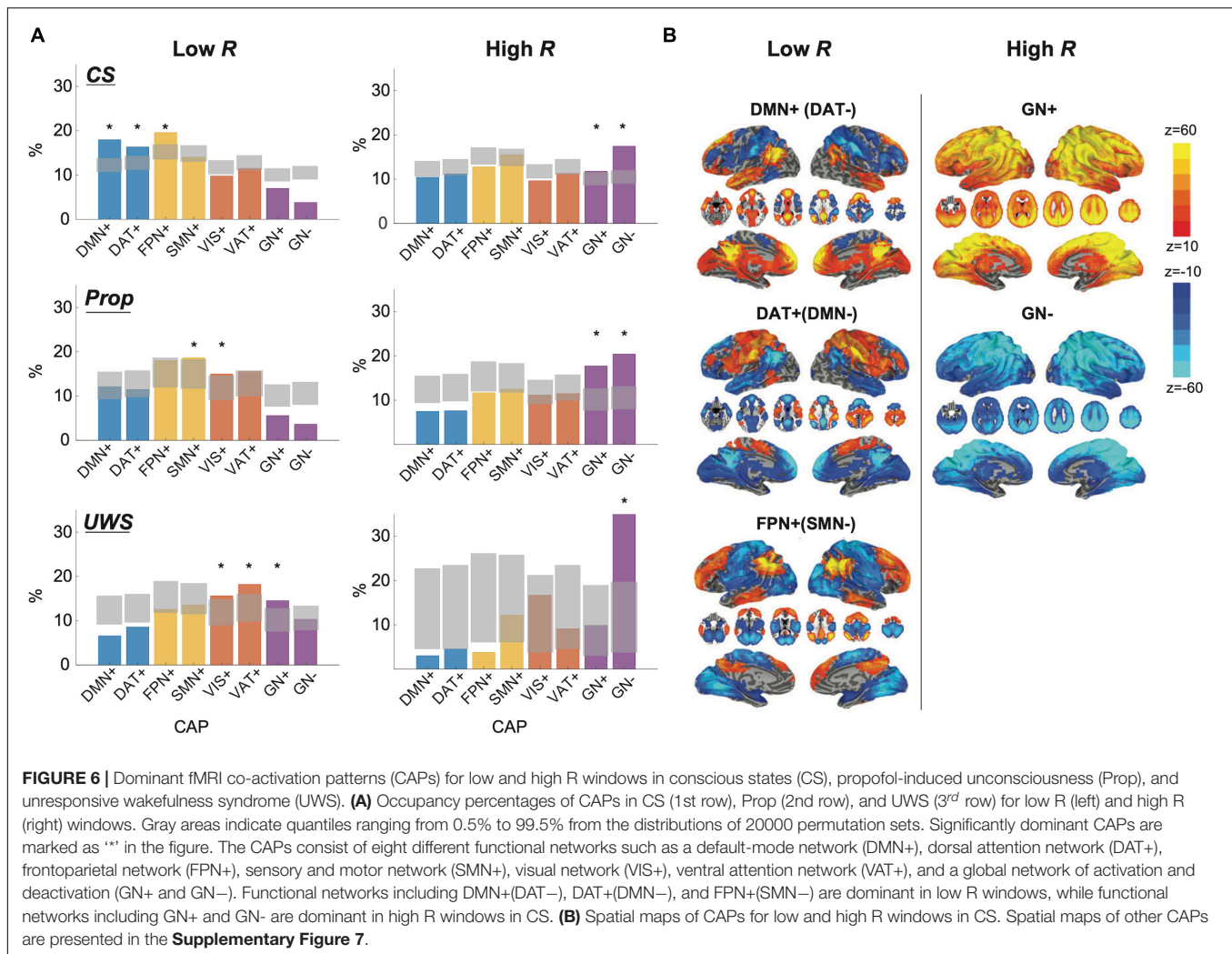
Finally, we tested whether the preference of high and low  $R$  windows can be confirmed using a different brain imaging modality such as fMRI BOLD signals, investigating an association with well-known fMRI networks that are activated while the brain carries out tasks requiring internal and external information processing. We predicted that the fMRI networks relevant to internal or external information processing may be preferentially

activated in high or low  $R$  windows and such preferences may change in disrupted states of consciousness.

To determine which fMRI networks are dominant in high and low  $R$ s, we first classified human fMRI BOLD signals from conscious states (CS,  $n = 42$ ), propofol-induced unconscious states (Prop,  $n = 14$ ), and UWS ( $n = 13$ ) into low and high  $R$  windows with the thresholds  $R_{th} = \langle R \rangle_{total} \pm 0.5\sqrt{\langle R^2(t) \rangle - \langle R(t) \rangle^2}$ , where  $\langle R \rangle_{total}$  is an average network synchronization across all subjects and all states. We then investigated the occurrence rates of co-activation patterns (CAPs) for the low and high  $R$  windows, comparing them with the occurrence rates of a null CAP time-series generated by 20,000 permutations of BOLD signals (Figure 6A). The CAPs were defined as 8 typical patterns of BOLD co-activation across voxels such as default mode network (DMN+), dorsal attentional network (DAT+), frontoparietal network (FPN+), sensory and motor network (SMN+), visual network (VIS+), ventral attention network (VAT+), and a global network of activation and deactivation (GN+ and GN−). More detailed explanation of the CAP analysis is in “Materials and Methods” and Supplementary Figure 7. According to our hypothesis, the low  $R$  windows should be characterized by networks that are sensitive to external stimuli, whereas the high  $R$  windows should be more relevant to networks that are insensitive to external stimuli. Our results show that in CS the low  $R$  windows are dominated by DMN+, DAT+, and FPN+, which are sensitive to external stimuli (Sadaghiani et al., 2015), and the high  $R$  windows are dominated by GN+ and GN−, which are known as arousal networks that are insensitive to external stimuli (Figure 6B; Chang et al., 2016; Liu et al., 2018; Turchi et al., 2018). The dominant CAPs in CS were replaced with other CAPs in Prop and UWS, mostly deactivating DMN+. Thus, the distinct preferences of high and low  $R$ s for internal information of the brain network and external stimuli observed near criticality in the conscious state can be confirmed with fMRI BOLD signals. The disrupted preferences were also observed universally in the brain network that distant from criticality and the EEG and fMRI data in different states of consciousness.

## DISCUSSION

With a computational model, we demonstrated that when the brain network resides near a critical state, the network synchronization  $R$  correlates with the amount of shared information ( $SMI$ ) within the network and the network susceptibility ( $\chi$ ) to external stimuli, respectively. The highly synchronized states prefer internal information integration within the network (large  $SMI$  and small  $\chi$ ), while the incoherent states prefer external stimuli (small  $SMI$  and large  $\chi$ ). However, the correlations between  $R$ ,  $SMI$ , and  $\chi$  were diminished when the brain network deviated from criticality. The modeling study also showed that a highly synchronized brain network displays predominant hub activities, providing a network condition favorable for internal information integration. The modeling results were compared with empirical data across diverse states of consciousness. First, the brain network during conscious states



exhibited the largest synchronization fluctuation compared to the different states of consciousness (ISO, KET, PSY, MCS, and UWS), which confirms criticality of the conscious brain. Second, the EEG functional brain networks during CS showed temporally fluctuating preferences between the internal information of the network and external stimuli, presenting a large  $SMI$ /small  $\chi$  at high  $R$  and a small  $SMI$ /large  $\chi$  at low  $R$ . These preferences were significantly diminished in the different states of consciousness. Additionally, the dwell-time of the hub-dominant network configurations at high  $R$  windows were organized in a scale-free manner. The power-law distribution of the dwell-time changed into a random organization (i.e., exponential distribution) in unconscious states (**Supplementary Figure 8**). Finally, the most frequently observed CAPs at high (low)  $R$  in the fMRI signals corresponded to internally integrative (highly susceptible) brain states.

In sum, distinct network preferences for the internal information within the network or external stimuli spontaneously emerge when the brain network is positioned near criticality and in conscious states. The brain loses such distinct preferences far from criticality and in different states of

consciousness (ISO, KET, PSY, MCS, and UWS), which results in a disruption of temporal information processing capability.

## Criticality's Novel Role in Conscious Brain Function

Biological systems can obtain many functional benefits from operating near a critical point of a phase transition. Criticality in biological systems, conjectured to emerge as the result of adaptive and evolutionary processes, produces an optimal balance between stability and instability, optimal computational capacity, large dynamical repertoires, and greater sensitivity to stimuli (Beggs, 2008; Kitzbichler et al., 2009; Tagliazucchi et al., 2012; Haimovici et al., 2013; Cocchi et al., 2017; Kim and Lee, 2019). The brain near criticality should display characteristic features that are empirically measurable: maximal sensitivity, large spatiotemporal correlations, and large variances in synchronization. Maximal sensitivity is an important property for sensory systems in the brain, such as the olfactory, visual, and auditory systems, optimizing responses to environmental cues (Chialvo, 2006; Kinouchi and Copelli, 2006; Bushdid et al., 2014; Hudspeth, 2014;

Shew et al., 2015). A large spatiotemporal correlation is crucial for the brain to induce coordinated neural activities across space and time (Tagliazucchi et al., 2012), which is a useful mechanism for the generation of long-lasting and slow-decaying memories at multiple timescales (Deco and Jirsa, 2012). A large variance of synchronization induces large statistical complexity and large repertoires in the brain, which is due to the maximal variety of attractors and metastability (Haldeman and Beggs, 2005; Deco and Jirsa, 2012; Haimovici et al., 2013) in corresponding state spaces. It facilitates the spontaneous generation of complex patterns required for optimizing the brain's capability for storing and processing information, enabling the brain to constantly traverse different network configurations, which is associated with cognitive flexibility. Based on our findings in the model study and the empirical data analyses, we propose a novel role for criticality in brain function, which is the provision of a platform that enables the brain to rapidly transition between distinct temporal preferences for internal information within the network and external stimuli. The large network synchronization fluctuation that spontaneously emerges near criticality enables the brain network to integrate internal information or be sensitive to external information in the time domain. Furthermore, the temporally variable network preferences might play the role of a functional platform for continuous switching between internal and external information modes in the brain, which is essential for constructing inner models of the outside world through the recursive learning process (Honey et al., 2017) as well as for the emergence of consciousness. In addition, as the network becomes distant from a critical state along with reduced synchronization fluctuation, such distinct preferences vanish as altered and depressed states of consciousness occur. Our model simulation and empirical data analyses explicitly demonstrate that the variance of synchronization (PCF) is maximal at a critical state (Figure 2B) and in conscious states (Figure 5A). By contrast, the variance is significantly diminished in different states of consciousness such as general anesthesia, psychedelic experiences, and pathological disorders of consciousness (Figure 5A).

## Brain Network Preferences Vanish in Different States of Consciousness

In different states of consciousness (ISO, KET, PSY, MCS, and UWS), brain networks no longer possessed distinct preferences for the internal information of the network or external stimuli. It is important to consider how brain networks in different states of consciousness lose such preferences and what differentiates altered vs. depressed states of consciousness. Decreased *SMI*, especially in high *R* windows, is a common feature across altered and depressed states of consciousness regardless of the type of anesthetic or traumatic injury. However, network susceptibility  $\chi$  was relatively unchanged across different states (ISO, KET, PSY, MCS, and UWS). These results imply that the brain network preferentially loses its capability to integrate internal information, while its susceptibility to external stimuli remains relatively intact. In other words, the brain network may be able to receive external stimuli, but it cannot be globally integrated

during different states of consciousness, which may be specifically related to functional deafferentation and disconnection from the external world due to the isolation of the thalamocortical network (Mhuirheartaigh et al., 2013). In addition, loss of consciousness accompanied the disruption of the posterior hub-dominant network configuration (Supplementary Figure 5), which impairs the functional role of hubs for integrating and transmitting information within the hierarchical brain network (Stam and van Straaten, 2012; van den Heuvel and Sporns, 2013). This hub disruption also causes the selective inhibition of top-down processes by preferentially impeding information flow from hub nodes to peripheral nodes in the brain network (Marinazzo et al., 2012; Stam and van Straaten, 2012; Cohen et al., 2018). Numerous studies have demonstrated that anesthetics selectively inhibit higher-order information integration in top-down processes while preserving the bottom-up and primary sensory processes, highlighting the importance of top-down processes for the emergence of consciousness (Liu et al., 2012; Schroter et al., 2012). Our model and empirical data analyses suggest that such preferential inhibition of internal information integration in the brain network, specifically, via the disruption of the hub-dominant network configuration in high *R* windows, which is likely associated with high-order information integration, is a characteristic phenomenon that can occur when a complex network deviates from criticality. If the network moves to states far from criticality (super- or subcritical states), it will preferentially hamper the normal functions of hubs that have dense connections.

The altered and depressed states of consciousness were associated with distinctively impaired network properties, *SMI* and  $\chi$ , in high and low *R* windows, all of which differentiate these states. Interestingly, the psychedelic and minimally conscious states showed opposite network properties (Figure 5D). The psychedelic state preserved the network preference for integrating the internal information of the network (i.e., positive correlation between *R* and *SMI*); however, the network preference for susceptibility to external stimuli between high and low *R* windows vanished. The low *R* windows no longer retained the bias toward external stimuli. In contrast, the MCS lost the capability for integrating internal information (i.e., it exhibited a small *SMI* in a high *R* window), but instead presented larger network susceptibility in high *R* windows, which is the opposite compared to consciousness (Figure 5E). In other words, the brain network of the MCS disrupted the network's capability for internal information integration with an abnormal relationship between network synchronization and susceptibility. Conversely, the brain network of the psychedelic state still functioned for internal information integration, however, it did not respond to external stimuli properly, which suggests an impaired network preference for external stimuli. Finally, we found that the dwell-time for the hub-dominant network configuration (a positive correlation between node degree and amplitude;  $S^{amp} > 0$ ) follows a power-law distribution while conscious but follows an exponential distribution during ISO and KET (Supplementary Figure 8). Considering the association of a hub-dominant network configuration ( $S^{amp} > 0$ ) with high *R* and *SMI* in both critical and conscious states (Figures 3, 4),



the power-law of the dwell times implies that when the brain integrates internal information (higher *SMI*) in high *R* windows, the hub structure dominates the integration process in the brain network. Also, the posterior hub-dominant brain network configuration is organized temporally in a scale-free manner, which enables the brain to efficiently process higher-order information integration at various time scales. In contrast, the anesthetized states (ISO and KET) lose the brain's capacity to process internal information (low *SMI* and *R*) and their temporal organization becomes close to random processes. More specifically, even with similar levels of *R*, the *SMI* values in high *R* windows are different across states, revealing a state dependence of the brain in relation to its capability for internal information integration in high *R* windows and emphasizing the significance of functional brain network configurations (Supplementary Figure 6).

### Dominant Functional Magnetic Resonance Imaging Co-activation Patterns at High and Low Levels of Synchronization

Functional magnetic resonance imaging studies have demonstrated that conscious states are correlated with critical dynamics (e.g., see Tagliazucchi et al., 2016). Dominant fMRI co-activation patterns for incoherent and synchronized windows in conscious states were presumably associated with brain networks that are sensitive and insensitive to external stimuli, respectively. We showed that, in conscious states, the co-activation patterns for incoherent windows were dominated by DMN+, DAT+, and FPN+. These networks have complex interactions and support higher-order cognitive functions. For example, the DMN engages in a variety of processes such as autobiographical memory, imagination, and self-referencing (Raichle, 2015). The DAT mediates cognitive processes such as goal-driven attention, inhibition, and top-down guided voluntary control (Corbetta and Shulman, 2002; Vossel et al., 2014). The FPN+ is known to flexibly alter its functional connections dynamically according to current task demands (Cole et al., 2013) and has a strong association with working memory (Murphy et al., 2020). Furthermore, these three networks are at a high position of a representational hierarchy, relatively far from the sensory and motor systems in terms of both functional connectivity and anatomical distance (Margulies et al., 2016). Such a hierarchical disposition is thought to allow these networks to process transmodal information in a way that is unconstrained by immediate external stimuli. We suggest that these spatially segregated co-activation patterns (DMN+, DAT+, and FPN+) may be associated with high sensitivity to stimuli. As supported by a previous study by Sadaghiani et al. (2015), pre-stimulus brain states with higher modularity (i.e., higher spatial segregation and lower global integration) bias toward detecting external stimuli, whereas pre-stimulus brain states with lower modularity (i.e., higher spatial homogeneity and higher global integrity) bias toward misses.

In contrast, we found that highly synchronized states during consciousness are dominated by GN+ and GN-. The two co-activation patterns are correlated with global EEG

synchronization in the alpha frequency band (Liu et al., 2018) and associated with arousal fluctuations regulated by subcortical-cortical connectivity (Chang et al., 2016; Liu et al., 2018; Turchi et al., 2018). Particularly, it has been suggested that GN+ is specific to lapses in alertness associated with a transition to a state of lower arousal (Liu et al., 2018), rendering the neural system less sensitive to external stimuli. Taken together, our results suggest that the distinct preferences of the brain network for internal information or external stimuli in low and high levels of network synchronization, regardless of imaging modalities, are common network properties near critical states.

Notably, the temporal scales of the network preference switching in EEG and fMRI data are inherently different due to the measurement of different proxies of brain activity (EEG: neuronal processing; fMRI: blood-oxygenation level). Because of the lack of time resolution of fMRI data, we focused on identifying the relevant brain network structures in high and low *R* windows. However, since the fMRI signals near and far from criticality can be modeled by networked oscillators (Deco et al., 2017, 2018), we expect the general relationship between *SMI*, *R*, and  $\chi$  near criticality we found with the networked canonical oscillator model also holds for the fMRI signals. Further study will be needed to confirm this.

### Potential Mechanisms for the Emergence of Distinct Preferences in the Brain Network

recent studies of brain network characteristics near criticality have provided some evidence of the potential mechanism for the emergence of continuous switching between internal and external modes. One of the potential mechanisms is a global fluctuation in neural gain mediated by ascending neuromodulatory nuclei such as the pontine locus coeruleus (Lee and Dan, 2012; Shine et al., 2018). A series of studies has suggested that the modulation of neural gain with the dynamic changes in noradrenaline results in a large fluctuation between network integration and segregation in the brain (Shine et al., 2018; Li et al., 2019). When the neural gain dynamics (presynaptic afferent input) of the locus coeruleus reside near criticality, a small fluctuation produces a sharp transition between network integration and segregation, which respectively accompanies optimized information transfer and optimized information storage in the brain (Li et al., 2019). Network integration and segregation have also been associated with an increased spatial correlation (correlated with elevated information transfer) and increased autocorrelation times (correlated with increased information storage) with correspondence to high and low levels of phase synchronization (Shine et al., 2018; Li et al., 2019).

A previous computational model study from our research team found that brain network responsiveness significantly depends on the level of network synchronization as well as the distance from criticality when a stimulus is applied (Kim and Lee, 2020). Based on these results, we suggested that a potential mechanism for the relationship between brain network synchronization and responsiveness is the phase response curve in physics, which is a general property of networked oscillators ubiquitously observed in physical and biological systems. The

phase response curve describes the way that a system with a collective periodic behavior (for instance, circadian rhythms, cardiac rhythms, and spiking neurons) responds to external stimuli (Granada et al., 2009; Hannay et al., 2015). The response of an oscillating system is measured by the phase shift from the original phase. This phase shift (i.e., advancing or delaying the original phase) is an inherent characteristic of the oscillatory system, which is determined by given network configurations. Previous analytic studies discovered that a low (high) phase synchronization induces a large (small) phase response to a stimulus, proving that stimulating the phases around a stable fixed point of the phase response curve increases phase response, whereas stimulating the phases around an unstable fixed point decreases phase response (Hannay et al., 2015). These properties generally hold for networks with different coupling functions, network structures, and connectivity (Levnajić and Pikovsky, 2010).

The present study advances previous results by analyzing the role of synchronization fluctuations in the brain network, modulating the distance from criticality in the human brain network model and analyzing the empirical data from EEG and fMRI for various states of consciousness. We directly calculated the amount of information sharing (*SMI*) and network susceptibility ( $\chi$ ) with the network synchronization *R* in the time domain for both the model and data. We propose the large synchronization fluctuation is a functional platform for creating temporal windows that process internal or external information. In particular, the temporal windows characterized by large *SMI* and small  $\chi$  play a role in internal information integration while defending against interventions from the outside world. These characteristics of a highly synchronized network create temporal windows shielded from external perturbations and simultaneously enable the brain to integrate globally distributed information across brain regions. Empirically, we demonstrated that such selective roles for temporal windows, observed in human brain networks in conscious states, are disrupted in different states of consciousness (ISO, KET, PSY, MCS, and UWS). Notably, in our model study, we did not consider the neuromodulation system, which implies that the temporal windows for internal or external information processing can arise through the interplay between regional brain activities while the interactions are close to criticality. In other words, the emergence of the network preference for internal or external information and the continuous fluctuation between these two modes originate as a generic network feature near criticality, regardless of the specific neurobiology.

## Limitations

There are several limitations in this study. First, our brain network model simulated source signals of the brain network, not the EEG signals themselves in conscious and unconscious states. Instead, we focused on identifying the generic features of brain networks near and far from criticality to study the alteration of the brain network preference for internal/external modes in conscious and unconscious states. Second, we quantitatively evaluated the preferences for internal information and external stimuli without external stimulation. Since we cannot directly

measure the response of unconscious subjects for cognitive tasks, which is beyond the scope of the current study, quantitative evaluations were limited to investigating the relationship between *R*, *SMI*, and  $\chi$  at a system level. However, we expect that the results would be similar to the model results where we have already shown the relationship between *R* and responsiveness with external stimuli (Kim and Lee, 2020). Third, we focused on 4–12 Hz because this frequency band demonstrates global network features in conscious states that are suitable for the application of our brain network model (Zhang et al., 2018). However, how the EEG of 4–12 Hz is globally networked in the cortex and the relevant neuroanatomy is still unclear. Furthermore, higher frequency oscillations—which might be important for information processing—were excluded. Fourth, in our previous model study, we found more specific amplitudes and phases of oscillators that determine large and small responsiveness to external stimuli. However, in this study, we only used the level of synchronization as a criterion to classify the temporal windows due to the complexity of the application to the EEGs of various states of consciousness. Fifth, we limited our fMRI analysis to the comparison of co-activation patterns between high and low *R* windows due to the small data length for measuring critical dynamics in fMRI data. Future studies should directly compare the relationships between *R*, *SMI*, and  $\chi$  in both EEG and fMRI data with longer fMRI data.

## CONCLUSION

Based on both a computational model and empirical data analyses from EEG and fMRI, we propose a novel role for criticality in facilitating continuous switching between internal and external modes in the brain. We found that a large network synchronization fluctuation that emerges near criticality creates continuous switching between distinct network preferences for internal information and external stimuli. Specifically, we found high *SMI* and posterior hub dominant network configuration in high *R* windows, which naturally occur near criticality, play essential roles in maintaining consciousness in conscious states and psychedelic states. The distinct network preferences at high and low levels of synchronization that we found with a canonical networked oscillator model at criticality and EEGs during conscious states supported our argument together with the results of the fMRI co-activation pattern analysis and the EEG analysis of different states of consciousness (ISO, KET, PSY, MCS, and UWS). The computational model study and empirical data analyses lead us to propose that criticality creates a functional platform for network transitions between internal and external modes in the brain, which presumably play an essential role in constructing models of the self and the world.

## DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because participants in these empirical studies did not give consent to have data made publicly available. Requests to access these datasets should be directed to UL, [uclee@umich.edu](mailto:uclee@umich.edu).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Michigan Medical School Institutional Review Board, Institutional Review Board of Medical College of Wisconsin, and Institutional Review Board of the Huashan Hospital, Fudan University. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

UL, MK, and HK designed the research. MK and HK simulated the model. MK analyzed the data. GM, ZH, DJ, and RI curated the data. GM, UL, MK, HK, and ZH wrote the manuscript. All authors reviewed the manuscript.

## FUNDING

This research was supported by the National Institute of General Medical Sciences of the National Institutes

of Health (Bethesda, MD, United States) under award numbers R01GM111293 (ketamine data) and R01GM098578 (computational model), and the James S. McDonnell Foundation, St. Louis (isoflurane data), and by the Department of Anesthesiology, University of Michigan, Ann Arbor, MI, United States.

## ACKNOWLEDGMENTS

The authors thank the Klinikum rechts der Isar, Technische Universität München (Munich, Germany) and the Therapiezentrum Burgau (Burgau, Germany) for providing the MCS and UWS EEG data.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnsys.2021.657809/full#supplementary-material>

## REFERENCES

- Alstott, J., Bullmore, E., and Plenz, D. (2014). Powerlaw: a python package for analysis of heavy-tailed distributions. *PLoS One* 9:85777. doi: 10.1371/journal.pone.0085777
- Bastos, A. M., Vezoli, J., and Fries, P. (2015). Communication through coherence with inter-areal delays. *Curr. Opin. Neurobiol.* 31, 173–180. doi: 10.1016/j.conb.2014.11.001
- Beggs, J. M. (2008). The criticality hypothesis: how local cortical networks might optimize information processing. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 366, 329–343. doi: 10.1098/rsta.2007.2092
- Bressler, S. L., and Tognoli, E. (2006). Operational principles of neurocognitive networks. *Int. J. Psychophysiol.* 60, 139–148. doi: 10.1016/j.ijpsycho.2005.12.008
- Bushdid, C., Magnasco, M. O., Vossell, L. B., and Keller, A. (2014). Humans can discriminate more than 1 trillion olfactory stimuli. *Science* 343, 1370–1372. doi: 10.1126/science.1249168
- Chang, C., Leopold, D. A., Schölvinck, M. L., Mandelkow, H., Picchioni, D., Liu, X., et al. (2016). Tracking brain arousal fluctuations with fMRI. *Proc. Natl. Acad. Sci. U.S.A.* 113, 4518–4523. doi: 10.1073/pnas.1520613113
- Chialvo, D. R. (2006). Psychophysics: are our senses critical? *Nat. Phys.* 2, 301–302. doi: 10.1038/nphys300
- Cocchi, L., Gollo, L. L., Zalesky, A., and Breakspear, M. (2017). Criticality in the brain: a synthesis of neurobiology, models and cognition. *Prog. Neurobiol.* 158, 132–152. doi: 10.1016/j.pneurobio.2017.07.002
- Cohen, D., van Swinderen, B., and Tsuchiya, N. (2018). Isoflurane impairs low-frequency feedback but leaves high-frequency feedforward connectivity intact in the fly brain. *eNeuro* 5:1103923. doi: 10.1523/ENEURO.0329-17.2018
- Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., and Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* 16, 1348–1355. doi: 10.1038/nn.3470
- Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755
- Deco, G., Cabral, J., Saenger, V. M., Boly, M., Tagliazucchi, E., Laufs, H., et al. (2018). Perturbation of whole-brain dynamics in silico reveals mechanistic differences between brain states. *Neuroimage* 169, 46–56. doi: 10.1016/j.neuroimage.2017.12.009
- Deco, G., Cabral, J., Woolrich, M. W., Stevner, A. B. A., van Hartevelt, T. J., and Kringelbach, M. L. (2017). Single or multiple frequency generators in ongoing brain activity: a mechanistic whole-brain model of empirical MEG data. *Neuroimage* 152, 538–550. doi: 10.1016/j.NEUROIMAGE.2017.03.023
- Deco, G., and Jirsa, V. K. (2012). Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. *J. Neurosci.* 32, 3366–3375. doi: 10.1523/JNEUROSCI.2523-11.2012
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011
- Gong, G., He, Y., Concha, L., Lebel, C., Gross, D. W., Evans, A. C., et al. (2009). Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cereb. Cortex* 19, 524–536. doi: 10.1093/cercor/bhn102
- Granada, A., Hennig, R. M., Ronacher, B., Kramer, A., and Herzel, H. (2009). Phase response curves: elucidating the dynamics of coupled oscillators. *Methods Enzymol.* 454, 1–27. doi: 10.1016/S0076-6879(08)03801-9
- Haimovici, A., Tagliazucchi, E., Balenzuela, P., and Chialvo, D. R. (2013). Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Phys. Rev. Lett.* 110:178101. doi: 10.1103/PhysRevLett.110.178101
- Haldeman, C., and Beggs, J. M. (2005). Critical branching captures activity in living neural networks and maximizes the number of metastable states. *Phys. Rev. Lett.* 94:058101. doi: 10.1103/PhysRevLett.94.058101
- Hannay, K. M., Booth, V., and Forger, D. B. (2015). Collective phase response curves for heterogeneous coupled oscillators. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* 92:022923. doi: 10.1103/PhysRevE.92.022923
- Hasselmo, M. E. (1995). Neuromodulation and cortical function: modeling the physiological basis of behavior. *Behav. Brain Res.* 67, 1–27. doi: 10.1016/0166-4328(94)00113-T
- Hinton, G. E., and McClelland, J. L. (1987). *Learning Representations by Recirculation*. *Neural Inf. Process. Syst.* 358–366. Available online at: <http://papers.nips.cc/paper/78-learning-representations-by-recirculation.pdf> (accessed October 12, 2020).
- Honey, C. J., Newman, E. L., and Schapiro, A. C. (2017). Switching between internal and external modes: a multiscale learning principle. *Netw. Neurosci.* 1, 339–356. doi: 10.1162/netn\_a\_00024
- Huang, Z., Zhang, J., Wu, J., Mashour, G. A., and Hudetz, A. G. (2020). Temporal circuit of macroscale dynamic brain activity supports human consciousness. *Sci. Adv.* 6, 87–98. doi: 10.1126/sciadv.aaz0087
- Hudspeth, A. J. (2014). Integrating the active process of hair cells with cochlear function. *Nat. Rev. Neurosci.* 15, 600–614. doi: 10.1038/nrn3786
- Imperatori, L. S., Betta, M., Cecchetti, L., Canales-Johnson, A., Ricciardi, E., Siclari, F., et al. (2019). EEG functional connectivity metrics wPLI and wSMI account



- for distinct types of brain functional interactions. *Sci. Rep.* 9:8894. doi: 10.1038/s41598-019-45289-7
- Kim, H., and Lee, U. (2019). Criticality as a determinant of integrated information  $\Phi$  in human brain networks. *Entropy* 21:981. doi: 10.3390/e21100981
- Kim, H., Moon, J.-Y., Mashour, G. A., and Lee, U. (2018). Mechanisms of hysteresis in human brain networks during transitions of consciousness and unconsciousness: theoretical principles and empirical evidence. *PLoS Comput. Biol.* 14:e1006424. doi: 10.1371/journal.pcbi.1006424
- Kim, M., and Lee, U. (2020). Alpha oscillation, criticality, and responsiveness in complex brain networks. *Netw. Neurosci.* 4, 155–173. doi: 10.1162/netn\_a\_00113
- King, J. R., Sitt, J. D., Faugeras, F., Rohaut, B., El Karoui, I., Cohen, L., et al. (2013). Information sharing in the brain indexes consciousness in noncommunicative patients. *Curr. Biol.* 23, 1914–1919. doi: 10.1016/j.cub.2013.07.075
- Kinouchi, O., and Copelli, M. (2006). Optimal dynamical range of excitable networks at criticality. *Nat. Phys.* 2, 348–352. doi: 10.1038/nphys289
- Kitzbichler, M. G., Smith, M. L., Christensen, S. R., and Bullmore, E. (2009). Broadband criticality of human brain network synchronization. *PLoS Comput. Biol.* 5:1000314. doi: 10.1371/journal.pcbi.1000314
- Lee, H., Golkowski, D., Jordan, D., Berger, S., Ilg, R., Lee, J., et al. (2019). Relationship of critical dynamics, functional connectivity, and states of consciousness in large-scale human brain networks. *Neuroimage* 188, 228–238. doi: 10.1016/j.neuroimage.2018.12.011
- Lee, S. H., and Dan, Y. (2012). Neuromodulation of brain states. *Neuron* 76, 209–222. doi: 10.1016/j.neuron.2012.09.012
- Levnajić, Z., and Pikovsky, A. (2010). Phase resetting of collective rhythm in ensembles of oscillators. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* 82, 056202. doi: 10.1103/PhysRevE.82.056202
- Li, M., Han, Y., Aburn, M. J., Breakspear, M., Poldrack, R. A., Shine, J. M., et al. (2019). Transitions in information processing dynamics at the whole-brain network level are driven by alterations in neural gain. *PLoS Comput. Biol.* 15:e1006957. doi: 10.1371/journal.pcbi.1006957
- Liu, X., De Zwart, J. A., Schölvinck, M. L., Chang, C., Ye, F. Q., Leopold, D. A., et al. (2018). Subcortical evidence for a contribution of arousal to fMRI studies of brain activity. *Nat. Commun.* 9:395.
- Liu, X., Lauer, K. K., Ward, B. D., Rao, S. M., Li, S.-J., and Hudetz, A. G. (2012). Propofol disrupts functional interactions between sensory and high-order processing of auditory verbal memory. *Hum. Brain Mapp.* 33, 2487–2498. doi: 10.1002/hbm.21385
- Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., et al. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12574–12579. doi: 10.1073/pnas.1608282113
- Marinazzo, D., Wu, G., Pellicoro, M., Angelini, L., and Stramaglia, S. (2012). Information flow in networks and the law of diminishing marginal returns: evidence from modeling and human electroencephalographic recordings. *PLoS One* 7:e45026. doi: 10.1371/journal.pone.0045026
- Mhuiricheartaigh, R. N., Warnaby, C., Rogers, R., Jbabdi, S., and Tracey, I. (2013). Slow-wave activity saturation and thalamocortical isolation during propofol anesthesia in humans. *Sci. Transl. Med.* 5:208ra148. doi: 10.1126/scitranslmed.3006007
- Muñoz, M. A. (2018). Colloquium: criticality and dynamical scaling in living systems. *Rev. Mod. Phys.* 90:031001. doi: 10.1103/RevModPhys.90.031001
- Murphy, A. C., Bertolero, M. A., Papadopoulos, L., Lydon-Staley, D. M., and Bassett, D. S. (2020). Multimodal network dynamics underpinning working memory. *Nat. Commun.* 11, 3035–3035. doi: 10.1038/s41467-020-15541-0
- O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: the generalized recirculation algorithm. *Neural Comput.* 8, 895–938. doi: 10.1162/neco.1996.8.5.895
- Raichle, M. E. (2015). The brain's default mode network. *Annu. Rev. Neurosci.* 38, 433–447. doi: 10.1146/annurev-neuro-071013-014030
- Roberts, J. A., Gollo, L. L., Abeyesuriya, R. G., Roberts, G., Mitchell, P. B., Woolrich, M. W., et al. (2019). Metastable brain waves. *Nat. Commun.* 10:1056. doi: 10.1038/s41467-019-08999-0
- Sadaghiani, S., Poline, J. B., Kleinschmidt, A., and D'Esposito, M. (2015). Ongoing dynamics in large-scale functional connectivity predict perception. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8463–8468. doi: 10.1073/pnas.1420687112
- Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., et al. (2009). Early-warning signals for critical transitions. *Nature* 461, 53–59. doi: 10.1038/nature08227
- Schroter, M. S., Spoormaker, V. I., Schorer, A., Wohlschläger, A., Czisch, M., Kochs, E. F., et al. (2012). Spatiotemporal reconfiguration of large-scale brain functional networks during propofol-induced loss of consciousness. *J. Neurosci.* 32, 12832–12840. doi: 10.1523/JNEUROSCI.6046-11.2012
- Shew, W. L., Clawson, W. P., Pobst, J., Karimipour, Y., Wright, N. C., and Wessel, R. (2015). Adaptation to sensory input tunes visual cortex to criticality. *Nat. Phys.* 11, 659–663. doi: 10.1038/nphys3370
- Shine, J. M., Aburn, M. J., Breakspear, M., and Poldrack, R. A. (2018). The modulation of neural gain facilitates a transition between functional segregation and integration in the brain. *Elife* 7:e31130.
- Stam, C. J., and van Straaten, E. C. W. (2012). Go with the flow: use of a directed phase lag index (dPLI) to characterize patterns of phase relations in a large-scale model of brain dynamics. *Neuroimage* 62, 1415–1428. doi: 10.1016/j.neuroimage.2012.05.050
- Tagliazucchi, E., Balenzuela, P., Fraiman, D., and Chialvo, D. R. (2012). Criticality in large-scale brain fMRI dynamics unveiled by a novel point process analysis. *Front. Physiol.* 3:15. doi: 10.3389/fphys.2012.00015
- Tagliazucchi, E., Chialvo, D. R., Siniatchkin, M., Amico, E., Brichant, J.-F., Bonhomme, V., et al. (2016). Large-scale signatures of unconsciousness are consistent with a departure from critical dynamics. *J. R. Soc. Interface* 13:20151027. doi: 10.1098/rsif.2015.1027
- Turchi, J., Chang, C., Ye, F. Q., Russ, B. E., Yu, D. K., Cortes, C. R., et al. (2018). The basal forebrain regulates global resting-state fMRI fluctuations. *Neuron* 97, 940–952.e4. doi: 10.1016/j.neuron.2018.01.032
- Van De Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., et al. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proc. Natl. Acad. Sci. U.S.A.* 111, 87–92. doi: 10.1073/pnas.1312141110
- van den Heuvel, M. P., and Sporns, O. (2013). Network hubs in the human brain. *Trends Cogn. Sci.* 17, 683–696. doi: 10.1016/j.tics.2013.09.012
- Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F., and Pennartz, C. M. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage* 55, 1548–1565. doi: 10.1016/j.neuroimage.2011.01.055
- Vossel, S., Geng, J. J., and Fink, G. R. (2014). Dorsal and ventral attention systems: distinct neural circuits but collaborative roles. *Neuroscientist* 20, 150–159. doi: 10.1177/1073858413494269
- Yoon, S., Sorbaro Sindaci, M., Goltsev, A. V., and Mendes, J. F. F. (2015). Critical behavior of the relaxation rate, the susceptibility, and a pair correlation function in the Kuramoto model on scale-free networks. *Phys. Rev. E* 91:32814. doi: 10.1103/PhysRevE.91.032814
- Zhang, H., Watrous, A. J., Patel, A., and Jacobs, J. (2018). Theta and alpha oscillations are traveling waves in the human neocortex. *Neuron* 98, 1269–1281.e4. doi: 10.1016/j.neuron.2018.05.019

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AM-W declared a past co-authorship with one of the authors GM to the handling editor.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Kim, Kim, Huang, Mashour, Jordan, Ilg and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# EEG-Microstates Reflect Auditory Distraction After Attentive Audiovisual Perception Recruitment of Cognitive Control Networks

Ute Korn<sup>1,2\*†</sup>, Marina Krylova<sup>2,3†</sup>, Kilian L. Heck<sup>1</sup>, Florian B. Häußinger<sup>2,4</sup>, Robert S. Stark<sup>2</sup>, Sarah Alizadeh<sup>2,3</sup>, Hamidreza Jamalabadi<sup>2,5</sup>, Martin Walter<sup>3</sup>, Ralf A. W. Galuske<sup>1</sup> and Matthias H. J. Munk<sup>1,2\*</sup>

<sup>1</sup> Systems Neurophysiology, Department of Biology, Darmstadt University of Technology, Darmstadt, Germany, <sup>2</sup> Department of Psychiatry and Psychotherapy, University of Tuebingen, Tuebingen, Germany, <sup>3</sup> Department of Psychiatry and Psychotherapy, Jena University Hospital, Jena, Germany, <sup>4</sup> NTT DATA Deutschland GmbH, Munich, Germany, <sup>5</sup> Department of Psychiatry and Psychotherapy, Philipps-University, Marburg, Germany

## OPEN ACCESS

### Edited by:

Maurizio Mattia,  
National Institute of Health (ISS), Italy

### Reviewed by:

Pascal Steullet,  
Centre Hospitalier Universitaire  
Vaudois (CHUV), Switzerland  
Yoonsuck Choe,  
Texas A&M University, United States

### \*Correspondence:

Ute Korn  
ute.korn@bio.tu-darmstadt.de  
Matthias H. J. Munk  
matthias.munk@uni-tuebingen.de

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

**Received:** 31 July 2021

**Accepted:** 12 November 2021

**Published:** 09 December 2021

### Citation:

Korn U, Krylova M, Heck KL, Häußinger FB, Stark RS, Alizadeh S, Jamalabadi H, Walter M, Galuske RAW and Munk MHJ (2021) EEG-Microstates Reflect Auditory Distraction After Attentive Audiovisual Perception Recruitment of Cognitive Control Networks. *Front. Syst. Neurosci.* 15:751226. doi: 10.3389/fnsys.2021.751226

Processing of sensory information is embedded into ongoing neural processes which contribute to brain states. Electroencephalographic microstates are semi-stable short-lived power distributions which have been associated with subsystem activity such as auditory, visual and attention networks. Here we explore changes in electrical brain states in response to an audiovisual perception and memorization task under conditions of auditory distraction. We discovered changes in brain microstates reflecting a weakening of states representing activity of the auditory system and strengthening of salience networks, supporting the idea that salience networks are active after audiovisual encoding and during memorization to protect memories and concentrate on upcoming behavioural response.

**Keywords:** electroencephalography, microstates, audiovisual, crossmodal, attention, distraction, resting states, background music

## 1. INTRODUCTION

Background noise is a ubiquitous phenomenon which increasingly infiltrates processes of our daily life, e.g., background music in supermarkets, street noise and traffic noise. It can also be found in current media, e.g., by unwanted advertisements, but also when watching movies and during reports and news, when attention should be focused maximally on relevant information and complex processes. In order to elucidate the effects of background noise several studies investigated students' performance and attention while doing their homework when simultaneously distracting sounds in the background were presented. It was shown that these inhibited their performance depending on the kind of background noise (Furnham and Bradley, 1997; Pool et al., 2000).

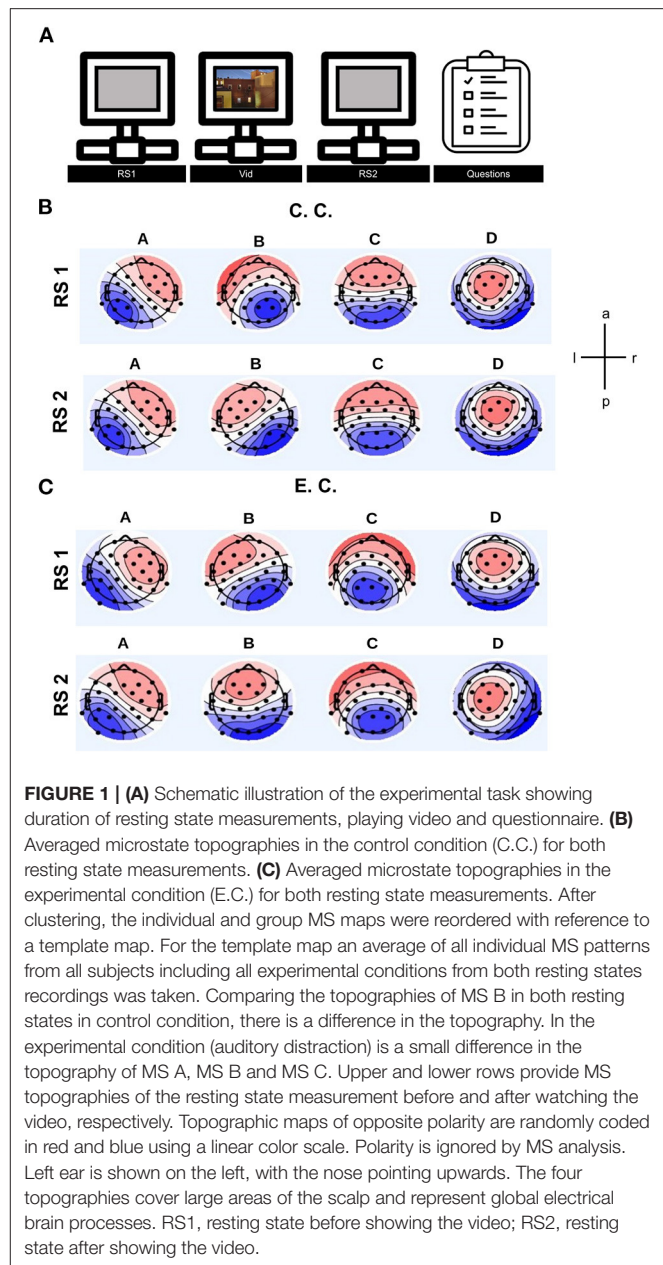
However, the effects of sensory distraction have barely been investigated with respect to their influence on brain states supporting multimodal perception and working memory. To date, much is known about cross-modal interference during stimulus processing (Spence et al., 2000; Calvert, 2001; Weissman et al., 2004), but hardly any data are available on the dynamics of brain states before and after encoding of simultaneously perceived information in different sensory modalities. In particular, the respective resting state activity might be of fundamental importance. Generally speaking, resting state (RS) activity in the electroencephalography (EEG) can be described as a

short time period of semi-stable electric field topographies, called microstates (MS) (Lehmann et al., 1987; Lehmann, 1990). These small numbers of topographies are stable for tens of milliseconds and transition into another semi-stable pattern (Koenig et al., 2002). General data analysis approaches of MS consider duration, occurrence, contribution and transition rate in order to describe MS topographies and commonly at least four different MS topographies can be differentiated (Koenig et al., 2002; Michel and Koenig, 2018). These MS are labelled with letters and each letter reflects one topography: MS A resembles a left frontal and to right occipito-temporal dipole, MS B resembles a right frontal and to left occipito-temporal dipole, MS C resembles an anterior-posterior dipole and MS D resembles a fronto-central and occipital-temporal dipole. Some studies have established relations between MS dynamics and various RS brain networks as revealed by functional magnetic resonance imaging (fMRI) (Britz et al., 2010) or in electrophysiological RS networks (Custo et al., 2017, for Review see Michel and Koenig, 2018).

Moreover, Weissman et al. (2004) investigated cross-modal interference between visual and auditory stimuli using fMRI and found that the sensory cortices play novel roles in increasing attention to goal-relevant stimuli. More precisely, the dorsolateral prefrontal cortex (DLPFC) and the anterior cingulate cortex (ACC) are involved in increasing attention when distracting auditory stimuli conflict with visual stimulation (Weissman et al., 2004). Seeley et al. (2007) identified the salience network, which is anchored in dorsal anterior cingulate and orbital fronto-insular cortices (Seeley et al., 2007). This network is being involved in controlling and coordinating domains of cognitive control, e.g., attention, working memory, inhibition and planning (Seeley et al., 2007; Niendam et al., 2012; Breukelaar et al., 2017). Britz et al. (2010) found that MS C is associated with the salience network (Britz et al., 2010). Since certain MS reflect network activity in the brain areas found by Weissman et al. (2004), e.g., ACC and DLPFC, it is most promising that we choose MS analysis as evaluation method for the RS recordings in our cross-modal task (Weissman et al., 2004).

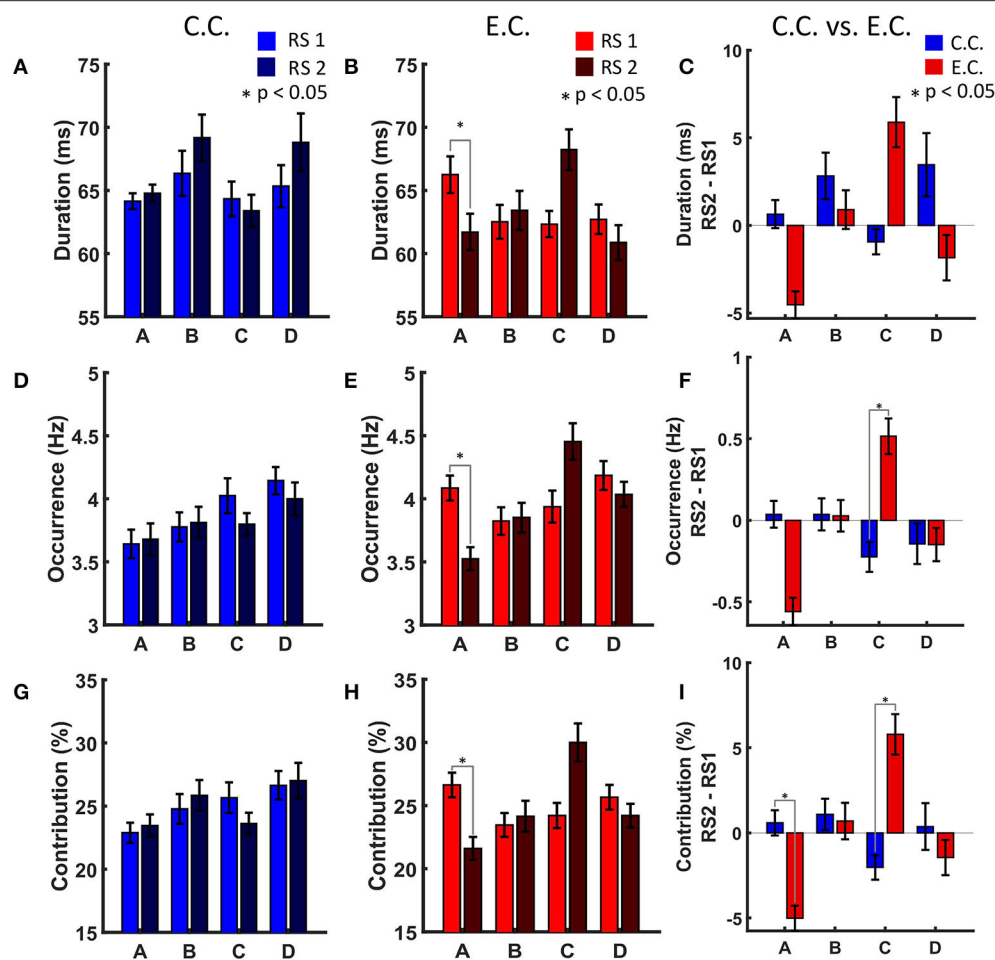
We wanted to investigate the influence of auditory interference in an audio-visual, attention-demanding paradigm to get a sense of the day-to-day environmental influences on the brain. Therefore, we investigated the effect of distraction by auditory background noise rather than visual noise interfering with auditory information. For this purpose, we developed a paradigm in which subjects were instructed to watch a video and memorise as many auditory and visual items as possible in order to recall content later in a questionnaire, not knowing that the effect of auditory distraction would be the target of the project. To this end, we prepared an audiovisual documentation which was presented to the subjects with or without background “noise” under varying levels of impact (experimental condition: auditory distraction, control condition: without auditory distraction) while measuring brain states from EEG recordings (Figure 1A). MS analyses were applied to two RS measurements, one recorded before and one after video presentation.

Based on previous work (Weissman et al., 2004), we propose that auditory distraction by loud background noise should modify ongoing brain activity captured in RS recordings



after active perception and encoding of the audio-visual task. Topographically, two MS are located frontally and occipitally, but laterally shifted [clockwise (MS A)/counterclockwise (MS B)]. Both include parts of the visual and auditory cortex. For this reason, we assume that our audio-visual paradigm with strong auditory background distraction has an impact on these two MSs.

Furthermore, we predict an increase in microstate parameters (contribution, occurrence and duration) and their transition rates in the experimental condition (with auditory distraction) of MS C, which seems to reflect the salience network or parts of the intrinsic default mode network (Britz et al., 2010; Custo et al., 2017; Michel and Koenig, 2018). Topographically, MS C shows opposite patterns between frontoparietal and occipital regions,



**FIGURE 2 |** MS parameters duration, occurrence and contribution. **(A)** MS parameter duration in the control condition. **(B)** MS parameter duration in the experimental condition with auditory distraction. MS A shows a significant decrease ( $p < 0.05$ ) in RS2. **(C)** MS parameter duration in comparison of both experimental conditions. Light colour indicates resting state recording before showing the video. Dark colour indicates resting state recording after showing the video. **(D)** MS parameter occurrence in the control condition. **(E)** MS parameter occurrence in the experimental condition. MS A shows a significant decrease ( $p < 0.05$ ) in RS2. **(F)** MS parameter occurrence in comparison of both experimental conditions. MS C shows a significant increase ( $p < 0.05$ ) after auditory distraction. **(G)** MS parameter contribution in the control condition. **(H)** MS parameter contribution in the experimental condition. MS A shows a significant decrease ( $p < 0.05$ ) in RS2. **(I)** MS parameter contribution in comparison of both experimental conditions. MS A shows a significant decrease ( $p < 0.05$ ) in RS2. MS C shows a significant increase ( $p < 0.05$ ) after auditory distraction. Fourteen Subjects were analysed for the experimental condition and 13 Subjects for the control condition. The results of the RS recordings were subtracted from each other to obtain a RS independent comparison between the two conditions. Blue, control condition (C.C.); red, experimental condition (E.C.); x-axis, MS classes; y-axis, each MS parameter. Error bars indicate standard error of the mean (\* indicates  $p < 0.05$ ).

which include the visual cortex, DLPFC, posterior cingulate cortex (PCC), precuneus and ACC (Britz et al., 2010; Custo et al., 2017). Regarding the behavioural test results, we propose that auditory background distraction would have a negative impact on subjects' recall performance (Furnham and Bradley, 1997; Pool et al., 2000).

## 2. METHODS

### 2.1. Paradigm

The background-music paradigm was developed to test the influence of auditory distraction on brain electrical states. Subjects were shown a TV feature about global inequality (poverty, wealth gap) of 5 min duration. The control group

(C.C.) watched the video with informative and non-vocal, decent background sounds, while the experimental group (E.C.) watched the video with informative sound and mostly vocal, loud and popular music (however, the lyrics mostly fitted well with current information context). We considered this musical background as distracting. Subjects were instructed at the beginning of the experiment to attentively watch and listen, as they would subsequently have to answer questions about the content. The RS recordings (5 min) were taken just before and immediately after showing the video (Figure 1A). During RS recording, subjects fixed their gaze at the centre of a grey screen with eyes open. Subjects were assigned to the groups (C.C. and E.C.) in a pseudo-randomised manner.

Subjects were asked eight questions about the content of the video (Questionnaire is available in **Supplementary Section 1.1**). Here, 6 binary items and 2 multinomial items were used. For subsequent data analysis operations, these multinomial variables were binarized (dummy-coding) for ensuring comparability. One questionnaire from the experimental group with auditory distraction was lost, only 13 questionnaires could be evaluated for this group.

## 2.2. Subjects

30 subjects volunteered to perform our experiment. Written informed consent was obtained from all participants. After pre-processing, data sets from 13 subjects who performed the control condition and from 14 subjects who performed the experimental condition qualified for further analysis. The age range of all subjects was 22–33 years (mean = 25, SD = 3.1, median = 24). In the control condition 5 male and 8 female subjects and in the experimental condition 7 male and 7 female subjects were measured.

## 2.3. Registering Brain Activity

30 EEG-electrodes (actiCap, Brain Products GmbH, Germany) were placed in standard positions (10-20-system) on the head (see **Supplementary Figure S1**). Electrode impedance was improved until  $\leq 5 \text{ k}\Omega$ . Data were sampled with 250 or 500 Hz.

## 2.4. EEG Pre-processing

Pre-processing functions of the MATLAB toolbox EEGLAB (v. 2019\_1) were used to cope with known signal components related to biological / technical noise, e.g., eye blinks and line noise (Delorme and Makeig, 2004). Each dataset was first bandpass filtered between 0.3 and 200 Hz and subsequently segmented into epochs of 1 s duration. Also, muscle and motion artefacts as well as open ( $> 120 \mu\text{V}$ ) and flat channels ( $< 5 \mu\text{V}$ ) were removed.

Furthermore, independent component analysis (ICA) was employed to extract signal components reflecting eye movement and continuous muscle activity (Cohen, 2014). Components can be interpreted based on their topographies, time courses and frequency spectra. Components which contain flat dynamics, sporadic high-amplitude spikes and were spatially located in anterior brain regions were removed as eye blinking. Whereas, components which contain bursts between 20 to 40 Hz activity and with a high amplitude, located near the face, neck or ears were removed as muscle activity (Cohen, 2014).

## 2.5. Microstate Analysis

The EEGLAB plugin for microstates (v1.1) by Thomas Koenig<sup>1</sup> was used for MS analysis. Pre-processed EEG data were further bandpass-filtered between 2 and 20 Hz and re-referenced to average reference. Due to the fact that in some experiments sampling rates were 500 Hz and in others 250 Hz, we reduced sampling rate for all data sets to 250 Hz to ensure comparability. The entire dataset consisted of recordings with one or the other experimental/control condition (C.C./E.C.) and two RS recordings (RS1/RS2) per subject. We analysed exactly 5 min of

both RS recordings, as this was the smallest common recording duration in all subjects.

In accordance with Murray et al. (2008), we calculated global field power peaks (GFP) for each epoch, which constitutes a reference-independent measure of response strength (Murray et al., 2008). Subsequently, we followed the methodology of Krylova et al. (2021) for MS analysis, for more details see **Supplementary Section 1.2** (Krylova et al., 2021).

We derived the following dynamic parameters for each MS providing averaged (arithmetic mean) values:

- duration (period in which consecutive maps were assigned to the same MS class),
- occurrence (mean number of MS per second),
- contribution (percentage time occupied in each MS) and finally
- transition rate (original/delta) (Koenig et al., 2002).

Contribution is a percentage representation of the MSs over the period of the recording duration. The transition rate between all MSs was expressed as probability with respect to all possible transitions (original transition rate). The original transition rate was then related to the theoretical transition model of Lehmann et al. (2005), the latter providing the observed deviation from equal transition probability for all possible transitions (delta transition rate) (Lehmann et al., 2005).

## 2.6. Statistical Analysis

Statistical analysis was performed in MATLAB [version: 9.5.0.1178774, (2018b), Natick, Massachusetts: The MathWorks Inc.].

To test for equal medians between C.C. and E.C., the two-sided Wilcoxon rank sum test was used (function *ranksum*), which is very close to the Mann-Whitney U test (MWU). Wilcoxon signed rank tests (function *signrank*) were conducted for testing RS1 and RS2 within each separate experimental condition (Hollander and Douglas, 1999; Gibbons and Chakraborti, 2010). For further details see **Supplementary Section 1.3**.

The Bonferroni Correction was used to adjust for multiple comparisons of the MS parameters duration, occurrence and contribution, whereas the False Discovery Rate (FDR) was used for the delta and original transition rate of MSs (Genovese et al., 2002; Cohen, 2014; Fachada and C. Rosa, 2018).

For the evaluation of the questionnaires, we used non-parametric statistical methods (Mann-Whitney U test,  $\chi^2$ -test) for assessing the differences of central tendencies between control and experimental groups.

## 3. RESULTS

We analysed 30-channel resting state EEG-recordings of 27 subjects before and after they memorised visual and auditory objects which they were asked for in a questionnaire. Both groups (C.C.  $n = 13$ / E.C.  $n = 14$ ) performed indistinguishably well, on average recalling 8 items (C.C.: 8.2; E.C.: 7.6), the two groups

<sup>1</sup><http://www.thomaskoenig.ch/index.php/software/microstates-in-eeqlab>



differed only by a non-significant trend ( $\chi^2$ -test:  $p = 0.77$ ; MWU-test: tied- $p = 0.49$ ) that distracted observers tended to memorise less well.

After clustering and reordering steps, the final topographies of all MSs were estimated. The averaged topographies per condition and resting state recording were analysed (Figures 1B,C). Further description and discussion about the topographies is available in the Supplementary Sections 1.4 and 1.5.

### 3.1. MS Parameters and Transition Rates of the C.C.

We could hardly see any significant changes in MS parameters for the control group after watching and memorising the video in comparison of the two RS recordings (Figures 2A,D,G). No significant results were also observed for MS transitions in the control condition (Figure 3A, as we have already illustrated in Krylova et al., 2021).

### 3.2. MS Parameters and Transition Rates of the E.C.

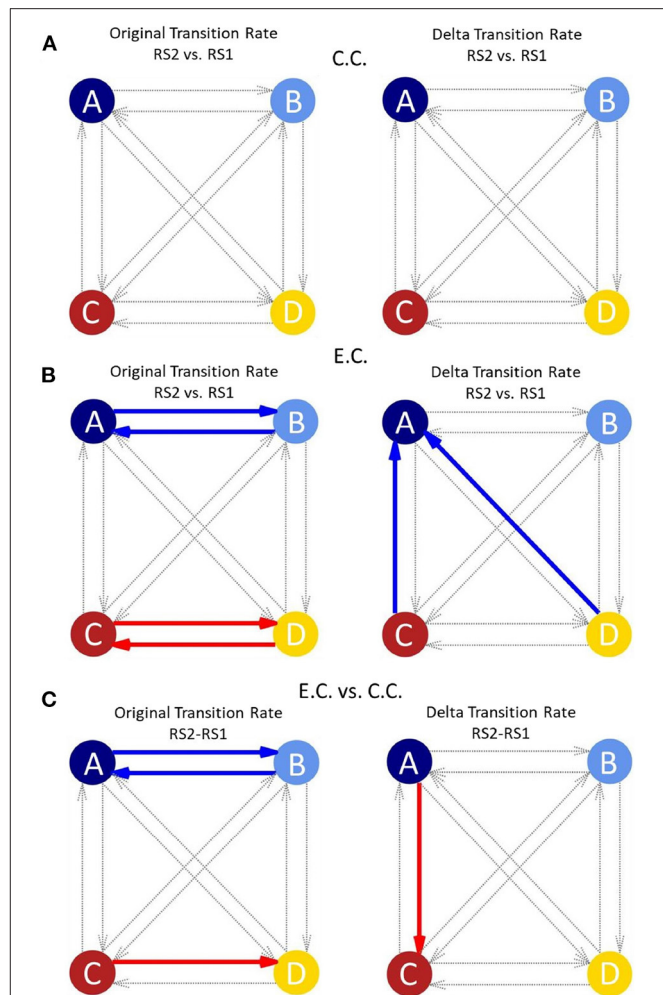
We found that distracting music after active audiovisual perception had a significant impact on all MS parameters of two of the four investigated MSs. Statistical analysis revealed that these effects occurred both for MS A and MS C. MS A showed a significant decrease for the parameter's duration, occurrence and contribution during RS2 as compared to RS1 for the E.C. (Figures 2B,E,H).

The relative frequencies of transitions between different MSs provide information on the overall timing of brain states and preponderance of networks: here, MS A, MS C and MS D showed significant changes in transition rate, which indicates the timing of the MSs (Figure 3B). This allows for statements to be made about which MS occurs more often and whether a transition to a specific MS is preferred. As displayed in Figure 3B on the left, the original transition rate shows significant results for the transitions between MS A and MS B and for the transitions between MS C and MS D during RS2 as compared to RS1 for the E.C.

Transitions from MS C to MS A and from MS D to MS A show significant results for the delta transition rate during RS2 as compared to RS1 in the E.C. (Figure 3B on the right). The delta transition rate is estimated as the difference between the observed and the theoretical transition rate (Lehmann et al., 2005).

### 3.3. MS Parameters and Transition Rates in Comparison Between C.C. and E.C.

Comparing E.C. and C.C. (with RS2 minus RS1) auditory distraction influences MS C more than MS A (Figures 2C,F,I). Contribution of MS A shows a significant decrease (Figure 2I), while occurrence and contribution of MS C show significant increases for the condition with background music distraction (Figures 2F,I). For the original transition rate we found significant decreases for the transitions between MS A and MS B (in both directions) and a significant increase for the transitions



**FIGURE 3 | (A)** Network graphs of the transition rate (original/delta) of the four canonical MSs for the control condition comparing the two resting state measurements (modified from Krylova et al., 2021). In the control condition, no significant results could be obtained for any of the transition rates between MSs (thin grey arrows). The original transition rate is influenced by the occurrence of MSs, whereas, the delta transition rate is independent of the occurrence of MSs. **(B)** Network graphs of the transition rate (original/delta) of the four canonical MSs for the experimental condition (with auditory distraction) comparing the two resting state measurements. Clearly differences between the original and delta transition rate are visible. Original transition rate shows significant shifts (coloured thick arrows) between MS A and MS B and between MS C and MS D. The delta transition rate shows significant shifts from MS C and MS D to MS A. **(C)** Network graphs of the transition rate (original/delta) in comparison of the experimental and control condition and the resting state recordings. There are differences for the transitions between the original and delta transition rates. The original transition rate shows significant shifts between MS A and MS B and from MS C to MS D. Whereas the delta transition rate only shows a significant shift from MS A to MS C. Each circle indicates one MS (A–D); red arrows: increased transition rate; blue arrows: decreased transition rate; thick arrows:  $p$ -value between 0.05 and 0.01; dotted grey arrows: transitions between MSs with  $p > 0.05$ ; E.C., experimental condition; C.C., control condition; RS1, resting state before showing the video; RS2, resting state after showing the video; MS, Microstate.

between MS C to MS D (also in both directions in Figure 3C on the left). The delta transition rate shows a significant

increase for the transitions from MS A to MS C (**Figure 3C** on the right).

## 4. DISCUSSION

Distracting auditory stimuli impact MSs occurrence and contribution during RS2 while keeping visual and auditory items in memory. To our surprise, MS A occurred less frequently and was shorter when distracting music interfered during encoding, while MS C increased in duration and occurrence. It was not expected that the audiovisual paradigm had no impact on the parameters of MS B.

### 4.1. Cross-Modal Interaction Effects on MS A

In MS A, the topography is slightly shifted and covers upon distraction more temporal areas, as shown in the red- and blue-coloured patterns in **Figures 1B,C**. The primary areas of the visual system and the auditory system are located in these regions.

Weissman et al. (2004) demonstrated that the DLPFC was active when subjects had to solve a cross-modal task with distraction. The DLPFC is located in the active areas of MS A and MS B, given their topography. Weissman et al. (2004) created two parts of the cross-modal task, one with visual stimuli and auditory distraction and the other with auditory stimuli and visual distraction. When participants had to solve a visual task while listening to distracting auditory input, they found only activation of the left DLPFC for this type of task. Conversely, for an auditory task with visually distracting stimuli, no significant activation of the right DLPFC was found (Weissman et al., 2004). We expected for our paradigm to see stronger impact on MS B representing activity in the visual system in RS2. Nevertheless, we saw a decrease of MS parameters for MS A. The DLPFC is involved in directing attention to task-relevant stimuli (Weissman et al., 2004). Although we only evaluated the RS before and after the attention-demanding task, the DLPFC seems to be active. Therefore, the demands on DLPFC processes appear to be higher when attention is directed towards a visual process with a concurrent auditory distracting component (Weissman et al., 2004). More plausible seems to be the interpretation of Britz et al. (2010), who associated MS A with the auditory network (Britz et al., 2010). Thus, the decrease in duration, occurrence and contribution during RS2 could also be due to fatigue or adaptation of this network after auditory distraction, in particular, since these effects were not observed in RS2 of the C.C. in which only informative sound was provided.

### 4.2. MS C and the Cognitive Control Network

As several studies reported, MS C is associated with parts of the cognitive control network (Seeley et al., 2007; Britz et al., 2010; Custo et al., 2017). Britz et al. (2010) associated MS C with a fMRI RS network, named salience network (Britz et al., 2010). The salience network provides functions such as attention, planning or working memory in order to enable appropriate behaviour

or achieve a specific goal (Breukelaar et al., 2017). In contrast to Britz et al. (2010) and Custo et al. (2017) estimated sources of the MS topographies with topographic electrophysiological state source-imaging (Britz et al., 2010; Custo et al., 2017). They determined seven mean MS patterns in their dataset, where MS C was split into two spatially correlated MS topographies (MS C and MS F) (Custo et al., 2017). Generators of MS C localised in the PCC and the precuneus, whereas MS F was generated in the dorsal ACC extending to the superior frontal gyrus, the bilateral middle frontal gyrus and bilateral insula. They identified these two MSs as part of the intrinsic default mode network, which deals with various aspects of self-related information (self-related evaluations, voluntary actions, episodic memory and planning) (Custo et al., 2017). For an extensive review, see Salomon et al. (2014).

After RS2 recording, no significant differences were found between C.C. and E.C. groups in questionnaire performance. However, brain areas and networks relevant for task solving were also associated with MS C and were activated after auditory distraction (Niendam et al., 2012; Breukelaar et al., 2017). According to our results, this seems to have taken up more resources for the video with auditory distraction than for the C.C.. The topography of MS C also supports this finding, as fMRI studies have found that cognitive control is induced by simultaneous activation of frontal and parietal cortex. Involved are the dorsal anterior cingulate cortex, the dorsolateral prefrontal cortex and the dorsal/posterior parietal cortex (Niendam et al., 2012; Breukelaar et al., 2017). Britz et al. (2010) found a relation between MS C and positive BOLD activations in the posterior part of the bilateral ACC, the bilateral inferior frontal gyri, the right anterior insula and the left claustrum (Britz et al., 2010). Other authors assume that MS C is associated with a portion of the anterior default mode network and not with the salience network (Custo et al., 2017; Seitzman et al., 2017; Michel and Koenig, 2018). Custo et al. (2017) suggested that the salience network described by Britz et al. (2010) rather corresponds to their MS F (Britz et al., 2010; Custo et al., 2017). They suggested that if only four microstate maps were used for the clustering procedure, MS C should be seen as a combination of the topographic patterns C and F, including the anterior and posterior cingulate cortices (Custo et al., 2017).

### 4.3. Transition Rates Show Interaction Between Salience and Attention Network After E.C.

Comparing the two transition rates in **Figures 3B,C** (original/delta), we noticed that the transitions in the original rate (left column) were dominated by transitions between MS A and MS B, or MS C and MS D. In the delta transition rate (right column), we observed differences of transitions from MS D and MS C to MS A in the E.C. and from MS A to MS C in comparison of the C.C. and E.C.. This suggests that the two transition rates should be considered separately. The delta transition rate was being calculated in combination with the theoretical transition model of Lehmann et al. (2005) and is

independent of contribution and occurrence of the respective MSs (Lehmann et al., 2005).

As described above, the original transition rate decreases in RS2 for the transitions between MS A and MS B after auditory distraction (**Figure 3B**, left plot, blue arrows). Previous studies associated MS A with activity in the auditory and MS B with activity in the visual network (Britz et al., 2010; Custo et al., 2017; Seitzman et al., 2017). This suggests that the transition decrease between MS A and MS B (**Figure 3B**, left plot, blue arrows) reflects that auditory distraction override visual input. However, this interpretation is in contrast to the results of the other MS parameters. Here, a lower duration, occurrence and contribution for MS A can be seen. The original transition rate is influenced by the occurrences and contribution of the MSs and reflects the decrease of these parameters and this may also be an indication of fatigue in the auditory network.

A significant increase in the original transition rate in RS2 was also found for the transitions between MS C and MS D after auditory distraction (**Figure 3B**, left plot, red arrows). As mentioned above, these MSs are associated with the salience (MS C) and attention network (MS D) (Britz et al., 2010; Seitzman et al., 2017). Britz et al. (2010) found a correlation between MS D topography and ventral fronto-parietal areas being involved in switching and focusing of attention (Corbetta and Shulman, 2002; Britz et al., 2010). Thus, the increased transitions between these two MSs could reflect alternating activation of these two resting state networks, which corresponds to our hypotheses. As mentioned above Custo et al. (2017) split MS C in two separate topographies, where only one corresponds to the salience network (Britz et al., 2010; Custo et al., 2017). Mishra et al. (2020) found that transitions between MSs are continuous and do not necessarily represent discrete transitions between MSs. Therefore, they suggested MS C as an interim state and not a main MS pattern (Mishra et al., 2020). The assumption of both studies relates to our results, as it argues for a close interaction between MS C and MS D.

The delta transition rate shows significant decreases (right column in **Figure 3B**) between the transitions from MS C to MS A and from MS D to MS A (vertical and oblique arrows, respectively). Some studies associated MS D with the dorsal attention network (Britz et al., 2010; Custo et al., 2017; Seitzman et al., 2017). This network is relevant for detecting behaviourally relevant stimuli. The topography of MS D includes more dorsal areas in the fronto-parietal cortex which are involved in switching and reorientation of attention (Corbetta and Shulman, 2002; Britz et al., 2010). MS C is associated with the salience network or the intrinsic default mode network (Britz et al., 2010; Custo et al., 2017). The decrease in the transition rate to the auditory network (MS A) and the decrease of MS parameters duration, contribution and occurrence of MS A, indicate that auditory influences are actively limited by the salience/ intrinsic default mode network (MS C) and attention network (MS D).

#### 4.4. Transition Rates Are Influenced by the Salience and Attention Networks

The influence of the audiovisual paradigm on RS activity can be unravelled when comparing E.C. and C.C.. The results of

the RS recordings were subtracted from each other to obtain a RS independent comparison between the two conditions. Transitions between MS A and MS B show significant decreases in the original transition rate after auditory distraction (**Figure 3C**, left plot, blue arrows). This suggests that the auditory distraction negatively affects the transitions between the two MSs. In contrast to the RS following the auditory distraction, only the transitions from MS C to MS D are significantly increased for the original transition rate in the comparison between E.C. and C.C. (**Figure 3C**, left plot, red arrows). Thus, the attention network seems to be more active, suggesting that the transitions between MS A and MS B are actively influenced by auditory distraction.

Comparison of the two conditions revealed significantly higher transition rates from MS A to MS C, using the delta transition rate (**Figure 3C**, right plot, red arrow). These result again suggest involvement of the salience network: we proposed that the salience network most likely actively filters out auditory distraction. This interpretation is also more plausible with regard to the other MS parameters, as the occurrence and contribution show an increase for MS C, while MS A showed a decrease in contribution (**Figure 2I**).

In summary, the salience, attention and intrinsic default mode networks are activated by our paradigm and appear to compensate for auditory distraction in order to protect memory content for the successful completion of the questionnaire. We would rather not assume that those networks are actively blocking any further influence from new sensory input in favour of memory maintenance, as no significant differences in questionnaire performance were seen between C.C. and E.C.. This is also underlined by the comparison between RS1 and RS2 in the C.C., which showed no significant differences in terms of MS parameters.

## 5. CONCLUSION

Distracting music during active audiovisual perceptions had a significant impact on microstates. In particular we observed effects on MS A and MS C, reflecting activity changes in the auditory system and in the salience or intrinsic default mode network. The control group expressed hardly any changes in MS parameters as a consequence of watching and memorising the video. The experimental group, being exposed to auditory distraction, showed effects on both MS A and MS C, the latter effects being more pronounced. The uniformity of response among subjects in the experimental groups is a strong indication that these measurements are reliable. The most plausible interpretation is that brain activity related to auditory system activity (MS A) is fatigued by auditory distraction. Furthermore, the salience network and the attention network are active in order to memorise the answers for the following questionnaire and to minimise the influences of the distracting background music. Increased activity of the salience network combined with changes in transition rates between MS C and MS A and between MS D and MS A suggests active suppression of auditory distraction. Thus, MS analysis can provide important information about the cerebral processes that occur in everyday life.



## DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available as they are being used for further analysis. Requests to access the datasets should be directed to Matthias H. J. Munk, [matthias.munk@uni-tuebingen.de](mailto:matthias.munk@uni-tuebingen.de), Ute Korn, [ute.korn@bio.tu-darmstadt.de](mailto:ute.korn@bio.tu-darmstadt.de).

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

FH and MM designed research. FH, UK, MM, and RG acquired the data. MK, KH, SA, RS, and HJ provided hardware/software support for data acquisition and analysis. UK and MK analysed the data. KH provided support for statistical analysis. UK

prepared the first draft of the manuscript. MK, KH, RG, and MM corrected and improved the draft. All authors have read, commented, and approved the final version of the manuscript prior to submission.

## FUNDING

UK was funded by the Hubert Markl-scholarship of the Carlo and Karin Giersch Foundation at Technische Universität Darmstadt.

## ACKNOWLEDGMENTS

The authors thank for excellent technical help from B. Schopp, R. Täglic, and K. Wehner. Also special thanks to Dr. D. Hofmann for the support of EEG preprocessing.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnsys.2021.751226/full#supplementary-material>

## REFERENCES

- Broekelaar, I. A., Antees, C., Grieve, S. M., Foster, S. L., Gomes, L., Williams, L. M., et al. (2017). Cognitive control network anatomy correlates with neurocognitive behavior: a longitudinal study. *Hum. Brain Mapp.* 38, 631–643. doi: 10.1002/hbm.23401
- Britz, J., Van De Ville, D., and Michel, C. M. (2010). Bold correlates of eeg topography reveal rapid resting-state network dynamics. *Neuroimage* 52, 1162–1170. doi: 10.1016/j.neuroimage.2010.02.052
- Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cereb. Cortex* 11, 1110–1123. doi: 10.1093/cercor/11.12.1110
- Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. Cambridge, MA: MIT Press.
- Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755
- Custo, A., Van De Ville, D., Wells, W. M., Tomescu, M. I., Brunet, D., and Michel, C. M. (2017). Electroencephalographic resting-state networks: Source localization of microstates. *Brain Connect.* 7, 671–682. doi: 10.1089/brain.2016.0476
- Delorme, A., and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Fachada, N., and Rosa, C. A. (2018). micomp: a matlab/octave toolbox for multivariate independent comparison of observations. *J. Open Source Softw.* 3, 430. doi: 10.21105/joss.00430
- Furnham, A., and Bradley, A. (1997). Music while you work: The differential distraction of background music on the cognitive test performance of introverts and extraverts. *Appl. Cogn. Psychol.* 11, 445–455. doi: 10.1002/(SICI)1099-0720(199710)11:5<445::AID-ACP472>3.0.CO;2-R
- Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15, 870–878. doi: 10.1006/nimg.2001.1037
- Gibbons, J. D., and Chakraborti, S. (2010). *Nonparametric Statistical Inference*. New York, NY: Chapman and Hall/CRC. doi: 10.1201/9781439896129
- Hollander, M., and Douglas, W. A. (1999). *Nonparametric Statistical Methods*. Chichester: Wiley-Interscience.
- Koenig, T., Prichep, L., Lehmann, D., Sosa, P. V., Braeker, E., Kleinlogel, H., et al. (2002). Millisecond by millisecond, year by year: normative eeg microstates and developmental stages. *Neuroimage* 16, 41–48. doi: 10.1006/nimg.2002.1070
- Krylova, M., Alizadeh, S., Izyurov, I., Teckentrup, V., Chang, C., van der Meer, J., et al. (2021). Evidence for modulation of eeg microstate sequence by vigilance level. *Neuroimage* 224:117393. doi: 10.1016/j.neuroimage.2020.117393
- Lehmann, D. (1990). "Brain electric microstates and cognition: the atoms of thought," in *Machinery of the Mind: Data, Theory, and Speculations About Higher Brain Function*, eds E. R. John, T. Harmony, L. S. Prichep, M. Valdés-Sosa, and P. A. Valdés-Sosa (Birkhäuser Boston, Boston, MA), 209–224.
- Lehmann, D., Faber, P. L., Galderisi, S., Herrmann, W. M., Kinoshita, T., Koukkou, M., et al. (2005). Eeg microstate duration and syntax in acute, medication-naïve, first-episode schizophrenia: a multi-center study. *Psychiatry Res. Neuroimaging* 138, 141–156. doi: 10.1016/j.pscychres.2004.05.007
- Lehmann, D., Ozaki, H., and Pal, I. (1987). Eeg alpha map series: brain microstates by space-oriented adaptive segmentation. *Electroencephalogr. Clin. Neurophysiol.* 67, 271–288. doi: 10.1016/0013-4694(87)90025-3
- Michel, C. M., and Koenig, T. (2018). Eeg microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review. *Neuroimage* 180, 577–593. doi: 10.1016/j.neuroimage.2017.11.062
- Mishra, A., Englitz, B., and Cohen, M. X. (2020). Eeg microstates as a continuous phenomenon. *Neuroimage* 208:116454. doi: 10.1016/j.neuroimage.2019.116454
- Murray, M. M., Brunet, D., and Michel, C. M. (2008). Topographic erp analyses: a step-by-step tutorial review. *Brain Topogr.* 20, 249–264. doi: 10.1007/s10548-008-0054-5
- Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., and Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn. Affect. Behav. Neurosci.* 12, 241–268. doi: 10.3758/s13415-011-0083-5
- Pool, M. M., van der Voort, T. H. A., Beentjes, J. W. J., and Koolstra, C. M. (2000). Background television as an inhibitor of performance on easy and difficult homework assignments. *Commun. Res.* 27, 293–326. doi: 10.1177/009365000027003002
- Salomon, R., Levy, D. R., and Malach, R. (2014). Deconstructing the default: cortical subdivision of the default mode/intrinsic system during self-related processing. *Hum. Brain Mapp.* 35, 1491–1502. doi: 10.1002/hbm.22268
- Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci.* 27, 2349–2356. doi: 10.1523/JNEUROSCI.5587-06.2007



- Seitzman, B. A., Abell, M., Bartley, S. C., Erickson, M. A., Bolbecker, A. R., and Hetrick, W. P. (2017). Cognitive manipulation of brain electric microstates. *Neuroimage* 146, 533–543. doi: 10.1016/j.neuroimage.2016.10.002
- Spence, C., Ranson, J., and Driver, J. (2000). Cross-modal selective attention: On the difficulty of ignoring sounds at the locus of visual attention. *Percept. Psychophys.* 62, 410–424. doi: 10.3758/BF03205560
- Weissman, D. H., Warner, L., and Woldorff, M. (2004). The neural mechanisms for minimizing cross-modal distraction. *J. Neurosci.* 24, 10941–10949. doi: 10.1523/JNEUROSCI.3669-04.2004

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Korn, Krylova, Heck, Häußinger, Stark, Alizadeh, Jamalabadi, Walter, Galuske and Munk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Holographic Duality and the Physics of Consciousness

Uziel Awret\*

Inspire Institute, Alexandria, VA, United States

This paper introduces a novel dual-aspect theory of consciousness that is based on the principle of holographic-duality in modern physics and explores the prospects of making philosophically significant empirical discoveries about the physical correlates of consciousness. The theory is motivated by an approach that identifies certain anti-physicalist problem intuitions associated with representational content and spatial location and attempts to provide these with a consciousness-independent explanation, while suspending questions about the hard problem of consciousness and the more problematic “phenomenal character”. Providing such topic neutral explanations is “hard” enough to make a philosophical difference and yet “easy” enough to be approached scientifically. I will argue that abstract algorithms are not enough to solve this problem and that a more radical “computation” that is inspired by physics and that can be realized in “strange metals” may be needed. While speculative, this approach has the potential to both establish necessary connections between structural aspects of conscious mental states and the physical substrate “generating” them and explain why this representational content is “nowhere to be found”. I will end with a reconsideration of the conceivability of zombies.

## OPEN ACCESS

### Edited by:

Timothy L. Hubbard,  
Arizona State University, United States

### Reviewed by:

Ron Chrisley,  
University of Sussex, United Kingdom  
Rafi Malach,  
Weizmann Institute of Science, Israel

### \*Correspondence:

Uziel Awret  
awretu@gmail.com

**Received:** 25 March 2021

**Accepted:** 20 January 2022

**Published:** 15 March 2022

### Citation:

Awret U (2022) Holographic  
Duality and the Physics  
of Consciousness.  
*Front. Syst. Neurosci.* 16:685699.  
doi: 10.3389/fnsys.2022.685699

**Keywords:** holography, duality, coding, meta-problem, physical theory of consciousness, strange metals, dual-aspect

*But if there is a kind of inherent necessary connection between being the kind of cognitive system we are and possession of the requisite experience, then the problem seems much less serious. What kind of metaphysics this would involve is not clear. The idea would be that there is something in the nature of intentional “scripts” of the kind we instantiate that necessitates an experiential reflection of it — a “script” that is necessarily produced as a movie. In a way, this might be a form of “panprotopsychism” as applied to certain informational states.*

Levine (2019)

## PROLOG

We seem to have good reasons to believe that consciousness is physical and good reasons to believe that it is not. This paradoxical situation is referred to by Stoljar (2009) as our “Epistemic situation” (or predicament) and following Post-Newtonians like Priestly who claimed that we don’t know enough about the physical to conclude that consciousness is not physical, Stoljar solution to this epistemic dilemma is to separate the consciousness problem in two, a scientific empirical one (Stoljar, 2006), and a philosophical one aiming to explain our anti-physicalist intuitions. Explaining away such problem intuitions is crucial to Non-Reductive Physicalism, Eliminativism and Strong Illusionism (Frankish) (“The Meta-Problem Challenge”) and also to the conceivability of zombies. Such explanations can be psychological, computational,

structural, neurological or cultural, one satisfying class of explanations appeals to the “riches of physics” and embrace a version of Stoljar’s (2006) Ignorance Hypothesis (IH) holding that we are ignorant of physical facts relevant to consciousness whose knowledge would explain away our anti-physicalist intuitions.

The IH comes in different versions that depend on how broad we take the “physical” to be. The “physics” may refer to our standard physics or to a non-standard physics that includes intrinsic or protophenomenal properties. In this paper, unless I say otherwise, I will embrace a first order approximation of the IH that takes the “physical” to be standard. Let’s call the physical facts that the IH appeals to—S-facts (after Stoljar). The discovery of such facts is important in its own right because there are good reasons to believe that providing a physical explanation of our anti-physicalist intuitions will constrain solutions to the hard problem (Stoljar’s Epistemic View does not search for possible S-facts but only claims that the IH is the best default explanation of our epistemic situation).

S-facts are not easy to imagine but easier to imagine than the facts needed for a physical explanation of consciousness. To see why S-facts are hard to conceive of consider Papineau’s (2007) critique of the IH:

*Stoljar is here placing strong demands on the content of our ignorance. It must be such that, if only we knew the relevant non-experiential facts, this would render zombies inconceivable. However, it is not clear that any non-experiential facts could play this role. By their nature, non-experiential facts would seem to be third-personal, objective, and non-perspectival, while experiential facts are first-personal, subjective, and perspectival. It is hard to see how knowledge of the former could automatically render the absence of the latter inconceivable.*

It is therefore quite possible that no physical facts can be both strange enough to explain our problem intuitions and “standard-physical” enough to count as standard.

The Strange Metal theory of consciousness (section “Strange Metals”) is an attempt to provide a physical explanation to some of our essential anti-physicalist intuitions and especially, an attempt to conceive of possible empirical findings about the neural correlates of consciousness (NCC) that can make a “philosophical difference” (section “NCC Correlations and Empirical Equivalence”) I will argue that unlike classical physics and QM, which lack the resources to provide S-facts, Quantum Gravity (especially with its unexpected connections to condensed matter theory) does harbor such resources (section “Physics with Resources to Explain the SMP and ~P Problem Intuitions”).

When it comes to formalizing consciousness by system theorists and mathematicians one can adopt top-down approaches inspired by symmetry (Kleiner, 2020) or by fundamental postulates (Tononi and Koch, 2015). Here, using a physics inspired approach, I begin by searching for a physics that can explain some strategically chosen anti-physicalist intuitions (section “SMP and ~P Problem Intuitions”) only to discover a highly abstract physics in which information is more fundamental than spacetime and “matter” and “identity” prove subtler than we realized (section “Physics with Resources to Explain the SMP and ~P Problem Intuitions”). The mathematical

structure that is exposed by this appeal to the physics of the meta-problem is relevant to attempting a mathematical formulation of consciousness that expands what we mean by physical system such as to endow those with rich private “inner-world” that is theoretically inaccessible to other systems.

To appreciate the unexpected relevance of quantum gravity to our anti-physicalist intuitions (4-2) I will consider Susskind’s (2017) radical ontological interpretation of the “Anti-de Sitter Space - Conformal Field Theory (AdS-CFT)” correspondence (4-5). In Feynman’s spirit of – “Don’t ask yourself whether it is *too strange* but whether it is *strange enough*” – I argue that Susskind’s interpretation is both physically “standard enough” and “strange enough” to explain relevant anti-physicalist intuitions.

For Susskind entanglement in  $d$  dimensions is equivalent to a wormhole in  $d+1$  dimensions, one cannot exist without the other. What is important for our purposes is his thought experiment that includes a spherical shell described by a two dimensional conformal QFT (quantum field theory). Imagine a large quantum computer implementing complex intractable computation (4-6). According to the AdS-CFT correspondence the spherical shell is dual to a three dimensional inner “bulk” AdS space (4-1). Susskind notes that this dual bulk space is different than the laboratory space. The bulk space implements what is essentially the same computation as the quantum shell but in a completely different way. While there is an information theoretic sense in which the dual spaces are identical, physically they could not be more different. Each element in one space maps unto a unique element in the dual space (bijection) and while an element may be fundamental, or local, in one space, its dual may be composite (relativity of fundamentals) or non-local (4-1). For our purposes what is important is that the only way to access the rich representation in the bulk is to have a “technician” uploaded unto the surface quantum computer. Susskind goes on to speculate that such a technician, once entering the bulk, could find ways to communicate with the lab technicians. However, the bulk space which is real, having a novel relationship to the lab spacetime, can also provide us with a physical subjective space harboring rich representations that evolve in parallel to the on goings of the quantum shell computer.

The question I will ask is whether consciousness can inhabit such “bulk” AdS spaces and whether such “physics” can explain why it is so hard to believe that consciousness is identical to anything in the brain [and is nowhere to be found in Leibniz Mill (section “SMP and ~P Problem Intuitions”) Such physics may also provide necessary connections between the more structural aspects of our phenomenology and corresponding structural aspects of the brain. One way or another a putative “physics of consciousness” will have to explain the relationship between the “space of phenomenology” and the “phenomenology of space” (section “Physics with Resources to Explain the SMP and ~P Problem Intuitions”) or what Chalmers (2020) describes as reconciling the *scientific* and *manifest* images of space] and here AdS-CFT has one more connection to our phenomenology that should interest system-theorists, It enables us to view rusted pieces of copper oxide as computing devices that “convert” hard, intractable, quantum, information inaccessible by perturbative methods, into easy, classical, *geometrical*, phenomenological

information accessible to perturbative methods (4-4). If we were to discover that our minimal PCC is describable by such CFT it would be hard to reject the possibility that the centrality of geometry to our spatiotemporal phenomenology results from such dual bulk AdS spaces. System theoretic approaches to consciousness need to take a stand on the privacy or “radical interiority” of consciousness. Some like Tononi (2008) claim that any system with non-zero Phi has unique (extra-theoretical, non-falsifiable) access to itself not available to any other system, while some like Clark (2019) use predictive coding to attribute our anti-physicalist intuitions to the unavailability of interoception to experience.

Another way in which I hope this paper will be relevant to system theorists is as an example of a meta-theoretic strategy conjoining putative solutions to the hard problem with topic neutral explanations of our anti-physicalist intuitions (section “The Meta-Problem of Consciousness”). As we will see system theorists interested in the hard problem may want to first construct (honest) system-theoretic models that generate anti-physicalist intuitions similar to ours and only then use this construction to constrain the hard problem (section “The Meta-Problem Challenge”). While the first step is empirical the second step is more philosophical and embraces Chalmers claim that the Meta problem and hard problem of consciousness are “almost independent” but co-constraining. As a matter of fact a fruitful way to think of a solution to the Meta problem of consciousness is that of showing that a “smart enough” honest embedded AI is likely to generate conscious and problem reports similar to ours.

## INTRODUCTION

The biggest obstacle standing in the way of a “mathematical formalization of experience” is probably related to David Chalmers’ “Structure and Dynamics Argument” (Chalmers, 2002; Alter, 2015) and similar to:

- a) Mathematics is structural.
- b) Structure can only yield more structure.
- c) Consciousness is not structural.
- d) Conclusion: Mathematics cannot describe consciousness.

One way to reject this conclusion is to reject the first premise by arguing that mathematics may indeed include non-structural entities too. Another way out is to reject the third premise and search for relevant structural aspects of conscious experience. I will follow the second path and concentrate on three points of contact between mathematics, modeling and experience:

- 1) Distinguishing two major properties of consciousness—The non-structural “phenomenal character” common to all conscious states and the more structural “representational content” that differentiates conscious states.
- 2) Distinguishing between the “hard problem of consciousness” and the more empirical “meta-problem of consciousness.”

- 3) Searching for physics that is rich enough to solve the meta-problem.

Sec. “Ineffable and Manageable Anti-physicalist Intuitions” explores the first point of contact. “SMP and ~P Problem Intuitions” presents what I term the “Structural Mismatch Problem” (SMP) and the “~P Problem.” “Magic and Necessity” presents a thought experiment correlating holography and necessity, while “Coding and Necessity” asks whether “deep-learning-based” coding strategies can solve SMP. Sec “NCC Correlations and Empirical Equivalence” explores causation and correlation in the NCC.

Sec. “Empirical Challenges” explores the second point of contact. “The Meta-Problem of Consciousness” introduces the meta-problem of consciousness, “The Meta-Problem Challenge” considers more empirical approaches to the meta-problem and gauges its philosophical impact. “The Meta-Problem Challenge” finally argues that the meta-problem pressures Russellian Monism.

Sec. “Holographic Duality and Strange Metals” combines the first two sections to search for physics that solve the SMP and the ~P Problem, concluding that the phenomenon of “Holographic Duality” in modern physics has the resources to explain these intuitions. I shall introduce some concepts in and around the modern physics’ treatment of duality in order to promote, in the final discussion, a “holographic theory of consciousness.” “Duality” introduces the principle of duality, “Physics with Resources to Explain the SMP and ~P Problem Intuitions” argues that “Holographic Duality” is suitable to the presented problem. “Strange Metals” introduces the concept of “Strange Metals” and the “AdS-CMT correspondence,” “AdS-CFT Correspondence as Computation” introduces the AdS-CFT correspondence as pushing the envelope of computation, “Einstein-Rosen bridge (ER) = einstein podolsky and rosen (EPR)” portrays the “ER = EPR interpretation of entanglement” and a more radical version of duality, and “Holographic Duality and Quantum Error Correction Codes” describes the connection between quantum error-correcting codes and the constitution of spacetime.

I will end with a discussion in which I will ask whether the suggested “holographic” theory of consciousness can handle zombies.

## INEFFABLE AND MANAGEABLE ANTI-PHYSICALIST INTUITIONS

### Structural Mismatch Problem and ~P Problem Intuitions

Consciousness is roughly attributed two kinds of properties: *phenomenal character*, common to all conscious states, and *representational content*, specifying the difference between such states.

The phenomenal character consists of properties common to all conscious states, including the feeling that *there is something it is like to be conscious*, or that consciousness is given to a self, or to itself, or that it is directed, or transparent to



introspection, or that it is self-affirming. The representational content refers to the difference between such states.<sup>1</sup> In section “Coding and Necessity” I will present recent work suggesting brain-bound explanations of some of the structural aspects of our phenomenology.

The question “What is it about the way phenomenal experience *is* that is made necessary by the way the brain *is*?”<sup>2</sup> can be broken into two:

- a) What is it about the way phenomenal character *is* that is made necessary by the way the brain *is*?
- b) What is it about the way representational content *is* that is made necessary by the way the brain *is*?

The first point of contact between mathematics and consciousness that I want to consider is question (b) – because, unlike (a), it relates two domains with structural properties. Hence, in this section I evaluate the prospects of establishing necessary connections between the structural aspects of representational content and the structure of the brain states that generate (or correlate with) it.

Already in 1714, in Sec. 17 of his *Monadology*, Leibniz (1960) argues that perception cannot be given a “mechanical” explanation:

One is obliged to admit that perception and what depends upon it is inexplicable on mechanical principles, that is, by figures and motions. In imagining that there is a machine whose construction would enable it to think, to sense, and to have perception, one could conceive it enlarged while retaining the same proportions, so that one could enter into it, just like into a windmill. Supposing this, one should, when visiting within it, find only parts pushing one another, and never anything by which to explain a perception. Thus it is in the simple substance, and not in the composite or in the machine, that one must look for perception.

There are two things that you will not find in the brain. First, consciousness itself, which is only given to the “owner” of the brain (i.e., the phenomenal character of consciousness). Second, and just as importantly for our purpose, representational content; if the owner of the brain is conscious of three blue goats, anyone in the windmill shall be hard pressed to find a blurry imprint of three goats or anything resembling that. If the image is physical how can it be identical to something in the brain? I will concentrate on two “problem intuitions” that are related to

Leibniz’s Mill. The first is the “The Structural Mismatch Problem” (SMP), following Chalmers’ (2016) use of the term in the context of the “combination problem” of constitutive Russellian Monism, yet applying it more generally. The SMP is the intuition that the structure of the representational content of phenomenal states lacks any necessary connections to the brain structure that generate them. This phenomenal field that can harbor a huge number of “qualia” combinations can be considered as a system of differences endowed with structure and information laden. Even eliminative materialists that view qualia as illusionary would agree that it is at least a richly structured illusion utilized by the brain for executive function. At the same time, the physical or neural substrate generating such “phenomenal fields” has its own structure that is completely different from the phenomenal structure. All major theories of mind suffer from an inability to establish or even conceive of necessary connections between these two structural domains, or sets of differences, despite their exquisite correlations.

The second central anti-physicalist problem intuition is what I call the “~P Problem,” or “not Physical” problem intuition, to borrow from Levine (2019)—*why mind-body identities provoke cognitive dissonance in a way standard theoretical identities don’t*.

The ~P problem intuition is also related to the “other minds problem”: how can complex physical systems generate complex physical self-representations accessible to those systems but completely concealed from “without”? As we will see, the same physics that I argue to be rich enough to solve the SMP is also rich enough to solve the ~P Problem intuition, providing the only physical account that I know of such “private” spaces.

Both are important anti-physicalist problem intuitions and, as we will see, constructing consciousness-independent explanations to these problem intuitions is an important part of a multidisciplinary meta-problem research program concerning philosophers, system theorists and computational neuroscientists.

One way of explaining these problem intuitions is explaining why the necessary connections between mentality and the brain are hidden. Let us consider the next “holographic metaphor.”

## Magic and Necessity

Uri Geller hears about a planet that is pretty advanced but with inhabitants that have not yet discovered holography and the Gabor transform (a linear transformation relating a 2 dimensional hologram to its associated 3 dimensional holograph). He loads his spaceship with holograms and a laser and takes off. As soon as he arrives, he announces that he will perform an act of real magic. The mostly naturalist inhabitants, embracing continuity and rejecting radical emergence, gather around with curiosity. Geller pulls out a thin glass plate (hologram) and shines a laser through it producing a 3D holographic image. The inhabitants are surprised but are sure that the holographic image emerges from the holographic plate; they tell Geller that, however, surprising this might be, this phenomenon must have a logical, physical explanation. Geller retorts that at the very least they have a serious structural mismatch problem, because the patterns of the face carved on the surface of the hologram plate have nothing to do with

<sup>1</sup> To a first order approximation I will assume that reality exists, that it is structural and that our perceptions of the world are veridical to some extent (say, because of evolutionary pressure). While a bee and a primate may have different phenomenal character to a consciousness realist “phenomenal character” must refer to some minimal set of properties.

<sup>2</sup> I believe that “biting the bullet” on strong emergentism should be an act of the last resort. This is certainly true for giving up on necessary connections between two structural domains that are exquisitely correlated. Also in Susskind’s cell the hardware of the quantum computer is not multiply realizable in the sense that a computer that simulates complex entanglement without itself being a similarly entangled quantum computer will fail to produce the bulk dual. Also the evolution of the relevant states of such a machine is only tractable to the bulk (Preskill). Similarly to Joseph Levine (2019). I loosen-up the metaphysical definition of identity to preserve necessity.

the holographic image and do not seem to necessitate it in any way. He proceeds to challenge them to conceive of any possible necessary connection between the two. He also shatters a hologram into pieces and shows them that shining the laser through the fragments produces low resolution holographs of the whole image, and then challenges again the locals to explain it [especially the large local panpsychist community who believe that their biggest problem is explaining how elementary “little subjects” combine into a large subject (Coleman, 2014)].

The inhabitants happen to have powerful computers with decent big data and deep learning capabilities, and confiscate the holograms to explore the correlations between the holograms and the holographs. After a couple of months, they are able to predict the images generated by holograms that they have not analyzed before (They kept a couple for that purpose).

The inhabitants of the planet tell Geller that they can generate these images by using their computers and that this is not real magic, to which he responds that the production of mere correlations does not provide necessary metaphysical connections between the plates and the holograms and that they still have a serious structural mismatch problem. The inhabitants go back to the lab and use information compression algorithms to search for the most efficient way of constructing holographic images from previously unseen holograms, with the constraint that the fragments produce lower resolution holographs of the whole image. After another month, they discover the Gabor transformation in which the hologram necessitates the holograph, learn about phase-front reconstruction and fine Geller in Bitcoin.

## Coding and Necessity

Here we can ask whether something similar can be done with neural decoding, in which deep learning and big data approaches can be used to determine what face a primate is looking at among thousands just by “observing its brain.” The last 20 years have witnessed a steady improvement in our ability to decode the subjective contents of the human brain, from *Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex* (Haxby et al., 2001) to *Identifying Natural Images from Human Brain Activity* (Kay et al., 2008), to *Reading the Mind's Eye: Decoding Category Information during Mental Imagery* (Reddy et al., 2010), to *Neural Portraits of Perception: Reconstructing Face Images from Evoked Brain Activity* (Cowan et al., 2014), to more recent work based on both single neuron measurements, e.g., *The Code for Facial Identity in the Primate Brain* (Chang and Tsao, 2017), and more global fMRI methods, *Reconstructing Faces from fMRI Patterns Using Deep Generative Neural Networks* (VanRullen and Reddy, 2019). These new decoding capabilities fall short of solving the SMP but are not just mere brute correlations. At their best, such deep-learning-based decoding can identify the optimal dimensionality of such spaces. In Chang and Tsao's (2017) code the “computer” identified a 50-dimensional face space, and facial images are thus expressed as points in a 50-dimensional space. What was surprising about this work is that identifying the dimensionality of the “face space” and the relevant neural correlates enabled recording from a few hundred neurons to accurately decode a large number of faces.

This may be attributed to the conjecture that deep learning and evolution settle on the same number of dimensions. It also seems that similar linear code strategies are ubiquitous to biological pattern recognition. While such codes do not help solve the meta-problem of consciousness or the SMP they do seem to support the claim that representational content possesses a structure that lends itself to scientific investigation.

The decoding of visual images like faces has the advantage that it yields a “metric” in which the degree of difference between images is commensurate with the difference in neural firing patterns that can be expressed as a distance between two points in some abstract space. When faces are similar, the activation patterns are similar and their distance is short. One way of constructing such “neural manifolds” that admit a metric and a distance formula is using *statistical geometry* (Roy and Kafatos, 2003).

The difference between faces is easier to formulate than the difference between scents. However, that has not stopped olfaction researchers from constructing a metric for the “distance” between different scents on an olfactory scale. Olfaction researchers deal with two transformations, from the physical space of odor molecules to the space of neural activity and from that to the space of odor perception. To formulate such transforms we need to first construct metrics for these three different spaces. In *Measuring Smells* (Haddad et al., 2008), the authors take two different approaches to constructing a metric for structural chemistry in which the “Dragon” software<sup>3</sup> is used to obtain 1664 molecular descriptors for more than 1500 odorants. Each odorant is described by a vector in a 1664-dimensional space and subjected to a principal component analysis as a well-established method for dimension-reduction. As for this study, each odorant in the physical space can be described in two ways, one where the distance between odorants is determined by their principal component score (PC1) and one by taking the full vectors and measuring the “geometric distance” between them (summing the squares of the differences between the components and taking their square root). The application of principal component analysis to all three sets of data showed that the key axis (PC1) was correlated across all three domains, providing a one-dimensional metric based on pleasantness that could also be extended to other species. The second multi-dimensional metric construction was successful in predicting neural activity and showing that the smaller the distance between odorants in the physico/chemical space, the more similar the neural responses. This metric too seems to be applicable to different species, suggesting that aspects of the structure of olfactory space are conserved across species based on the reliance on similar environmental regularities.

Olfactory decoding is not good at explaining why a rose smells the way it does, more importantly, nor does it help solve the SMP and ~P Problem intuitions. I presented these studies in the olfactory field as these findings are analogous to the visual field, which is more relevant to the purpose of this paper. This is because visual perception is more geometric than olfaction, and as we will see in Sec. “Holographic Duality and Strange

<sup>3</sup><http://www.taletale.mi.it>

Metals” where geometric structure emerges from “holographic computation.”

It seems as though such brute computerized decoding strategies can at most tighten the correlations between phenomenal states and the neural (physical) processes that generate them; yet, can they reveal, a deeper overarching small set of psychophysical laws? Perhaps because information compression is evolutionary advantageous and because Deep Learning neural nets and their layered architecture are good at optimizing information compression. Here is Joseph Levine (2019) on the “fine-tuned structure of experience”:

*“Among the materialist arguments that I find most compelling is the appeal to the myriad ways in which what I will call the ‘finetuned structure’ of our conscious experience can be explained to a very large extent by the functional profile of the underlying physical mechanisms. For instance, take color experience. Leaving aside the explanatory gap that attends there being any experience at all, or one’s color experience having the particular qualities that it does, there is clearly a lot about the structure of that experience that is explicable by appeal to underlying physical-functional mechanisms. ... It’s reasonable, then, to suppose that whatever psychophysical emergent laws there are possess a unified structure that makes sense of this relation between physical-functional architecture and the fine-tuned structure of experience.”*

Can deep-learning and big data provide us with that “unified structure” and proper psychophysical “laws of nature.” There is a sense in which improvements in such methods can at best make strong emergence harder to accept. Even cases in which deep learning can discover invariant properties in the data, like optimal dimensionality that both facilitates information processing and is shown to be harnessed by the brain, are “environmentally opportunistic”; thus, they lack the resources to reveal a mathematical transformation that explains away the SMP in the way that the Gabor transformation explains away the mystery of holography. So, while brute decoding lacks the resources to solve both the hard problem and the meta-problem of consciousness it does show that representational content can be accessed scientifically and modeled.

We are still nowhere close from an answer to the question that framed this section: What is it about the way the structure of our phenomenal states is that is necessitated by the structure of the physical substrate generating it? Nor are we closer to explaining away the SMP and ~P Problem intuitions.

The classical hologram-holograph relationship is based on phase front reconstruction and can perhaps be related to neuronal processing but still relates two physical domains. It builds as best an analogy with the relation between phenomenal states and their neural correlates. This begs the question, if the “Technicolor” phenomenal domain is indeed physical, we need to understand why it is so hard to accept as physical, and hopefully physics is rich enough to explain that. To try and answer this question I will argue that we need a more radical computation instantiated in a more radical system by more radical physics. In Sec. “Holographic Duality and Strange Metals” I will still appeal to a holographic correspondence of sorts and yet a radically different one.

## Neural Correlates of Consciousness Correlations and Empirical Equivalence

The claim that frames this whole discussion is that there are physical processes whose instantiation by the NCC could make a “philosophical difference” by explaining relevant anti-physicalist intuitions. The physics that I will rely on to do that allows for very strange correlations in which two complex entities A and B are highly correlated without A causing or constituting B and vice versa. Neither can such correlations be explained by appeal to third party explanation. To see why such physics can make a “philosophical difference” let’s consider Kriegel’s (2020) “Beyond the Neural Correlates of Consciousness” in which he argues that explaining the correlation between A and B is exhausted by:

- 1) A causes B or B causes A.
- 2) A constitutes B or B constitutes A.
- 3) A and B are caused by a third party.
- 4) A and B are constituted by a third party.

I will argue that explaining the correlations involved in “A is Dual to B” (loosely related to the correlations typical to QM entanglement) does not fit into any of these categories. Kriegel goes on to argue that:

- a) When it comes to mind-brain correlations we are not likely to discover empirical facts about the NCC that will favor causal explanations over constitutive ones or vice versa.
- b) The six possible explanations of the mind-brain correlations that appear in 1-4 map out our standard theories of mind.
- c) It is not likely that we will discover empirical facts about the NCC that will favor one theory of mind over another.

Kriegel’s “empirical equivalence” can be interpreted both pessimistically and optimistically, the pessimist may conclude that we are not likely to discover philosophically relevant empirical NCC facts while the optimist can accept premises a-c but argue that there are explanations of the correlation between A and B, missed by 1-4, that map unto theories different than standard theories of mind. Not only is it possible that we will find empirical facts about the NCC favoring such theories, those among us that search for a philosophically significant physical explanation to the mind-brain correlations seem to gain an important hint; the physics explaining these correlations cannot be based on causation or constitution nor on simple third party explanations. If one accepts 1-4, but refuses to believe that we will not discover philosophically relevant empirical facts about the NCC, such “novel” theories becomes especially attractive. This means that if condensed matter physicists were to discover materials that display such “exotic” correlations it would be worth asking whether similar physics is instantiated by the NCC (including its biomolecules and electron clouds).

Kriegel believes that the theories of mind mapped by 1-4 are empirically equivalent because in the case of NCC correlations it is hard to imagine empirical findings about the NCC that favor constitution over causation (or vice versa). Roughly, the reason for that is that the difference between causation and constitution boils down to establishing the presence, or absence, of either a



causal mechanism or time-delay and it is hard to think of an experiment that can establish the existence (or lack) of a causal mechanism or time delay in the case of mind-brain correlations. A similar argument applies to the third party explanations.

There are other reasons to think that the correlations between brain states and their correlated phenomenal states cannot be described by 1-4 and all this suggests that we need to consider bulk physics exhibiting correlations that do not fit comfortably into 1-4. This is precisely what happens with the physics of holographic duality and the AdS-CFT (“necessary correlations”) correspondence that not only transcends causation and constitution but has other philosophical advantages.

Discovering that the NCC harbors physical mechanisms instantiating holographic duality (like in the putative case of the Strange Metals) would provide an example of a philosophically significant empirical finding related to the NCC that is interesting *even* as a thought experiment because it shows that philosophically relevant empirical discoveries about the NCC are possible.

What is important for our purpose is to show that correlations typical of Holographic Duality cannot be described by 1-4 because such duals neither constitute nor cause each other, nor are they constituted or caused by a third party, convincing Vistarini (2017) that such a relation is better described as a Dual Aspect theory.

Radical Duality suggests that the same information is realized in completely different ways and “simultaneously” so...

In Sec “Duality” I will argue that CFT-AdS provides such a connection and constitutes such a metaphysics. To conclude, the “physics of consciousness” should explain correlations that are not explained by Kriegl’s 1-4. In Sec “ER = EPR” we will consider Leonard Susskind’s thought experiment, relating entanglement in QFT to the Einstein-Rosen Bridge connecting black holes as a radical example of such thinking.

## EMPIRICAL CHALLENGES

### The Meta-Problem of Consciousness

The second point of contact between physics and consciousness derives from the consideration of the “meta-problem of consciousness”—seeking to provide topic neutral explanation to what we say and know about consciousness—a strategy used by philosophers including Hobbes, Hume Spinoza and also Kant (1781/1999); Place (1956), Armstrong (1968); Dennett (1992), Rey (1996); Carruthers (2017), and lately Kammerer (2018) and Frankish (2019). The meta-problem is situated “in between” the hard and easy problems of consciousness. This *empirical* problem constrains the hard problem while lending itself to mathematical modeling.

We may be witnessing the beginning of a trend in the philosophy of mind in which metaphysical theories of consciousness aiming to solve the hard problem of consciousness (i.e., how matter gives rise to the mind) must also explain “conscious reports” and “problem reports,” or the problem of how consciousness acts back on the matter of the brain to become the source of what we know and report about it. A successful theory

of introspection should both secure the foundations of our self-knowledge and explain how consciousness manages to generate reports about itself. In the introduction to their *Introspection and Consciousness*, Smithies and Stoljar (2012) write:

Recent philosophy of mind has been dominated by metaphysical questions about the nature of consciousness and its place in the physical world, while much less attention has been devoted to questions about the epistemic role of consciousness as a source of knowledge and justified belief.

Recent work attempting to conjoin the metaphysics of consciousness and the epistemology of self-knowledge (Stoljar, 2016; Byrne, 2018) has culminated in Chalmers’ (2018) *The Meta-Problem of Consciousness*. According to Chalmers, the meta-problem of consciousness is the problem of why we think that the problem of consciousness is hard or why we think that the explanatory gap associated with consciousness is categorically different from explanatory gaps between different sciences. Chalmers sharpens this more unified approach by concentrating, as a first order approximation, on topic-neutral (independent of consciousness and its cognates) explanations of our *problem intuitions*. We can think of solutions to the meta-problem as accounts explaining why “smart enough” unconscious AI is likely to generate “conscious reports” like “I am conscious!” and problem reports like “I cannot believe that consciousness is physical”. “Problem intuitions” can be viewed as the underlying artificial machine states that cause such reports. A strong illusionist about consciousness like Frankish (2019) refers to those as “quasi-phenomenal,” whereas a strong emergentist like Joseph Levine (2019) refers to them as “intentional scripts,” being indeed directed either at the world or at other internal states of the machine, however, synthetic their origin.

The meta-problem is situated in between the hard and easy problems. It is “easy” because it is independent of consciousness, and yet it is “hard” because it requires explanations on how non-conscious systems can generate conscious reports and problem reports that are indistinguishable from ours. The meta-problem is thus not only the hardest “easy problem” but, as Chalmers (2018) shows, one constrains the possible solutions of the hard problem.

Chalmers’ procedure aims to test the coherence of theories of mind by demanding that metaphysical theories of consciousness clarify their position on the meta-problem. The existence of a solution to the meta-problem of consciousness is crucial to both eliminativists like Dennett and to proponents of the conceivability argument like Chalmers (because for a zombie to be conceivable it must generate conscious and problem reports). The meta-problem program is a multidisciplinary attempt to provide topic-neutral explanations to our problem intuitions, drawing from naturalistic, computational and philosophical insights. The problem intuitions giving rise to the hard problem can depend on culture, language and more. Naturalist explanations can appeal to the biological evolution of self-modeling, counter-factual emulation and the modeling of space and time.



The meta-problem also challenges computer scientists. In a recent paper, *Consciousness as Generative Entanglement*, Andy Clark (2019) lists the many successes of predictive processing in modeling perception, action and attention, adding:

But despite their clear promise as accounts of the neurocomputational origins of typical and atypical forms of human experience, they have not yet been leveraged so as to shed light on the so-called hard problem of consciousness. . .

Clark (2019) suggests that predictive processing is ideally suited to solve the meta-problem by explaining why it is likely to report “puzzling qualia”:

The upshot is that the contents that constitute our qualitative experiences are subtly responsive to our own reactive dispositions and current physiological state in ways that remain hidden from us.

These involve complex cascades of interoceptive and proprioceptive predictions whose inaccessibility to introspection causes us to attribute their origin to something non-existent similar to Armstrong’s “headless woman illusion,” where failing to see the woman’s head is confused with seeing the woman without a head. Here the fact that qualia seem to emanate from nowhere causes them to seem puzzling.<sup>4</sup>

Carl Friston’s (2018) *Am I Self-Conscious?* takes a different route, where the most essential property to conscious self-modeling is “temporal thickness” [also see Chouraqi (2011)]:

*The proposal on offer here is that the self-evidencing has a temporal thickness and depth, which underwrites inferences about the counterfactual consequences of action. This necessarily lends (active) inference a purposeful and self-centered aspect that has the hallmarks of consciousness.*

Friston is an eliminativist trying to provide a topic neutral explanation to the hard problem of consciousness and not to the meta-problem like Clark; however, what is interesting here from a modeling point is that Lisman (2005) and Lisman and Buzsáki (2008) “theta precession inspired” Temporal Coding mechanism seems to provide such temporal thickness that can also be related to Gregory’s “presence” and useful for modeling indexical concepts (“here” and “now”). These are all topic-neutral contributions and it is hard to see why an AI with an architecture inspired by place and grid neurons may suddenly spring to consciousness. Combining advanced versions of self-modeling, temporal modeling and the modeling of counter-factual emulation is part of the meta-problem of consciousness program.

The meta-problem strategy is ideal for modes of analysis that combine *a priori* analytic approaches (like the six types of theories of mind comprising Chalmers’ “A-F taxonomy”) with empirical ones. It is not a naturalistic strategy to reduce our immutable philosophical concepts to scientific reconstructions; rather, the

aim is to suspend such philosophical investigations, to first solve the meta-problem empirically, to account for how these appeared in beings like us (Awret, 2019).

## The Meta-Problem Challenge

The meta-problem of consciousness challenges theories of mind with a procedure termed the “meta-challenge”:

- a) Divide the theoretical space into realist and eliminativist theories, and then divide those into those accepting a solution to the meta-problem and those that do not.
- b) Challenge eliminativists to provide a topic-neutral solution to our anti-eliminativist intuitions.
- c) Challenge consciousness realists that accept a solution to the meta-problem to defend themselves from charges of unjustified belief.
- d) Challenge consciousness realists rejecting the existence of a solution to the meta-problem to explain their choice (as it would likely entail that zombies are inconceivable and that machines cannot pass the Turing Test).

Applying this procedure to different theories of mind is outside the scope of this paper. What is important to the purpose of this paper is that the meta-challenge puts real pressure on current theories of mind, so much so that it causes Chalmers to be pessimistic about their prospects.

Solutions to the meta-problem must be realized by some brain process that we might call the “meta-process,” just like solutions to the hard problem by some “consciousness process” (Chalmers, 2018). If the two processes are identical, one is forced to embrace eliminativism (Frankish, 2019); if they are different, one must defend charges of unjustified belief. Chalmers’ own preferred solution is “realizationism,” by which consciousness is realized by meta-processes or where phenomenal consciousness is realized by access-consciousness. In the next section I will argue against realizationism and for what Chalmers terms “Meta-Correlationalism,” by which the meta-process and the consciousness process are separate but perfectly correlated. These theories are problematic when the correlations are brute, like in dual-aspect theories of consciousness. Chalmers’ own dual-aspect information theory of consciousness, by which information has a phenomenal aspect and a material aspect, at the same time suggests that the same information is simultaneously realized mentally and materially. The holographic theory that I pursue here is partly similar to this, and yet very different as now *the two aspects are correlated and necessarily so*.

Even if the categorical basis of microphysical properties were phenomenal, that would not explain how consciousness starts flipping electrons so as to use the brain to proclaim its existence. However, “meta-challenging” Russellian Monism (RM) shows that it fails to overcome charges of epiphenomenalism (as it “promises” to do) since we still have no idea how an intrinsic and non-relational categorical basis may act on the “molecules of the brain” to announce itself relationally. To make things worse, solving the notorious combination problem of type-F theories does not seem to help with the meta-challenge. Both ways out for realizationism considered by Chalmers

<sup>4</sup>Sartre coherently linked the “being-for-itself” intrinsically to nothingness, as it was a hole in being, so to say. Also see Chalmers (2018) on attempts to solve the meta problem that appeal to “primitive property attribution.”

are problematic. Interactionist Dualism (Crick and Koch, 1990) has insurmountable problems with causal closure (introducing new metaphysical and explanatory gaps), and the attribution of problematic “phenomenal powers” to the categorical base (Morch, 2020) not only defeats the purpose of Russellian monism but raises a novel problem. Were phenomenal powers to exist (say they could flip an electron’s spin), then one could design experiments showing that during conscious reports it is possible to establish the existence of extra-theoretic influence on the results of measurements performed on the physical substrate underlying such reports. I do not see this happening anytime soon.

In the next section I will present some new and surprising connections between quantum gravity and condensed matter theory and argue that they possess the resources to provide topic-neutral explanations to the aforementioned problem intuitions.

## HOLOGRAPHIC DUALITY AND STRANGE METALS

### Duality

In their *Introduction to Special Issue on Dualities*, Castellani and Rickles (2017) begin with “dueling concepts” in the Chinese cosmology and compare these to an analogous duality in science:

In Chinese cosmology, the various phenomena of the universe are then viewed as an interplay of these dueling concepts. Further, the two sides are inextricably entangled and interdependent, neither making complete sense without the other: there is no sense of one side of the dual pair being more fundamental or superior. So seems to be the case with dualities in science, with a similar entanglement holding together dueling concepts such as “hot/cold,” “big/small,” “high-energy/low-energy,” “finite/infinite,” “composite/elementary,” “localized/delocalized,” etc., though in such a way that an equivalence holds between them at some level—in general, one finds that some quantity and its reciprocal are involved in a duality mapping. There is, then, mystery in dualities, in making sense of how there can be equivalence given such apparently stark differences (differences labeled, in this special issue, as “striking” by De Haro and “shocking” by Huggett). Yet, within this mystery there lies a golden opportunity for philosophers.

Duality in mathematics and physics is not a theorem or a law of nature but a deep principle (Atiyah, 2007) that excels at “carving nature at the joints,” so to speak; in mathematics it has been used for hundreds of years and continuously adapted and evolved, undergoing successive generalizations that guided progress in geometry, algebra and analysis. In physics, duality entered the scene with Maxwell’s equations’ invariance to a duality transformation that exchanges the electric field  $E$  with the magnetic field  $B$  and  $B$  with  $-E$  when no sources are present, later with wave-particle duality in quantum mechanics, and finally with the Kramers-Wannier duality in 1941 that yielded a

simple way to predict phase transitions in the 2D Ising model. Quantum field theory then presented us with new dualities, like the “Montonen Olive Duality” in 1977, while dualities became central to string theory—with S-Duality, T-Duality and especially the mysterious M-theory, in which all five different string theories were shown to be cross sections of the same currently unknown theoretical object-related to each other by duality transformation. The E–B and KW dualities are self-dual, in the sense that the duals in these theories are described by the same equations. However, most dual descriptions are not isomorphic and the duals that I will be interested in here, is the AdS–CFT correspondence (Maldacena, 1998), or “gauge-gravity duality,” relating a many-body strongly interacting quantum field theory on a  $d$ -dimensional surface to a gravitational theory in “the bulk” with  $d+1$  dimensions. The theory is related to the holographic principle in string theory, stating that the information of the physical “bulk” in  $d+1$  dimensions is inscribed on its  $d$ -dimensional boundary. This String–QFT duality is surprising; to quote Polchinski (2015):

Dualities between field theories, and dualities between string theories, are remarkable, taking QFT and string theory far beyond their perturbative descriptions. A duality between a field theory and a string theory might seem to be impossible, on several grounds. String theories require ten dimensions, whereas renormalizable field theories do not seem to exist in dimensions greater than four... String theories seem to contain many more degrees of freedom than QFT’s, from the infinite number of internal states of the string. And, string theories contain quantum gravity, with its many conceptual puzzles, while renormalizable QFT’s do not. Well, prepare to be amazed.

Duality in physics can be seen as a symmetrical transformation relating different *theories* of the same entity. Unlike ordinary global and local symmetries that leave physics invariant with respect to solutions to the same theory, under this duality the observable physics is invariant under exchange of theories. Duality is therefore a more radical symmetry that can relate different physical theories in different space dimensions, while maintaining a strict but highly counter-intuitive bijective mapping between fields, expectation values and other relevant physical variables.

The more we know about the structure of physics and math, the more central the role duality plays in acquiring this knowledge. However, the connections between the notion of duality and philosophy, especially philosophy of mind, are less well established (Castellani and Rickles, 2017):

Despite their ubiquity and importance in physics, mathematics, and logic, it is fair to say philosophers have yet to embrace dualities with the gusto they deserve. This is particularly unfortunate since dualities connect to a great many issues that philosophers have fully embraced. To name a few notable examples:

- Reduction, emergence, and fundamentality.
- Theoretical equivalence and synonymy.
- Underdetermination and empirical equivalence.

- Realism versus anti-realism.
- Unification in (and of) mathematics and physics.

To which one can add dual aspect theories (Vistarini, 2017), identity, and, which interests us here, the meta-problem of consciousness. One can also try to use duality as a meta-theoretic tool to capture the relationship between different theories of mind. For example, one might use the “relativity of fundamentality principle” (Castellani, 2016) by which what is fundamental on one side of the duality becomes composite on the other one.

The reason that the AdS-CFT correspondence is “holographic” is metaphorically represented by the holographs of 3D glasses; in the same way that the unintelligible interference patterns of the quantum fields on a two-dimensional surface are equivalent to, or dual to, an intelligible classical/phenomenological theory with an extra space dimension, the 2D glass hologram in linear optics, with its unintelligible printed interference patterns, is equivalent (in informational terms) to a 3D hologram made up of elements we may relate to. In a way, the 3D hologram is a geometric interpretation of a 2D field theory, serving, in that sense, a powerful parallelism for the AdS-CFT duality and the more general gauge-gravity duality. The difference is that, in the case of classical holography, the hologram with the use of a laser generates the holograph, which is not true in the case of more symmetric dualities like AdS-CFT.

## Physics With Resources to Explain the Structural Mismatch Problem and $\sim P$ Problem Intuitions

If the most concrete thing we know—consciousness—is physical, akin to a novel exotic phase of matter, then the “physics of consciousness” must look very strange, perhaps one that will deconstruct matter, discover novel reflexive states, rely on subtler notions of identity and clarify the relationship between spacetime, information and entanglement. The most pressing problem in this respect is probably to relate such a putative physics of consciousness to the physics of spacetime and information.

One can argue that classical physics, including non-linear dynamics and electromagnetic theory, lack the resources to describe such states of matter, as do QM and QFT. I am well aware of QM (Stapp, 2007) and QFT theories of consciousness (Ricciardi and Umezawa, 2004; Freeman and Vitiello, 2006), and yet I believe that they lack the resources to solve not just the hard problem but also the meta-problem. After all, one of the biggest obstacles on the path to a “physics of consciousness” is relating it to space and time: we have no idea how to do that. Can information be realized in ways that transcend our ordinary conceptions of space and time? However, even if we consider background-independent physics, it is hard to see how some stringy version of quantum gravity can be used in a direct assault on the hard problem.

This is where the meta-problem comes into play. We can ask an “easier” question that is still relevant, i.e., “Can QM and QFT solve the meta-problem by providing topic-neutral

explanations of our problem intuitions?” Chalmers (2018) suggests that they do not and as I argued in the prolog such S-facts are worth pursuing.

Is the physics of quantum gravity and the Planck scale up to the task and can it provide us with a solution to the meta-problem and tangible philosophical advantages?

The superstring revolution started in the 1960's with the hope of reaching a theory of everything, thus explaining the origin of spacetime by unifying the standard model, super-symmetry, QFT and local gauge theories with gravity and Einstein's equation. The theory had spectacular early success in areas like enumerative geometry, making substantial contributions to mathematics which won a string theorist like Ed Witten a Fields Medal. However, string theory was very frustrating i, because, despite its beauty and the deep connections it had with the foundations of mathematics and physics, the objects it predicted could not be verified experimentally. For example, experimental verification of the existence of mini black holes predicted by string theory would require a particle accelerator 100 times bigger than CERN. Any attempt to relate the physics of string theory and the AdS-CFT correspondence to the meta-problem and the hard problem must explain how Planck scale physics 20 orders of magnitude smaller than a proton may be relevant to the brain. After all, it is hard enough to establish the existence of non-trivial quantum effects in biological systems at room temperature. Even if Planck scale physics and quantum gravity did have the resources to solve the meta-problem, we would first need to show that this physics is relevant to macroscopic systems describable by Condensed Matter Theory (CMT), and then show that such peculiar CMT mechanisms are instantiated in the brain, preferably by processes that correlate with consciousness. In the last 20 years, the discovery of totally unexpected and even mysterious connections between CMT and quantum gravity has transformed both fields in a way that made them essential to each other. String theorists can finally perform laboratory experiments to verify their predictions and learn strange new facts about spacetime, black hole thermodynamics and the information paradox. Condensed matter theorists, in their turn, can use the sophisticated string theory mathematical machinery to understand highly entangled, strongly interacting novel states of matter that cannot be described by standard perturbative CMT. The possibility of a connection between such CMT states of matter and the brain is currently a useful thought experiment relevant to both the possibility of a physical solution to the meta-problem.

## Strange Metals

In 2007 Sachdev and his collaborators (Hartnoll, 2007) were trying to understand the properties of 2D high-temperature superconducting Copper-oxides named “Cuprates” (rusted copper) that displayed quantum critical behavior, namely scale invariance which made them describable by a conformal quantum field theory in two dimensions. When these materials were heated above their superconducting phase, they displayed unexpected transport properties (Nernst effect) and an electronic state of matter that can be described as “irreducibly many particle entangled compressible matter” (Zannan, 2018). These



“Cuprates” were dubbed “strange metals,” because this electronic phase of matter exhibited conduction without quasi-particles and fast hydrodynamics typical of quark-gluon liquids. The use of standard perturbative QFT approaches proved intractable but they realized that a conformal QFT has a gravitational dual with an AdS4-CFT3 correspondence. Defying reason, they decided to borrow from the mathematical machinery of string theory and perform the same calculation in the dual gravitational space with an extra dimension containing a dyonic black hole (with charge and magnetism), itself a solution of the Einstein-Maxwell equations. To their surprise, the calculation explained the peculiar transport of heat and electricity typical of these “strange metals” and agreed with experiments. In the beginning they must have thought that this is a case of the same math coincidentally describing completely different phenomena and that the dual gravitational space was not real in any sense, but their findings drew more attention and scrutiny—bringing a flood of results that is impossible to review here. Let us just say that these results connected the “strange metal” state of such “Cuprates” to many electron states with Planckian dissipation and minimal viscosity, which is typical of quark/gluon plasmas and fast hydrodynamic configurations, maximally entangled states, “instant thermalization” and maximal entropy production that sets limits on physical computability (Lloyd, 2000). On the AdS side of the duality, strange metals became important as quantum computing devices, enabling string theorists to test string-based hypotheses about quantum gravity, black hole thermodynamics, the information paradox, the emergence of spacetime and more. As an example of the unreasonable success attributed to what Zaanen et al. (2012) terms the AdS-CMT correspondence (with Condensed Matter Physics):

*It is perhaps the greatest success of the AdS/CFT correspondence that Einstein gravity can be used as “generating functional” to determine hydrodynamical equations. The above is of course a particular basic example but at present this “method” is used with significant practical consequence, by re-assessing particularly complicated forms of hydrodynamics.*

Strange metals also provide a strange and novel form of computation, to quote Zaanen (2018).

*We have come to the realization that holography is a mathematical machine that computes observable properties of stuff that is some kind of most extreme, “maximally” entangled form of this compressible quantum matter. Its observable properties do represent “physical” physics. However, this can be very different from anything that you learned in college.*

## Anti-de Sitter Space—Conformal Field Theory. Correspondence as Computation

Philosophy is written in this grand book — I mean the universe — which stands continually open to our gaze, but it cannot be understood unless one first learns to comprehend the language in which it is written. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures, without which it is humanly impossible to understand a single word

of it; without these, one is wandering about in a dark labyrinth (Einstein et al., 1935; Galileo, 1623).

As we said our search for the “mathematics of consciousness” did not appeal to novel mathematics and “first principles”; pretty much the opposite: first, we looked for physics that is rich enough to solve strategically chosen aspects of the meta-problem, and only then asked ourselves whether it has interesting computational or mathematical properties. I argued that the only physics that can solve these “meta problems” finding room for a non-spatial mind in a spatial world is a background independent physics like the one we find in quantum gravity. It turns out that at this level information becomes more basic than space and time and that’s an advantage for a possible “physics of consciousness.” Any physical process computes the temporal evolution of its own states and can be viewed as an analog computer. Useful computing devices operate on “unmanageable” input that we cannot directly relate to, to produce, or distil, output that is relevant to us and that we can relate to.

Shortly after Maldacena’s discovery of the AdS-CFT correspondence, Gubser, Klebanov and Polyakov, followed by Witten, discovered the GKPW rule (or transformation), providing a precise mapping between the physics on both sides of the duality and a universal dictionary of sorts.

*“...the dictionary is also greatly counterintuitive and after all these years still seen as in many regards quite mysterious even by the professional holographists” (Zaanen et al., 2012).*

This is also where the interest of the string theorists resides: “the dream is that condensed matter experiment might be used as an analog quantum computer to test the weak conjecture under circumstances where one does not know how to proceed mathematically” (Zaanen et al., 2012). We see then that a rusted piece of copper may be viewed as performing “analog quantum computation.” This can be used to not only test the validity of the dictionary, but also to make novel predictions.

The AdS-CFT correspondence acts like a computational device that transforms unintelligible “hard” quantum information into “easy” classical information. Commenting on the physicists’ “discovery” of mirror symmetry that Atiyah (2007) attributes to duality between complex geometry (for example, Riemann surfaces and algebraic varieties over the complex numbers) and symplectic geometry (for example, real manifolds that generalize the phase space of classical Hamiltonian mechanics), he concludes:

So this marvelous theorem tells us that easy information on one side (periodic matrices, that can be calculated by classical means) is equivalent to difficult information on the other side (algebraic curves, for whose determinations there is very little information). In physics language, the easy information is what is called classical and the difficult one is what is called quantum. We are thus getting information of a quantum character on one side out of classical calculations on the other side.

Similarly, Zaanen et al. (2012) describes holographic duality as transforming an intractable strongly interacting QFT into a more



phenomenological theory, based more on things that we know and can relate to:

Our interest is in the behavior of an infinite number of strongly interacting quantum degrees of freedom, and “holography” appears as a “generating functional” that is supposedly extremely powerful in revealing the principles controlling “deep emergence” physics, translating it into phenomenological theories of a Landau-esque quality.

Concluding:

From the viewpoint of the mathematically inclined string theorist, this is where the relevancy of the present flirtation with condensed matter physics resides: the dream is that condensed matter experiment might be used as an analog quantum computer to test the weak conjecture under circumstances where one does not know how to proceed mathematically.

## Einstein-Rosen Bridge = Einstein Podolsky and Rosen

The strongest version of the AdS-CFT correspondence is probably advocated by Leonard Susskind (2017), adopting a literal interpretation of the ER = EPR principle that relates EPR, (Einstein et al., 1935) and quantum entanglement to the Einstein-Rosen bridge. On this account, entangled fields in a  $d$ -dimensional QFT serve as boundary conditions to a dual  $d+1$  geometric space, with gravity and connected by wormholes. This is a case of metaphysical necessity and not a case of ordinary empirical equivalence, because entanglement on the surface cannot exist without the wormholes in the bulk and vice versa. The duals do not emerge from each other and are neither grounded in nor constituted by each other but share a more symmetric form of equivalence reminiscent of dual-aspect monism (Vistarini, 2017). For Susskind (2017), any entanglement is accompanied by a corresponding wormhole and he envisions: seeing quantum gravity in a lab equipped with quantum computers and expecting that these will become feasible sometime in the next decade or two.

He then considers a thought experiment including a spherical shell that operates as an elaborate quantum computer. It instantiates a conformal QFT and serves as a boundary to an equivalent “bulk” dual with an extra dimension. The dual extra dimensional space with gravity (“projected” by the boundary CFT) is a necessary consequence of the dual physics. This “bulk” dual space is different than lab-space, and Susskind speculates about the possibility of an observer that is simulated by, or “merges” with, the 2D quantum computer in the shell, thus also entering the dual space to interact back with us. This is why such strong holographic duality can explain the  $\sim P$  Problem intuition, as entering the shell in the lab does not grant access to the “bulk” space dual to the surface quantum computer; to gain access to the “bulk,” the technician must first be converted into a computer program that is inserted into the computer. For the same reason, if our (P)NCC were to consist of some complex conformally entangled CFT state, say with two-dimensional properties, it would be dual to a 3D “bulk” space that cannot be

detected in the “lab-space” of the brain – unless you manage to somehow incorporate an external observer into the 2D “quantum computer.” This works just as well with a 3D CFT and a 4D AdS. Susskind’s take on entanglement suggests viewing such ‘private’ AdS spaces as precisely the kind of “Island Universes” that Lewis argued against. Here consciousness is constituted like space (action of renormalization in CFT) and owes its robustness to a quantum error-correction code that may be different than the one instantiated by ordinary space.

## Holographic Duality and Quantum Error Correction Codes

Preskill (2012) defines “quantum information science” as exploration of the frontier of highly complex quantum states, the “entanglement frontier”. Such systems cannot be simulated by classical computation or given a tractable mathematical description (Feynman, 1982) and can only be simulated with a quantum computer consisting of entangled qubits.  $N$  qubits live in a  $2^N$ -dimensional Hilbert space and, if a gram of highly entangled matter contains Avogadro’s number ( $10^{23}$ ) of such qubits, it generates a fantastically big “frontier” with a ( $2^{10^{23}}$ ) dimensional Hilbert space. Achieving such “quantum supremacy” demands an efficient unitary quantum error correction code and several such codes are currently being considered. Here the AdS-CFT correspondence and holographic duality keep providing novel dazzling insights into the workings of reality. The central mystery of the AdS-CFT correspondence is the “emergence” of gravity and “projection” of the additional spatial dimension (or time dimension in dS-CFT) that dualizes into “renormalization flow” and coarse graining:

It is called “holographic” since the gravitational side has one extra dimension: this “radial direction” connects the boundary to the deep interior [of the de-Sitter space] and has the identification as the scaling direction in the field theory. The claim is that AdS/CFT geometrizes the renormalization group and upon descending deeper in AdS one “sees” the physics at longer times and distances. The deep interior codes for the macroscopic scale (“IR”) (Zannan, 2018).

Here Preskill and others (Patawsky et al., 2015) show that the holographic correspondence establishes an equivalence between quantum error correction codes operating on the boundary QFT and the “robustness” of spacetime. Quantum error correction codes are based on entanglement and protect the information in individual qubits from noise by embedding this information in the entanglement patterns of multiple qubits. Spacetime itself is seen as constituted by entangled qubits of sorts and holographic duality strongly suggests that it owes its stability to a very efficient error correction code: “If such a code exists it must be very special and DARPA, taking notice, is funding efforts to discover such codes with the hope of producing efficient quantum error correcting codes to achieve ‘quantum supremacy’” (Wolchover, 2019). The surprising connection between quantum computation, error correction codes and the constitution of spacetime, made clear by the AdS-CFT duality,

is just the latest demonstration of duality as a gift that keeps on giving. If there is a “physics of consciousness,” then it may be related to the physics of time.

## DISCUSSION

The Strange Metal Theory (SMT) is a consciousness realist theory that handles the meta-challenge by accepting a separate spatiotemporal unfolding of the “radically different” consciousness process and meta-process—while showing that they realize the same information. I believe that what we know and say about consciousness has something to do with the way consciousness *really is*, and showing that we would probably make conscious and problem reports even if we were not conscious threatens this belief and the very foundations of our self-knowledge. However, SMT relies on two parallel processes realizing the same information. Think of a hologram varying continuously in time, creating a 2D holographic movie of sorts, and consider its dual 3D holographic counterpart: in SMT these movies cannot act on each other in any way and a frame in one movie can only be influenced by other frames in the same movie. Entities in one space may map in a complex and non-local manner unto dual entities in the dual space that may be very different.

While it's true that the action of vocal cords generating “conscious reports” are not caused by consciousness but by very different non-conscious 2+1 or 3+1-dimensional processes describable by CFT, in the corresponding 3+1- or 4+1-dimensional dual AdS “phenomenal” space (here a topic-neutral space that can be given a more classical/geometrical interpretation), the conscious reports are caused by consciousness.

If parts of a minimal (P)NCC were shown to display “strange metal” dynamics, that would immediately suggest a solution to the SMP and ~P Problem intuitions. The SMP can be explained by the way that holographic duality manages to conceal the necessary connections between the duals (fact used by Uri Geller in Sec. “Ineffable and Manageable Anti-physicalist Intuitions”) and the ~P Problem (also see Sec. “Ineffable and Manageable Anti-physicalist Intuitions”) can be explained by the strange ontological nature of such dual spaces. Remember Susskind's ER = EPR sphere in which the entangled CFT on the shell in the lab-space demands the existence of a “bulk” physics with an extra spatial dimension and gravity that is just as real but not identical to the lab-space associated with the interior of the shell. When measuring devices that live inside the 2D space perform measurements in this space, dual corresponding measurements must occur in the “bulk”; in order to access that strange “bulk” space you have to live in the 2D shell space, and this is why Susskind's lab technician must merge with the 2D “quantum computer” shell that simulates her in two dimensions (must be uploaded into the 2-D quantum computer) to access the 3D “bulk.” If the NCC is governed by holographic duality and I will shortly consider a similar 2+1-D NCC governed by CFT with a 3+1-D dual, the only way to access the 3D dual space is “inhabiting” the 2D

space. In any event the dual 3D space is not identical to the laboratory space in the way the interior of our skull is. That would also explain why our phenomenal space can only be accessed by us, how two completely different things can be identical and the intuition that consciousness cannot be identical to anything in the brain. Susskind's radical interpretation of entanglement and the AdS-CFT correspondence is not shared by everyone. For example Verlinde (2011) holds that the duals are not completely symmetric. Granting this, the extra space dimension on the AdS side of the duality emerges from the boundary CFT in the same way that thermodynamic variables emerge from microscopic physics. However, Susskind's position is sound, and radical enough to provide a physical explanation to our ~P Problem—which is otherwise very difficult to explain.

Tononi (2008) claims that any system with non zero Phi has inner properties that are only accessible to the system itself but the claim is brute, put in by hand so to speak, and does not result from the physics: likewise Varela's autopoietic approach to biological organisms fails to establish any extra-theoretic inner goings on above and beyond those described by ordinary molecular biology (Dennett, 2011). However, if Susskind is right and the only way a “technician” can enter the “geometric” bulk space is by being “uploaded” unto the 2-Di quantum computer then such AdS spaces can harbor representationally rich “subjective physical facts.” Discovering that our minimal NCC harbors such facts can explain aspects of the privacy of our mental experiences and be philosophically significant. It seems though that we still face a hard problem of consciousness. The dual classical, geometric space is still physical, and we have to explain how consciousness can be identical to something physical whether in that space or in any space. The deep idea here is that scaling with a proper renormalization theory in the CFT creates space in the AdS and if the conscious field is a sort of space that does reside in the “lab-space” then this is the way to construct it. The theory also deconstructs matter and explore the prospects of a mathematical transform that actually solves the SMP by generating structural aspects of the original environmental input (we do not need different individuals to harbor identical phenomenology, rather only a one to one mapping between them that maps unto a common environment). Discovering such a transform would mean that we have managed to identify the isomeric physical correlates of representational content, thus putting us in a better position to understand the physics of phenomenal character. One of the advantages of a topic-neutral physical solution of the meta-problem over other structural topic-neutral solutions to the meta-problem is based on a parsimony argument. Both representational content and phenomenal character are special and novel phenomena. It is unlikely that two highly correlated strange and novel phenomena be generated by two unrelated novel and strange mechanisms. That is, if you discover the physics of representational content, it will probably contain important clues that can help with the phenomenal character problem. If, on the other hand, we find solutions to the meta-problem that are based on standard computational approaches, it is hard to see how the same

class of algorithms that would shed light on the SMP and representational content in general may put you in a better position to solve the phenomenal character problem. After all there is a difference between the map and the territory. Also, the physical solution to the  $\sim P$  problem intuition based on holographic duality cannot be simulated but only *generated* by the right quantum computer.

What about zombies? Suppose we find out that aspects of a minimal (P)NCC instantiate holographic duality *à la* Susskind: would that make zombies inconceivable? Suppose that we also discover that the same mathematical transform explaining the  $\sim P$  Problem intuition (by providing a private AdS space) also solves aspects of the SMP (by showing that the transform preserves elements of the environmental structure), would that make zombies inconceivable? Seems like we can still have holographic “strange-metal” zombies or “AdS zombies” but these are considerably less likely. One reviewer insisted, and rightfully so, that showing that zombies are “highly unlikely” is not enough to undermine physicalism. Here I think that mature sciences and epistemologies determine their scope and limitations from the inside, so to speak, Heisenberg’s Uncertainty and Gödel’s Undecidability are a case in point. Perhaps brain science needs to mature similarly by collaborating with condensed matter physics to establish the existence of (P)NCC systems that contain rich representations that cannot be accessed by ordinary physical measuring devices and necessarily so. I believe that while improving our “epistemic condition” such a scenario still leaves open questions about phenomenal character that makes “holographic zombies” conceivable. What I am suggesting is that discovering that the (P)NCC is one of those rare systems in which information can be realized subjectively can provide us with clues about phenomenal character and the nature of this exotic state of matter. The need to appeal to a background independent physics in order to “place” non-spatial mind in a spatial world suggests a neutral-monistic q-bit based token-physicalism in which the ultimates are not multiply realizable. Such possibly protophenomenal ultimates/qbits can be realized either objectively/scientifically or subjectively/manifestly *and for physical reasons!*. In Awret (2019) I consider the AdS duals of self-referential processes on the CFT surface such as Lloyd (2000) gates which emulate closed time-like curves. So yes, holographic zombies cannot be ruled out but discovering that a minimal (P)NCC instantiates AdS correspondence in a way that solves both the  $\sim P$  and SMP problems should point

us in the right direction. Parsimony suggests that the same novel mechanism responsible for the confinement of these subjectively realized q-bits is also responsible for their intimate and “luminous” self-access which should make us optimistic. If the same physical mechanism would explain the  $\sim P$  and SMP problems while also causing ultimates in the brain a novel kind of self-access AdS zombies would still be conceivable but these are not going to threaten physicalism in the same way if physics discovers private physical spaces associated with the (P)NCC and commensurate with its representational content that are inaccessible to it (external measuring devices). For the conceivability argument to work knowing the *totality* of physical facts implies knowledge of psychological facts but if some of the relevant facts are missing the argument fails. Speculating about the nature of inaccessible facts cannot be used to undermine physicalism.

We began with Castellani and Rickles (2017), tracing the origin of duality to the Yin Yang principle in ancient Chinese cosmology, and I will end with a related possibility. The Claustrum is a good NCC candidate (Crick and Koch, 2005; Torgerson et al., 2015; Reardon, 2017). Suppose that we discover that the Claustrum contains massively entangled electrons describable by a 2D CFT with a dual 3D AdS space in which our phenomenology unfolds. At some point we discover that the information in our 3D universe is actually inscribed on the former surface; then we begin to understand the Claustrum as a surface existing in our “imaginary” 3D space that is actually dual to that space. So, the 3D space harbors the 2D surface that generates it, a bit similarly to a Klein bottle (Brown, 2007). This kind of convoluted topology is reminiscent of the Yin Yang cosmology where each dual harbors a bit of the other dual.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## REFERENCES

- Alter, T. (2015). The structure and dynamics argument against materialism. *Nous* 50, 794–815. doi: 10.1111/nous.12134
- Armstrong, D. M. (1968). The headless woman illusion and the defence of materialism. *Analysis* 29, 48–49. doi: 10.1093/analysis/29.2.48
- Atiyah, M. (2007). *Duality in Mathematics and Physics. Special Lecture: Riemann's Influence in Geometry*. Barcelona: Facultat de Matemàtiques i Estadística, 69–82.
- Awret, U. (2019). Neo-naturalism, conciliatory explanations, and spatiotemporal surprises. *Front. Psychol.* 9:2506. doi: 10.3389/fpsyg.2018.02506
- Brown, C. (2007). Chaos and graphics Taiji variations: Yin and Yang in multiple dimensions. *Comput. Graph.* 31, 142–146. doi: 10.1016/j.cag.2006.10.005
- Byrne, A. (2018). *Transparency and Self-Knowledge*. Oxford: Oxford University Press.
- Carruthers, P. (2017). The illusion of consciousness. *J. Conscious. Stud.* 24, 228–252.
- Castellani, E. (2016). Duality and ‘Particle’ democracy. *Stud. Hist. Philos. Sci. B* 59, 100–108.
- Castellani, E., and Rickles, D. (2017). Introduction to special issue on dualities. *Stud. Hist. Philos. Sci. B* 59, 1–5. doi: 10.1111/j.2151-6952.2011.00116.x
- Chalmers, D. (2002). “Consciousness and its place in nature,” in *Guide to the Philosophy of Mind*, eds S. Stich and T. Warfield (Cambridge: Blackwell), 247–272. Reprinted in D. J. Chalmers, *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press.

- Chalmers, D. J. (2016). "The combination problem for panpsychism," in *Panpsychism*, eds L. Jaskolla and G. Bruntrup (Oxford: Oxford University Press), 179–214.
- Chalmers, D. (2018). The meta-problem of consciousness. *J. Conscious. Stud.* 25, 6–61.
- Chalmers, D. (2020). "Finding space in a Nonspatial World," in *Philosophy Beyond Spacetime*, eds C. Wüthrich, B. Le Bihan, and N. Huggett (Oxford: Oxford University Press).
- Chang, L., and Tsao, D. (2017). The code for facial identity. I the primate brain. *Cell* 169, 1013–1028. doi: 10.1016/j.cell.2017.05.011
- Chouraqi, F. (2011). Temporal thickness in Merleau-Ponty's notes of May 1959. *Chiasmi Int.* 13, 407–427. doi: 10.5840/chiasmi20111324
- Clark, A. (2019). Consciousness as generative entanglement. *J. Philos.* 116, 645–662. doi: 10.1016/j.jphil.2019.1161241
- Coleman, S. (2014). Real combination problem: panpsychism, micro-subjects, and emergence. *Erkenn* 79, 19–44. doi: 10.1007/s10670-013-9431-x
- Cowen, A. S., Chun, M. M., and Kuhl, B. A. (2014). Neural portraits of perception: reconstructing face images from evoked brain activity. *Neuroimage* 94, 12–22. doi: 10.1016/j.neuroimage.2014.03.018
- Crick, F., and Koch, C. (1990). Towards a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275.
- Crick, F. C., and Koch, C. (2005). What is the function of the claustrum? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1271–1279. doi: 10.1098/rstb.2005.1661
- Dennett, D. C. (1992). *Consciousness Explained*. Boston, MA: Little-Brown.
- Dennett, D. C. (2011). Shall we tango? No, but thanks for asking. *J. Conscious. Stud.* 18, 23–34.
- Drake, S. (1957). *Galilei, Galileo 'The Assayer' (Il Saggiatore) in Discoveries and Opinions of Galileo, trans., with Introduction and Notes*. New York, NY: Doubleday Anchor Books.
- Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* 47:696. doi: 10.1103/PhysRev.48.696
- Feynman, R. P. (1982). Simulating physics with computers. *Int. J. Theor. Phys.* 21, 467–488. doi: 10.1007/BF02650179
- Frankish, K. (2019). The meta-problem is the problem of consciousness. *J. Conscious. Stud.* 26, 83–94.
- Freeman, W., and Vitiello, G. (2006). Nonlinear brain dynamics as macroscopic manifestation of underlying many-body dynamics. *Phys. Life Rev.* 3, 93–118. doi: 10.1016/j.plrev.2006.02.001
- Friston, K. (2018). Am i self-conscious? (Or does self-organization entail self-consciousness?). *Front. Psychol.* 9:579. doi: 10.3389/fpsyg.2018.00579
- Galileo, G. (1623). "Il Saggiatore," in *The Philosophy of the Sixteenth and Seventeenth Centuries (1966)*, trans. R. H. Popkin, 65. doi: 10.15585/mmwr.mm7019e3
- Haddad, R., Lapid, H., Harel, D., and Sobel, N. (2008). Measuring smells. *Curr. Opin. Neurobiol.* 18, 438–444. doi: 10.1016/j.conb.2008.09.007
- Hartnoll, S. (2007). Theory of the Nernst effect near quantum phase transitions in condensed matter and in dyonic black holes. *Phys. Rev. B* 76:144502.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430. doi: 10.1126/science.1063736
- Kammerer, F. (2018). Can you believe it? Illusionism and the illusion metaproblem. *Philos. Psychol.* 31, 44–67. doi: 10.1080/09515089.2017.1388361
- Kant, I. (1781/1999). *The Critique of Pure Reason*, eds P. Guyer and A. Wood (Cambridge: Cambridge University Press).
- Kay, K. N., Naselaris, T., Prenger, R. J., and Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature* 452, 352–355. doi: 10.1038/nature06713
- Kleiner, J. (2020). Mathematical models of consciousness. *Entropy* 22:609. doi: 10.3390/e22060609
- Kriegel, U. (2020). "Beyond the neural correlates of consciousness," in *The Oxford Handbook of the Philosophy of Consciousness. Philosophy of Mind*, ed. U. Kriegel (Oxford: Oxford University Press).
- Leibniz, G. W. (1960). *Die Philosophischen Schriften von Gottfried Wilhelm Leibniz*, Vol. 7, ed. C. I. Gerhardt (Hildesheim: Olms).
- Levine, J. (2019). On the meta-problem. *J. Conscious. Stud.* 26, 148–159.
- Lisman, J. (2005). The theta/gamma discrete phase code occurring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus* 15, 913–922. doi: 10.1002/hipo.20121
- Lisman, J., and Buzsáki, G. (2008). A neural coding scheme formed by the combined function of gamma and theta oscillations. *Schizophr. Bull.* 34, 974–980. doi: 10.1093/schbul/sbn060
- Lloyd, S. (2000). Ultimate physical limits to computation. *Nature* 406, 1047–1054. doi: 10.1038/35023282
- Maldacena, J. (1998). The large N limit of superconformal field theories and supergravity. *Adv. Theor. Math. Phys.* 2, 231–252.
- Morch, H. (2020). The phenomenal powers view and the meta-problem of consciousness. *J. Conscious. Stud.* 27, 131–142.
- Papineau, D. (2007). Review of Daniel Stoljar, ignorance and imagination: the epistemic origin of the problem of consciousness. *Notre Dame Philos. Rev.* 465:117. doi: 10.1093/mind/fzn022
- Patawki, F., Yoshida, B., Harlow, D., and Preskill, J. (2015). Holographic quantum error-correcting codes: toy models for the bulk/boundary correspondence. *arXiv [Preprint]*. arXiv:1503.06237v2 doi: 10.1126/sciadv.aaw0092
- Place, U. T. (1956). Is consciousness a brain process? *Br. J. Psychol.* 47, 44–50.
- Polchinski, J. (2015). Dualities of fields and strings. *arXiv [Preprint]*, 24. arXiv:1412.5704v3
- Preskill, J. (2012). Quantum computing and the entanglement frontier. *arXiv [Preprint]*. arXiv:1203.5813
- Reardon, S. (2017). A giant neuron found wrapped around entire mouse brain. *Nature* 543, 14–15. doi: 10.1038/nature.2017.21539
- Reddy, L., Tsuchiya, N., and Serre, T. (2010). Reading the mind's eye: decoding category information during mental imagery. *Neuroimage* 50, 818–825. doi: 10.1016/j.neuroimage.2009.11.084
- Rey, G. (1996). "Towards a projectivist account of conscious experience," in *Conscious Experience*, ed. T. Metzinger (Paderborn: Ferdinand-Schoeningh Verlag).
- Ricciardi, L., and Umezawa, H. (2004). "Brain physics and many-body problems," in *Brain and Being*, eds G. G. Gordon, K. H. Pribram, and G. Vitiello (Amsterdam: John Benjamins Publ. Co.), 255–266.
- Roy, S., and Kafatos, M. (2003). Geometroneurodynamics. *arXiv [Preprint]*. arXiv:quant-ph/0311110
- Smithies, D., and Stoljar, D. (2012). *Introspection and Consciousness*. Oxford: Oxford University Press.
- Stapp, H. (2007). "Quantum mechanical theories of consciousness," in *The Blackwell Companion to Consciousness*, eds M. Velmans and S. Schneider (Oxford: Blackwell), 300–312.
- Stoljar, D. (2006). *Ignorance and Imagination*. Oxford: Oxford University Press, 41–74.
- Stoljar, D. (2009). Précis of ignorance and imagination. *Philos. Phenomenol. Res.* LXXIX, 755–756.
- Stoljar, D. (2016). Introspection and necessity. *Nous* 50, 1–22.
- Susskind, L. (2017). Dear Qubitizers, GR=QM. *arXiv [Preprint]*. arXiv:1708.03040
- Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* 215, 216–242. doi: 10.2307/25470707
- Tononi, G., and Koch, C. (2015). Consciousness: here, there and everywhere? *Philos. Trans. R. Soc. B* 370:20140167.
- Torgerson, C. M., Irimia, A., Goh, S. Y., and Van Horn, J. D. (2015). The DTI connectivity of the human claustrum. *Hum. Brain Mapp.* 36, 827–838. doi: 10.1002/hbm.22667
- VanRullen, R., and Reddy, L. (2019). Reconstructing faces from fMRI patterns using deep generative neural networks. *Commun. Biol.* 2:193.
- Verlinde, E. (2011). On the origin of gravity and the laws of Newton. *J. High Energy Phys.* 2011:29.
- Vistarini, T. (2017). Holographic space and time: emergent in what sense? *Stud. Hist. Philos. Mod. Phys.* B 59, 126–135. doi: 10.1016/j.shpsb.2016.07.002
- Wolchover, N. (2019). *How Space and Time Could Be a Quantum Error-Correcting Code*. Quanta Magazine. Available online at: <https://www.quantamagazine.org/how-space-and-time-could-be-a-quantum-error-correcting-code-20190103/#> (accessed January 29, 2022).
- Zaanan, J., Sun, Y., Liu, Y., and Schlam, K. (2012). *The AdS/CMT Manual for Plumbers and Electricians. Lecture Notes*. Leiden: Universiteit Leiden, 3–40.



Zannen, J. (2018). Planckian dissipation, minimal viscosity and the transport in cuprate strange metals. *arXiv [Preprint]*. arXiv:16–1807.10951. doi: 10.21468/SciPostPhys.6.5.061

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of

the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Awret. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Brain Activity Characteristics of Patients With Disorders of Consciousness in the EEG Resting State Paradigm: A Review

Anna Duszyk-Bogorodzka<sup>1\*</sup>, Magdalena Zieleniewska<sup>2</sup> and Kamila Jankowiak-Siuda<sup>1</sup>

<sup>1</sup> Behavioural Neuroscience Lab, Institute of Psychology, SWPS University of Social Sciences and Humanities, Warsaw, Poland, <sup>2</sup> Faculty of Physics, University of Warsaw, Warsaw, Poland

## OPEN ACCESS

### Edited by:

Olivia Gosseries,  
University of Liège, Belgium

### Reviewed by:

Andrea Piarulli,  
Università degli Studi di Pisa, Italy  
Romulo Fuentes Flores,  
University of Chile, Chile

### \*Correspondence:

Anna Duszyk-Bogorodzka  
aduszyk1@swps.edu.pl

**Received:** 16 January 2021

**Accepted:** 25 April 2022

**Published:** 27 May 2022

### Citation:

Duszyk-Bogorodzka A,  
Zieleniewska M and  
Jankowiak-Siuda K (2022) Brain  
Activity Characteristics of Patients  
With Disorders of Consciousness in  
the EEG Resting State Paradigm:  
A Review.  
Front. Syst. Neurosci. 16:654541.  
doi: 10.3389/fnsys.2022.654541

The assessment of the level of consciousness in disorders of consciousness (DoC) is still one of the most challenging problems in contemporary medicine. Nevertheless, based on the multitude of studies conducted over the last 20 years on resting states based on electroencephalography (EEG) in DoC, it is possible to outline the brain activity profiles related to both patients without preserved consciousness and minimally conscious ones. In the case of patients without preserved consciousness, the dominance of low, mostly delta, frequency, and the marginalization of the higher frequencies were observed, both in terms of the global power of brain activity and in functional connectivity patterns. In turn, the minimally conscious patients revealed the opposite brain activity pattern—the characteristics of higher frequency bands were preserved both in global power and in functional long-distance connections. In this short review, we summarize the state of the art of EEG-based research in the resting state paradigm, in the context of providing potential support to the traditional clinical assessment of the level of consciousness.

**Keywords:** electroencephalography, EEG connectivity, resting state EEG, disorders of consciousness, vegetative state, unresponsive wakefulness syndrome, minimally consciousness state

## 1. INTRODUCTION

After severe brain damage, some patients suffer from disorders of consciousness (DoC), such as coma, unresponsive wakefulness syndrome (UWS), or minimally conscious state (MCS). In patients with UWS, no behavioral signs of consciousness are observed (Giacino et al., 2004). Those in MCS manifest volitional reactions, such as visual pursuit, pain localization (MCS), and/or command following movements (MCS+) but no communication is possible (Bruno et al., 2011). However, as Gosseries et al. (2014) and Thibaut et al. (2021) pointed out, some patients with UWS have brain activity patterns characteristic of MCS but no traces of consciousness are observed in their behavior. These patients are named non-behavioral MCS (MCS \*). Establishing communication and/or regaining the skill of functional use of an object indicates the emergence of MCS (eMCS) and recovery of consciousness (Giacino et al., 2002, 2004). Electroencephalography (EEG) is a common diagnostic technique in routine clinical neurological management. Due to its mobility and wide availability, it offers an indication of a patient's brain functioning, allowing for diagnosis and regular monitoring of their condition. In the case of patients with DoC, brain activity shows huge inter-patient variability. Following a brain injury, EEG can be altered and many abnormal patterns can be observed, e.g., diffuse slowing of activity, polymorphic focal

delta rhythm over damaged regions (Brenner, 2005), asymmetries related to brain damage, burst-suppression pattern, and epileptiform activity (Lehembre et al., 2012b). The clearest way to show the presence of consciousness is to use active experimental procedures, where a patient's direct engagement in following instructions is required. However, in the case of patients with DoC, this paradigm has several limitations, including the fluctuation of attention, aphasia, or sensory and motor deficits. These limitations are not relevant in the resting-state EEG analysis, which requires only maintaining an arousal state, without a necessity for language comprehension or patient's active cooperation. The application of resting-state EEG analysis makes it possible to determine the electrical activity of the awake brain in the absence of tasks and instructions (Bai et al., 2017).

The resting state networks were first described in 1995. Using functional magnetic resonance imaging (fMRI), it has been shown that the brains of people who were not given any task to do experienced slow and synchronous changes in sensorimotoric regions of the brain that showed so-called "connectivity" relation to one's activities while resting (Biswal et al., 1995). Similar patterns of resting state coherence have been documented in other brain areas, called the default mode network (DMN), and these brain regions are involved in self-referential thinking, emotional processing, and recalling memories (Raichle et al., 2001; Raichle, 2015; Shen, 2015). Using positron emission tomography (PET) and fMRI techniques, many other brain functional networks have been demonstrated at rest to show coherent fluctuations at low frequency (0.01–0.1 Hz), related to active sensory processing (Fox et al., 2005; Damoiseaux et al., 2006). However, it is indicated that the fMRI blood-oxygen-level dependent (BOLD) response may represent the summed neuronal activity observable with EEG signal (Whitman et al., 2013). Brain connectivity can be subdivided into structural and functional connectivity. The former focuses on neuroanatomical links between particular brain structures, the latter reflects the statistical dependencies between the activity of spatially distinct brain areas (Friston, 2011). Functional connectivity is usually measured by correlation, coherence, and information theory (Cao et al., 2021). While functional connectivity can be assessed based on BOLD and EEG signals, structural connections cannot be revealed by the EEG technique. Usually, functional connections are quantified either in frequency or time domain, in particular EEG bands, such as  $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$ , and  $\gamma$  or for range of the EEG signal, creating so-called EEG networks.

The analyses derived from the resting-state EEG are related to the fundamental state of the brain and seem to help monitor the patient's state of consciousness (Sitt et al., 2014; Bai et al., 2017, 2021; Sebastiano et al., 2021; Wutzl et al., 2021). Owing to the properties of the EEG measurement itself, the indices obtained reflect other aspects of brain functioning, e.g., the frequencies of characteristic rhythms. The most common approaches to the analysis of resting-state activity include the assessment of spectral power, signal complexity, and functional connectivity. The use of resting state EEG in DoC should make it possible to assess the severity of brain damage, bringing us closer to a prognosis about the patient's further condition and understanding the processes

acting between different states of consciousness. However, the heterogeneity of patients with DoC and the current means of application of analytical algorithms to resting EEG signals significantly impede data interpretation.

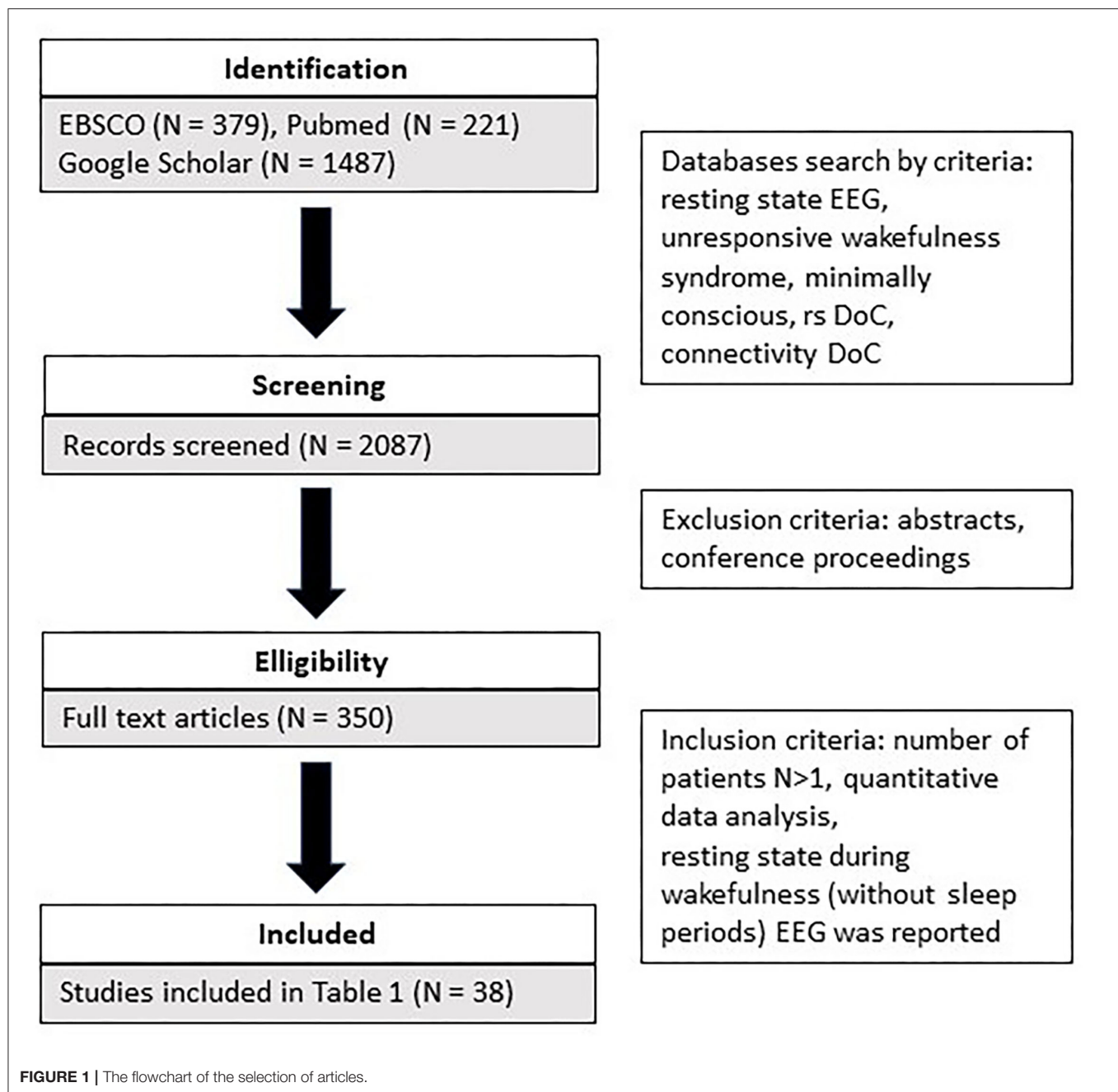
Therefore, the aim of this study is to summarize and discuss EEG resting data based on 38 studies related to the abovementioned entities. As well as outlining aspects of EEG activity that are crucial to the differentiation between unconscious and minimally conscious patients, we also show which of them have prognostic properties.

## 2. METHOD: SEARCH STRATEGY

The search covered three literature databases (EBSCO, Pubmed, Google Scholar). The following keywords were used as search criteria: resting state EEG and unresponsive wakefulness syndrome and minimally conscious, rs DoC, connectivity DoC. In addition, we identified additional references from those provided in the retrieved articles. A total of 2,081 articles were identified (refer to **Figure 1**). Subsequently, we screened the retrieved articles for the following inclusion criteria: (a) they were peer-reviewed or conference proceedings and abstracts; b) more than one patient participated in the study; c) only quantitative data analysis was used; d) resting state during wakefulness (without sleep periods) EEG was reported. A detailed comparison of all reviewed studies is summarized in **Table 1**.

## 3. SPECTRAL CHARACTERISTICS OF EEG IN DoC

One of the most popular methods used to assess oscillatory brain activity is the analysis of power spectrum density. The power of brain activity was reduced overall in patients with DoC compared with healthy controls and those with severe neurological cognitive damage (Leon-Carrion et al., 2008; Babiloni et al., 2009; Naro et al., 2016, 2018). A more detailed analysis showed an increase in slower band power ( $\delta$ ,  $\theta$ ) and a prominent drop-off in higher bands ( $\alpha$ ,  $\beta$ ) in the patient group, so the decrease of power is not uniform (Chennu et al., 2014; van den Brink et al., 2018), refer to **Table 1**. The power of faster oscillations is significantly reduced in patients with DoC (Coleman et al., 2005; Babiloni et al., 2009; Chennu et al., 2014; Naro et al., 2018; Stefan et al., 2018). In the case of patients with UWS, in comparison with those with MCS, brain activity is dominated by the low oscillation of the  $\delta$  band (Lehembre et al., 2012a; Höller et al., 2014; Sitt et al., 2014; Naro et al., 2016, 2018; Piarulli et al., 2016; Stefan et al., 2018; Lutkenhoff et al., 2019) (refer to **Figure 2**). A significant increase in the  $\delta$  band in patients with UWS in comparison with MCS and healthy controls was observed mainly in frontal areas (Lehembre et al., 2012a; Naro et al., 2016). As noted by Chennu et al. (2014) and Naro et al. (2018), about 80% of the overall spectral power was concentrated within the lowest EEG band, suggesting a severe impairment of cortico-thalamo connections (Gloor et al., 1977).



In turn, patients with MCS were usually characterized by stronger activity in higher bands ( $\theta$ ,  $\alpha$ ,  $\beta$ ) and weaker in the  $\delta$  band compared with UWS (Lechinger et al., 2013; Sitt et al., 2014; Naro et al., 2016, 2018; Lei et al., 2022). Thibaut et al. (2021) showed similar differences between patients with MCS\* and UWS. In particular, the presence of an  $\alpha$  band in the parietal areas seems to be crucial for consciousness (Babiloni et al., 2009; Lehembre et al., 2012a; Sitt et al., 2014; Naro et al., 2016) (refer to **Figure 2**). As shown by Babiloni et al. (2009), in the case of patients with UWS who had not recovered, the  $\alpha$  band in the posterior areas was

extremely weak, which seemed to be a reliable predictor of a bad outcome.

The results mentioned above are in line with the clinical observations based on visual inspection of the EEG signal of patients with DoC. Schiff (2016) proposed a theoretical framework to assess the integrity of thalamo-cortical and cortico-cortical networks. According to this theory, a higher level of anterior forebrain corticothalamic integrity is correlated with a higher level of consciousness, independent of the degree of injury to other brain areas (Schiff, 2016; Forgacs et al., 2017, 2020). The condition of thalamo-cortical integrity after severe



**TABLE 1 |** Summary of studies based on resting-state EEG.

Objectives	References	Patients	Method	Global results	$\delta$	$\theta$	$\alpha$	$\beta$	$\gamma$
Spectral characteristics	Coleman et al., 2005	6 UWS, 4 MCS	Power ratio index	UWS>MCS					
	Leon-Carrion et al., 2008	7 MCS, 9 SND	Power spectrum, Loreta		MCS>SND	MCS>SND		MCS<SDN	
	Babiloni et al., 2009	50 UWS, 30 HC	Power spectrum, Loreta				UWS<HC, UWS-REC>NON-REC		
	Lehembre et al., 2012a	10 UWS, 21 MCS	Power spectrum		UWS>MCS	UWS<MCS	UWS<MCS		
	Lechinger et al., 2013	8 VS, 9 MCS, 14 HC	Power spectrum		UWS>HC	UWS>HC	UWS, MCS<HC		
	Chennu et al., 2014	13 VS, 19 MCS, 26 HC	Power spectrum - power contribution		UWS, MCS>HC		UWS, MCS<HC	UWS, MCS<HC	
	Sitt et al., 2014	75 UWS, 68 MCS, 24 CS, 14 HC	Spectral power		UWS>MCS, CS	UWS<MCS, CS	UWS<MCS, CS	UWS, MCS<CS	
	Varotto et al., 2014	18 UWS, 10 HC	Power spectrum		DoC>HC	DoC<HC	DoC<HC	DoC<HC	
	Naro et al., 2016	6 UWS, 7 MCS, 10 HC	Source power (Loreta)		DoC<HC; UWS>MCS	DoC<HC; UWS<MCS	DoC<HC; UWS<MCS	DoC<HC; UWS<MCS	DoC<HC; UWS<MCS
	Piarulli et al., 2016	6 UWS, 6 MCS	Spectral power		UWS>MCS	UWS<MCS	UWS<MCS	UWS<MCS	
	Schorr et al., 2016	58 UWS, 15 MCS, 24 HC	Spectral power		DoC>HC		DoC<HC	DoC<HC	
	van den Brink et al., 2018	16 DoC, 16 HC	Spectral amplitude		DoC<HC	DoC<HC	DoC>HC	DoC>HC	
	Naro et al., 2018	17 UWS, 15 MCS	Spectral power		UWS>MCS		UWS>MCS		
	Stefan et al., 2018	51 UWS, 11 MCS	Spectral power		UWS>MCS		UWS<MCS		
	Lutkenhoff et al., 2019	37 UWS, 24 MCS	Spectral power	UWS<MCS	UWS>MCS-	UWS<MCS	UWS, MCS+<MCS-; UWS>MCS+		MCS+<MCS-
	Bai et al., 2021	37 UWS, 25 MCS, 25 HC	Spectral power				MCS-V>MCS-NV	MCS-M>MCS-NM	
	Thibaut et al., 2021	11 UWS, 15 MCS*, 54 MCS, 33 HC	Spectral power		UWS>MCS*	UWS<MCS*	UWS<MCS*		
	Lei et al., 2022	19 UWS, 21 MCS	Relative wavelet energy		UWS>MCS		UWS<MCS	UWS<MCS, EMCS	
Signal complexity	Schnakers et al., 2008	16 Coma, 13 UWS, 13 EMCS, 30 MCS	Bispectral index	UWS<MCS					
	Sarà and Pistoia, 2010	10 HC, 10 DoC	Approximate entropy	DoC<HC					
	Gosseries et al., 2011	6 Coma, 24 UWS, 26 MCS, 16 HC	Spectral entropy, State entropy, Response entropy	UWS, MCS<HC; UWS<MCS					
	Sarà et al., 2011	38 UWS, 40 HC	Approximate entropy	UWS<HC					
	Wu et al., 2011a	21 UWS, 16 MCS, 30 HC	Lempel-Ziv complexity (LZC), Approximate Entropy (AE), Cross-approximate entropy (CAE)	UWS, MCS<HC (LZC, AE, CAE)					
	Wu et al., 2011b	30 UWS, 20 MCS, 30 HC	Approximate entropy (AC), Cross-approximate entropy (CAE)	UWS<MCS, HC (CAE)					

(Continued)

TABLE 1 | Continued

Objectives	References	Patients	Method	Global results	$\delta$	$\theta$	$\alpha$	$\beta$	$\gamma$
	King et al., 2013	75 UWS, 68 MCS, 24 CS, 14 HC	Permutation entropy	UWS<MCS, CS					
	Sitt et al., 2014	75 UWS, 68 MCS, 24 CS, 14 HC	Kolmogorov-Chaitin complexity (KC), Permutation entropy (PE), Spectral entropy (SE)	UWS<MCS (KC), UWS<CS, MCS (SE)		UWS<MCS, CS (PE)			
	Piarulli et al., 2016	6 UWS, 6 MCS	Spectral entropy (SE), Wavelet decomposition of spectral entropy time-courses (TC)	UWS<MCS (SE); UWS<MCS (TC)					
	Stefan et al., 2018	51 UWS, 11 MCS	Approximate Entropy, Permutation Entropy (PEn)	ApEn: UWS<MCS			PEn: UWS<MCS		
	Lee et al., 2019a	30 isoflurane anesthesia, 15 ketamine anesthesia, 15 MCS, 27 UWS, 73 HC	Phase lag entropy (PLE): mean values and topographic similarities	mean PLE: UWS<MCS, topographic similarity of PLE: MCS, UWS<HC					
	Lei et al., 2022	19 UWS, 21 MCS	Approximate Entropy (AE), Sample Entropy (SA), Lempel-Ziv Complexity (LZC)	UWS<MCS (AE, SA, LZC); UWS<EMCS (AE, SA, LZC); MCS<EMCS (AE, SA, LZC)					
Functional connectivity networks	Pollonini et al., 2010	7 MCS, 9 SND	Coherence (C), Granger causality (GC)	MCS<SDN (C)	MCS<SDN (C); MCS>SND (GC)	MCS<SDN (C)	MCS<SDN (C)	MCS<SDN (C); MCS<SDN (GC)	
	Fingelkurts et al., 2012	14 UWS, 7 MCS, 5 HC	Operational Architectonics	UWS, MCS<HC; UWS<MCS				UWS<HC	
	Lehembre et al., 2012a	10 UWS, 21 MCS	Coherency (C), Imaginary part of coherency (IC), Phase Lag Index (PLI)		UWS<MCS (IC)	UWS<MCS (IC, PLI)	UWS<MCS (IC, PLI)		
	Leon-Carrion et al., 2012	7 MCS, 9 SND	Coherence (C), Granger causality	MCS<SND (C)	MCS<SND (C)	MCS<SND (C)	MCS<SND (C)	MCS<SND (C)	MCS<SND (C)
	King et al., 2013	75 UWS, 68 MCS, 24 CS, 14 HC	weighted symbolic mutual information (wSMI)	UWS<MCS, CS, HC					
	Cavinato et al., 2014	10 USW, 16 MCS, 15 HC	Coherence			UWS>MCS, HC	MCS, HC>UWS		
	Marinazzo et al., 2014	11 UWS, 10 MCS, 5 EMCS, 10 HC	Multivariate Granger Causality, Transfer entropy	UWS>MCS, EMCS, HC (TE); MCS>EMCS, HC; EMCS>HC					
	Höller et al., 2014	27 UWS, 22 MCS, 23 HC	44 biomarkers from resting EEG and machine learning		UWS>MCS	MCS>UWS	MCS>UWS	UWS>MCS	

(Continued)

TABLE 1 | Continued

Objectives	References	Patients	Method	Global results	$\delta$	$\theta$	$\alpha$	$\beta$	$\gamma$
	Sitt et al., 2014	68 MCS, 75 VS, 24 CS, 14 HC	Phase-locking index (PLI), Weighted symbolic mutual information (wSMI)		UWS>MCS (PLI)	UWS<MCS (wSMI)	UWS<MCS (wSMI)		
	Chennu et al., 2014	13 UWS, 19 MCS, 26 HC	Debiased weighted phase lag index		UWS, MCS>HC	UWS, MCS>HC	UWS, MCS<HC,		
	Varotto et al., 2014	18 UWS, 10 HC	Partial directed coherence		DoC<HC		DoC>HC		
	Schorr et al., 2016	58 UWS, 15 MCS, 24 HC	Coherence			MCS<HC	UWS, MCS<HC	UWS, MCS<HC	
	Chennu et al., 2017	23 UWS, 17 MCS-, 49 MCS+, 11 EMCS, 4 LIS, 26 HC	Debiased weighted phase lag index		MCS->MCS+		UWS<MCS-		
	Engemann et al., 2018	148 UWS, 179 MCS, 66 HC	Weighted symbolic mutual information			UWS≠MCS			
	Bareham et al., 2018	4 DoC, longitudinal study	Debiased weighted phase lag index		MCS-> MCS		UWS<MCS-		
	van den Brink et al., 2018	16 DoC patients, 16 HC	Correlation of orthogonalized amplitude envelopes		DoC>HC	DoC>HC	DoC<HC		
	Naro et al., 2018	17 UWS, 15 MCS	Debiased weighted phase lag index				UWS<MCS		UWS<MCS; UWS<MCS (in time-course)
	Stefan et al., 2018	51 UWS, 11 MCS	Coherence (C), Weighted symbolic mutual information (wSMI)				UWS>MCS (C)	UWS>MCS (C)	
	Rizkallah et al., 2019	9 UWS, 17 MCS-, 29 MCS+, 6 EMCS, 21 HC	Dynamic functional networks –segregation (C) and integration (P)	MCS-<HC (P)	UWS, MCS<HC (C, P)	EMCS, MCS, UWS<HC (C); MCS, UWS<HC; MCS-, UWS<EMCS; MCS+>MCS- (P)	EMCS, MCS+<HC (P)	MCS+<HC (P)	UWS, MCS<HC; MCS+<EMCS (P)
	Cai et al., 2020	35 UWS, 19 MCS, 23 HC	Multiplex graph metrics, Multiplex clustering coefficient (MCC), Multiplex participation coefficient (MPC)	UWS>MCS, HC; MCS>HC; (MPC); UWS<MCS, HC (MCC)					
	Bai et al., 2021	37 UWS, 25 MCS, 25 HC	Time-delay embedded Hidden Markov Model (TDE-HMM), Coherence (C)	DoC<HC; MCS<UWS (state dynamics);			DoC<HC (C)		
	Naro et al., 2021	17 UWS, 15 MCS	Multiplex and multilayer network metrics (network topology, NT; multiplex network heterogeneity, MNH)	UWS<MCS (NT)	UWS>MCS (MNH)	UWS>MCS (MNH)	UWS<MCS (MNH)	UWS<MCS (MNH)	
	Thibaut et al., 2021	11 UWS, 15 MCS*, 54 MCS, 33 HC	Debiased weighted phase index		UWS>MCS*	MCS* <MCS	UWS< MCS*, MCS, HC; MCS* <MCS; MCS<HC		

DoC, disorders of consciousness; HC, healthy controls; CS, conscious patients; SND, severe neurocognitive disorders; LIS, locked-in syndrome; MCS, minimally conscious state; MCS-V/MCS-NV, presence of purposeful/no purposeful visual responses; MCS-M/MCS-NM presence of purposeful/no purposeful motor responses; EMCS, emergence from MCS; UWS, vegetative state; REC/NON-REC, recovery/non-recovery patient.

brain injury is reflected in four different EEG regimes assessed *via* the signal spectrum related to predominant frequency power: A type—mostly  $\delta$  activity (below 4 Hz); B type— $\theta$  activity (5–7 Hz); C type— $\theta$ , and  $\beta$  (15–40 Hz) activity; D type— $\alpha$  (8–12 Hz) and  $\beta$  activity. Based on a model “ABCD” spectral pattern, “A” indicates complete loss of corticothalamic integrity; “D” indicates full recovery of corticothalamic integrity; while “B” and “C” represent interim regimes with distinct physiological foundations (Schiff, 2016; Forgacs et al., 2017). The low oscillations of  $\delta$  activity, uniform over the scalp, observed in patients with UWS correspond to an “A” type EEG regime and are interpreted as corticothalamic deafferentation—caused either by the impairment of synaptic transmission (van Putten and Hofmeijer, 2016) or by white matter lesions (Gloor et al., 1977). Regarding patients with MCS, the vast majority of studies showed stronger  $\theta$  activity in MCS in comparison with UWS. This seems to be related to the preservation of a certain amount of afferent input to cortical neurons, which corresponds to “B” and “C” type EEG regimes. Such a hypothesis confirms stronger  $\theta$  activity in patients with MCS in comparison to severe neurocognitive disorders (SND), which suggests that loss of consciousness is related to deeper brain injury, a low level of afferent input to cortical neurons, and stronger spontaneous oscillation of Layer V pyramidal cells. In comparison with patients with UWS, MCS brain activity is characterized by stronger  $\alpha$  band oscillations that reflect normal tonic firing, and seem to be related to preserved corticothalamic integrity but at a lower level than in the controls (refer to **Table 1**).

#### 4. SIGNAL COMPLEXITY IN PATIENTS WITH DoC

The concept of entropy, first proposed as a thermodynamic principle, is now used widely to quantify the effects of anesthetic drugs on brain activity. EEG signal entropy methods may be based on time series (e.g., Approximate Entropy) and phase space analysis (e.g., Phase Lag Entropy), or on the frequency spectrum (e.g., Spectral Entropy) (Bandt and Pompe, 2002). Although these algorithms are sensitive to different properties of the signal, such as spectral bandwidth and amplitude distribution, and are generally difficult to interpret, they tend to have lower values if the signal slows down and becomes less complex (Ferenets et al., 2006). Different measures of entropy were used to assess differences in signal complexity among patients with DoC (**Table 1**). In most of the studies, these methods show a coherent picture—the global EEG signal complexity was lower in patients with UWS than in those with MCS (Sarà and Pistoia, 2010; Gosseries et al., 2011; Sarà et al., 2011; Wu et al., 2011b; Sitt et al., 2014; Stefan et al., 2018; Lee et al., 2019a) (refer to **Figure 2**). In particular, patients with MCS were characterized by more complex interactions between the activity of brain regions (Lee et al., 2019a), and by higher variability and periodicity of spectral entropy over time (Piarulli et al., 2016). As shown by Lei et al. (2022) in the longitudinal scheme, different entropy measures (Approximate Entropy, Sample Entropy, and Lempel-Ziv complexity) increased non-monotonically with the inflection

point when the patients transferred from UWS to MCS. These results suggest that the transition from unconscious to conscious is not a linear phenomenon.

The best discrimination between UWS and conscious patient groups was obtained in  $\theta$  (Sitt et al., 2014) and  $\alpha$  bands (Stefan et al., 2018) (refer to **Table 1**). Moreover, lower values of complexity measures were observed in many studies on the effects of anesthetics (Liang et al., 2015) and sleep (Bruce et al., 2009) during the loss of consciousness.

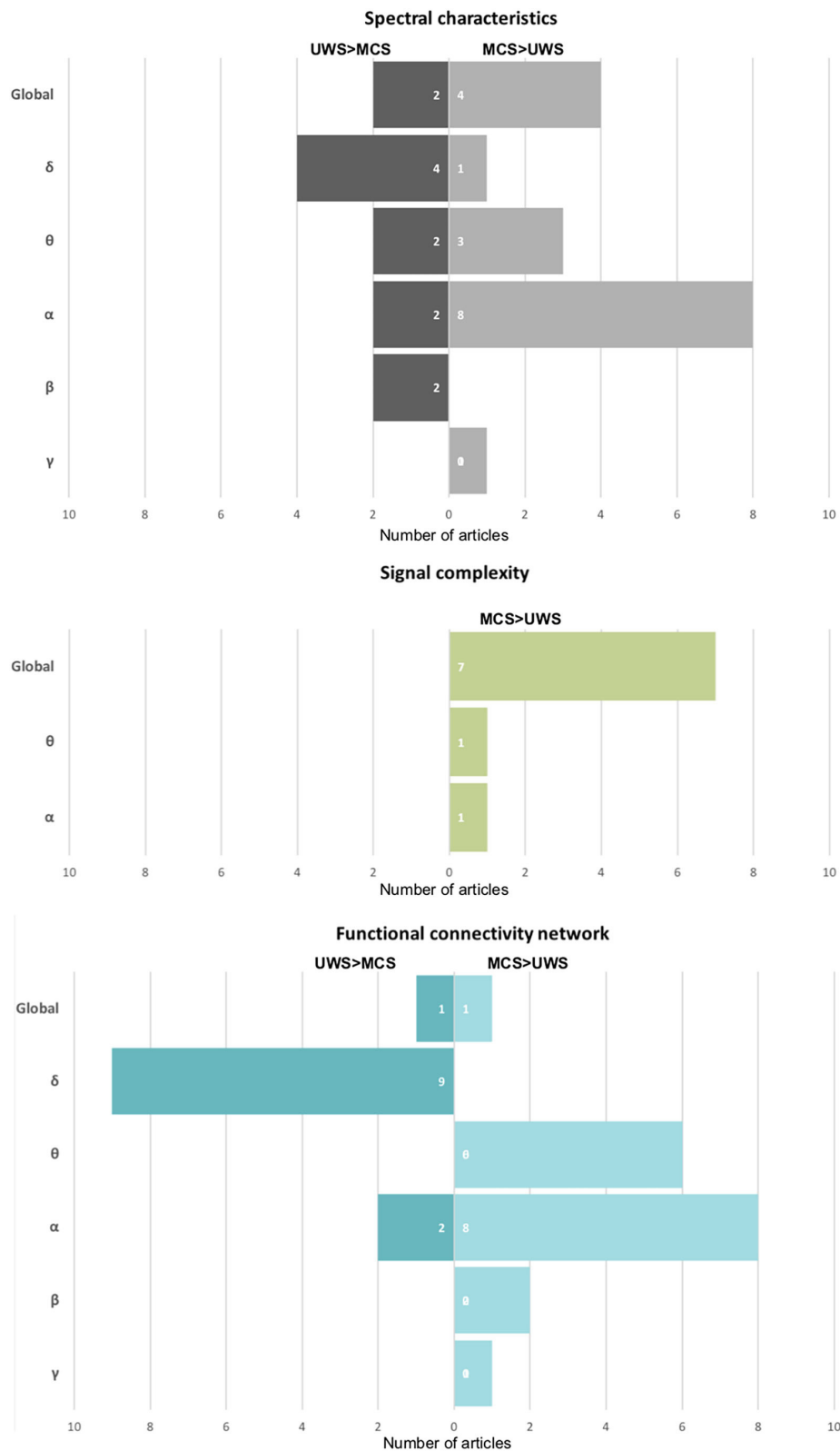
### 5. CHANGES IN NETWORK FUNCTIONAL CONNECTIVITY IN PATIENTS WITH DoC

Brain function relies on the co-activation of several structures in functional networks (Damoiseaux et al., 2006). Some of them (i.e., the default mode network, DMN) seem to be related to the mechanisms of consciousness (Mason et al., 2007; Boly et al., 2008; Boveroux et al., 2010). While the complexity of brain networks is indirectly reflected by entropy measures as discussed in the previous section, the most detailed information about the functioning of the brain is provided by an examination of network functional connections. In the case of the EEG signal, there are many approaches to the assessment of this phenomenon, which have been applied to assess brain function in patients with DoC (for review refer to Bai et al., 2017; Corchs et al., 2019; **Table 1**). In the following subsection, we turn our attention to network functional connectivity in UWS and MCS.

#### 5.1. Changes in Network Functional Connectivity in Patients With UWS

Considered at a global level, patients with UWS exhibit general deficits in information sharing, observed mainly over central and posterior areas and across medium-to-long distances (King et al., 2013). This has been interpreted as possibly related to disrupted thalamo-cortical and cortico-cortical interactions due to diffuse white matter lesions common in patients with persistent UWS and hypoactivation of the DMN, encompassing the precuneus and posterior cingulate cortex located over centro-parietal regions. This interpretation was supported by fMRI studies (Vanhaudenhuyse et al., 2009; Demertzi et al., 2015). Moreover, the analysis of particular bands showed more detailed, specific functional connectivity patterns related to the preserved level of consciousness. The most common pattern in patients with UWS, as compared with MCS and healthy controls, contains a decrease in the number of functional connections in higher bands (mostly  $\alpha$  and  $\theta$ ) and stronger communication in the  $\delta$  band—similar to the spectral power in these bands, as described in Section 3 (Höller et al., 2014; Sitt et al., 2014). Reduced functional connectivity patterns in frequencies higher than  $\delta$  (mostly  $\alpha$  and  $\theta$ ) in patients with UWS, compared with MCS, have been revealed in many studies (Lehembre et al., 2012a; Cavinato et al., 2014; Sitt et al., 2014; Chennu et al., 2017; Naro et al., 2018; van den Brink et al., 2018) (refer to **Figure 2** and **Table 1**). The most significant impairment of information flow was observed over fronto-posterior areas (Lehembre et al., 2012a) in both  $\theta$  (Lehembre et al., 2012a) and  $\alpha$  (Cavinato et al., 2014; Chennu





**FIGURE 2 |** The summary of the number of publications shows a particular effect on spectral characteristics, signal complexity, and functional connectivity networks. On the left side of the chart: a stronger effect in UWS in comparison to MCS, on the right one—opposite—a stronger effect in MCS than in patients with UWS.

et al., 2017), which are presumably related to the preserved consciousness (Sitt et al., 2014; Chennu et al., 2017) (discussed in the next Section 5.2).

In the study of Höller et al. (2014), patients with UWS showed increased long-distance functional connectivity between frontal and parietal/occipital regions not only in  $\theta$ , but, surprisingly, also in  $\beta$  frequencies, similar to the findings of Stefan et al. (2018), who reported stronger global coherence in the  $\beta$  band. It is difficult to interpret the preserved functional connectivity in higher bands in these patients, which is commonly linked with some functional networks observed in the resting state (Brookes et al., 2011). However, it should be noted that the EEG signal of patients with DoC has unusual characteristics and may be contaminated by massive artifacts, especially in the higher frequency range.

## 5.2. Changes in Network Functional Connectivity in Patients With MCS

Similar to healthy individuals, patients with MCS reveal the presence of functional connectivity networks, especially fronto-posterior (Bai et al., 2021; Naro et al., 2021), at higher frequencies. However, the strength of these connections was on average weaker in comparison with healthy controls (Fingelkurts et al., 2012; Chennu et al., 2014, 2017; Schorr et al., 2016) and patients with the severe neurocognitive disorder (Leon-Carrion et al., 2012). These results were also convergent with those of Chennu et al. (2014), who reported stronger functional connections in  $\alpha$  and weaker connections in  $\delta$  in healthy controls. In turn, the strength of fronto-parietal  $\alpha$  connections was correlated with the level of preserved consciousness (from MCS-, MCS+, eMCS, locked-in syndrome, to controls) (Chennu et al., 2017). Moreover, functional networks in the  $\alpha$  band were more pronounced on the local and global level in conscious controls than in patients with DoC (Chennu et al., 2017). However, some studies report preserved functional connectivity not only in  $\alpha$ , but also in two other bands ( $\theta$ ,  $\beta$ ) as a neural marker of consciousness in MCS (Lehembre et al., 2012a; Höller et al., 2014; Sitt et al., 2014; Naro et al., 2018). The stronger connectivity in  $\theta$  and  $\alpha$  bands is interpreted as the most important characteristic of the functional networks of conscious individuals (Lehembre et al., 2012a; Höller et al., 2014; Sitt et al., 2014). Regarding functional connections in  $\theta$ , two interpretations are possible: on the one hand, increased coherence in this band is observed in patients who have suffered structural brain injuries over the entire scalp (Schiff et al., 2014). This is in line with the observation of stronger functional connectivity measures in the  $\theta$  band observed in patients with DoC in general, in comparison with controls (Chennu et al., 2014; van den Brink et al., 2018). On the other hand, increased connectivity in  $\theta$  and  $\alpha$  in MCS compared to UWS occurs over mesio-parietal areas and may be related to the key role of fronto-parietal networks in a “global workspace” in generating conscious states, and it may, therefore, be expected that their activity is disrupted in patients with UWS (Sitt et al., 2014). These results are in line with those obtained by Bai et al. (2021), who showed reduced functional connectivity

in patients with DoC in comparison to HC, especially fronto-parietal connections in  $\alpha$  band. Moreover, they described four key transient brain states preserved in patients with DoC: anterior, posterior, sensorimotor, and visual. The dynamics of switching between these states were different in patients with UWS and MCS: the formers spent more time in the anterior state, while the latter—in sensorimotor, visual, and posterior ones. Moreover, the transition between the sensory and posterior states was suppressed in patients with UWS as compared to MCS. Among those with MCS, two subgroups of patients are distinguished: Patients with MCS— showing volitional reactions such as object localization and visual pursuit, and patients with MCS+ reacting to commands. A detailed examination of brain connections in these two subgroups was conducted by Chennu et al. (2017), who observed similar  $\alpha$  connectivity in patients with MCS— and MCS+, but, in the case of the latter group, they seemed to be situated over the frontal and parietal areas. Importantly, the strength of frontal, central, and parietal functional connections was related to brain metabolism as measured by positron emission tomography. Moreover, patients with MCS— connectivity in the  $\delta$  band was much stronger than in MCS+, which suggests a lower degree of cortical deafferentation in patients showing high-level behavioral responses. Similarly, the stronger connectivity in  $\alpha$  and weaker in  $\delta$  band was observed in patients with MCS\* in comparison to UWS. On the other hand, patients with MCS had stronger connectivity in  $\theta$  band than MCS\*, which is in line with the previous hypotheses on the role of particular EEG bands in mechanisms of consciousness. To summarize, the functional connectivity patterns of patients with MCS are characterized by preserved structure at higher frequencies. In comparison with healthy controls, however, their strength is reduced. This seems to be in line with the fMRI results that show that the strength of correlation in DMN is related to the level of consciousness (Vanhaudenhuyse et al., 2009).

## 6. RESTING-STATE EEG MEASURES IN SLEEP AND AFTER ANESTHETIC ADMINISTRATION

The patterns of brain activity described in patients with DoC corresponds to the characteristics of brain activity in loss of consciousness in other contexts. The increase in low frequency bands, observed particularly in patients with UWS, has been reported in other cases of loss or alteration of consciousness: during deep anesthesia (Steriade et al., 1993; Purdon et al., 2013), temporal lobe seizure (Englot et al., 2010), and slow wave sleep (Timofeev et al., 2001), seen as functional deafferentation. In numerous studies, the increase of low frequency power was associated with anesthetic administration and loss of consciousness, e.g., during induction with sevoflurane and propofol in combination with remifentanyl, there is a gradual increase in  $\theta$  and  $\delta$  bands in anterior areas, which spreads to posterior ones (Gugino et al., 2001). As isoflurane concentration increases, the power spectrum below 4 Hz strengthens (Hagihira, 2015). In turn, Purdon et al. (2013) and Banks et al. (2020) reported the propofol-induced increase of  $\delta$  and  $\alpha$  bands

in anterior regions. Similarly, stronger  $\delta$ -range activity was reported during a complex-partial seizure (Englot et al., 2010), accompanied by a decrease in cerebral blood flow in the bilateral frontal and parietal cortices, than during a temporal lobe seizure with preserved consciousness (Blumenfeld et al., 2004). The EEG signal of patients with UWS is characterized by lower values of complexity measures than in MCS and healthy controls. Similar levels of complexity and regularity of brain activity were observed in deep anesthesia (Liang et al., 2015; Shin et al., 2020) and during sleep (Burioka et al., 2005; Bruce et al., 2009). The comparison of 12 different entropy algorithms applied to EEG data registered in anesthesia induced by gamma-aminobutyric acid-ergic agents (isoflurane or sevoflurane anesthesia) showed in each case a similar effect of decreasing entropy measures as anesthesia deepened, which then increased during recovery (Liang et al., 2015). With higher doses of anesthetic drugs, the EEG signals become more regular and its frequency pattern simplifies. In turn, the analysis of regularity and spectral content of EEG signal during sleep showed lower values of the entropy measures in relation to slowing brain activity during the subsequent non-rapid eye movement (NREM) stages and gradual loosening of consciousness (Burioka et al., 2005), which was negatively correlated with the power of the  $\delta$  band (Bruce et al., 2009).

The vast majority of studies indicate stronger  $\delta$  functional connections in UWS than in MCS and the presence of higher frequency bands in patients with MCS, which are crucial for preserved consciousness. Such hypotheses are in line with the studies of Lee et al. (2019b), who showed emphasized phase-locking at low frequencies, particularly in posterior cortical areas, in conscious experiences during NREM sleep in comparison with the unconscious. These were interpreted as cortical bistability in thalamocortical circuits connected with loss of consciousness. Notably, the conscious experience was accompanied by higher phase-locked values in the  $\alpha$  and  $\beta$  bands. The importance of  $\delta$  band functional connectivity in loss of consciousness was indicated during propofol-induced sedation as well, demonstrating the lower level of responsiveness, stronger  $\delta$  power, and local  $\delta$  functional connectivity in parietal regions (Lee et al., 2017). On the other hand, a preserved pattern of  $\alpha$  activity seems to be connected with consciousness (Lee et al., 2019b; Banks et al., 2020). Loss of consciousness during the N2 and N3 stages of NREM and under anesthesia was associated with  $\alpha$ -band connectivity limited to anterior regions (Banks et al., 2020). In comparison, during wakefulness  $\alpha$ -band functional connectivity was mainly observed within and between temporal and parietal regions. Interestingly, the authors obtained no differences in long-distance connections between conscious and unconscious conditions, such as were observed in patients with MCS in comparison with UWS.

## 7. PROGNOSTIC VALUE OF RESTING-STATE EEG MEASURES

Finally, some of the resting-state EEG measures have prognostic value for patients' long-term recovery. One of them is the power of  $\alpha$  oscillations, which proved to be related to recovery

during follow-up (Babiloni et al., 2009; Fingelkurts et al., 2011; O'Donnell et al., 2021). To identify other prognostic indicators of a favorable outcome for patients with DoC, an analysis of brain oscillatory microstates was conducted (Fingelkurts et al., 2011). Its results indicated that a bad outcome was related to an increased rigidity in brain activity, which was reflected by a lower diversity and variability of EEG signal, and a higher probability of occurrence of  $\delta$ , as well as slow  $\theta$  oscillations in non-survivors. Similarly, Stefan et al. (2018), Sarà and Pistoia (2010), and Sarà et al. (2011) showed that the complexity of the EEG signal allowed prediction of patients' outcomes; however, these results are not coherent, and the study of Gosseries et al. (2011) did not confirm this hypothesis. Regarding the functional networks, Schorr et al. (2016), Chennu et al. (2017), Bareham et al. (2018), and Bareham et al. (2020) demonstrated that the functional connectivity may also have a prognostic value. As shown by Bareham et al. (2018) in a longitudinal examination of four patients with DoC, the process of regaining consciousness was accompanied by the establishment of the functional connectivity in the  $\alpha$  band and decreasing in the  $\delta$  range. These results were partially confirmed 2 years later by Bareham et al. (2020) who indicated functional connectivity in the  $\alpha$  band, the power of  $\theta$  band, age of a patient, and arousal levels as the most important predictors of behavioral recovery. On the other hand, Chennu et al. (2017) observed a clear relationship between EEG-based  $\delta$  connectivity and outcome—patients with negative outcomes had strong  $\delta$  functional connections across large parts of the central and parietal areas. In turn, Schorr et al. (2016) indicated that fronto-parietal (in  $\delta$ ,  $\theta$ ,  $\alpha$  and  $\beta$ ) and parietal (in  $\delta$  and  $\theta$ ) coherence could predict recovery from UWS, suggesting that both short and long-distance connections may have a prognostic value and are critical for the preservation of consciousness.

## 8. SUMMARY AND CONCLUSION

The extensive literature on the EEG-based resting state research in patients with DoC shows that the brain activity of patients with UWS is dominated by low  $\delta$  oscillations, both in spectral power and the connectivity domain. The flow of information in higher bands is severely reduced and the variability of brain activity time course is much lower than in patients with MCS. Additionally, the brain functional networks of patients with UWS are characterized by a less complex structure. Patients with MCS are characterized by preserved functional connectivity patterns in higher frequencies—in comparison with patients with UWS, stronger activity in higher frequencies and weaker activity in the  $\delta$  band are observed. It seems that patients with MCS have a more integral fronto-parietal network, and the overall pattern of connectivity, although weaker in higher frequencies, is more comparable to that of healthy controls. Moreover, the brain activity of patients with MCS is more complex and variable over time than that of patients with UWS.

The extensive literature on the EEG-based resting state research in patients with DoC shows quite coherent profiles of EEG functional connectivity in patients with UWS and MCS. However, some contrary observations are presented. This may be

due to the different data analysis approaches used in this field, the use of quite small and highly heterogeneous experimental groups, and the low quality of recorded data. It seems that the greatest variability of the results obtained is related to functional network analysis. Generally, methods of functional connectivity can be classified into two groups—model-free approaches (e.g., coherence, phase lag index) or model-based approaches dependent on the autoregressive model (Granger family indices). They all rely on different internal parameters and are sensitive to various characteristics of measured signals, and none is optimal for all types of signals with different signal-to-noise ratios, or different types of artifacts (Wang et al., 2014). Some research groups have also implemented customized algorithms that are difficult to interpret and compare with more standardized methods. The implementation itself of each method also needs to be considered, because different mathematical or implementation nuances may yield different results.

The EEG signal registered in patients with DoC is very often contaminated by different artifacts to a much higher degree than is the case in healthy subjects. Despite efforts to clean these data manually and the use of automated toolboxes, the quality of the data still seems questionable. It appears that muscle activity in particular, which is very common and strong in these patients, can influence the results obtained and their discrepancies between particular studies in higher frequency ranges.

Moreover, a large number of measures related to functional brain networks have been reported to be important for consciousness mechanisms and require further examination,

e.g., cross-frequency coupling (Naro et al., 2021), time-course characteristics of functional networks (Bai et al., 2021), or neural monitoring of visceral inputs (Candia-Rivera et al., 2021). Notably, the reviewed works are based on indicators (parameters) obtained from the EEG signal. However, new machine learning approaches are applied directly to the raw EEG signal (after filtering and artifacts rejection), showing successfully the most important features related to mechanisms of consciousness (Lee et al., 2022). Finally, another interesting avenue for further research is a combination of different techniques of neuroimaging for the examination of preserved connectivity in patients with DoC in a more complex manner, i.e., taking into account the presence of brain lesions or analysis of the patterns of functional networks obtained by fMRI.

## AUTHOR CONTRIBUTIONS

AD-B, MZ, and KJ-S reviewed the articles and wrote the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was partially supported by the Polish National Science Centre (UMO-2018/31/B/ST7/01888) and a Grant from SWPS University of Social Sciences and Humanities SUB/IPsy/04/2021/04.

## REFERENCES

- Babiloni, C., Sarà, M., Vecchio, F., Pistoia, F., Sebastiano, F., Onorati, P., et al. (2009). Cortical sources of resting-state alpha rhythms are abnormal in persistent vegetative state patients. *Clin. Neurophysiol.* 120, 719–729. doi: 10.1016/j.clinph.2009.02.157
- Bai, Y., Lin, Y., and Ziemann, U. (2021). Managing disorders of consciousness: the role of electroencephalography. *J. Neurol.* 268, 4033–4065. doi: 10.1007/s00415-020-10095-z
- Bai, Y., Xia, X., and Li, X. (2017). A review of resting-state electroencephalography analysis in disorders of consciousness. *Front. Neurol.* 8, 471. doi: 10.3389/fneur.2017.00471
- Bandt, C., and Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* 88, 174102. doi: 10.1103/PhysRevLett.88.174102
- Banks, M. I., Krause, B. M., Endemann, C. M., Campbell, D. I., Kovach, C. K., Dyken, M. E., et al. (2020). Cortical functional connectivity indexes arousal state during sleep and anesthesia. *Neuroimage* 211, 116627. doi: 10.1016/j.neuroimage.2020.116627
- Bareham, C. A., Allanson, J., Roberts, N., Hutchinson, P. J., Pickard, J. D., Menon, D. K., et al. (2018). Longitudinal bedside assessments of brain networks in disorders of consciousness: case reports from the field. *Front. Neurol.* 9, 676. doi: 10.3389/fneur.2018.00676
- Bareham, C. A., Roberts, N., Allanson, J., Hutchinson, P. J., Pickard, J. D., Menon, D. K., et al. (2020). Bedside eeg predicts longitudinal behavioural changes in disorders of consciousness. *Neuroimage Clin.* 28, 102372. doi: 10.1016/j.nicl.2020.102372
- Biswal, B., Zerrin Yetkin, F., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409
- Blumenfeld, H., McNally, K. A., Vanderhill, S. D., Paige, A. L., Chung, R., Davis, K., et al. (2004). Positive and negative network correlations in temporal lobe epilepsy. *Cereb. Cortex* 14, 892–902. doi: 10.1093/cercor/bhh048
- Boly, M., Phillips, C., Tshibanda, L., Vanhaudenhuyse, A., Schabus, M., Dang-Vu, T. T., et al. (2008). Intrinsic brain activity in altered states of consciousness: how conscious is the default mode of brain function? *Ann. N. Y. Acad. Sci.* 1129, 119–129. doi: 10.1196/annals.1417.015
- Boveroux, P., Vanhaudenhuyse, A., Bruno, M.-A., Noirhomme, Q., Lauwick, S., Luxen, A., et al. (2010). Breakdown of within-and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. *Anesthesiology* 13, 1038–1053. doi: 10.1097/ALN.0b013e3181f697f5
- Brenner, R. P. (2005). The interpretation of the eeg in stupor and coma. *Neurologist* 11, 271–284. doi: 10.1097/01.nrl.0000178756.44055.f6
- Brookes, M. J., Woolrich, M., Luckhoo, H., Price, D., Hale, J. R., Stephenson, M. C., et al. (2011). Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16783–16788. doi: 10.1073/pnas.1112685108
- Bruce, E. N., Bruce, M. C., and Venelaganti, S. (2009). Sample entropy tracks changes in eeg power spectrum with sleep state and aging. *J. Clin. Neurophysiol.* 26, 257. doi: 10.1097/WNP.0b013e3181b2f1e3
- Bruno, M.-A., Vanhaudenhuyse, A., Thibaut, A., Moonen, G., and Laureys, S. (2011). From unresponsive wakefulness to minimally conscious plus and functional locked-in syndromes: recent advances in our understanding of disorders of consciousness. *J. Neurol.* 258, 1373–1384. doi: 10.1007/s00415-011-6114-x
- Burioka, N., Miyata, M., Cornelissen, G., Halberg, F., Takeshima, T., Kaplan, D. T., et al. (2005). Approximate entropy in the electroencephalogram during wake and sleep. *Clin. EEG Neurosci.* 36, 21–24. doi: 10.1177/155005940503600106



- Cai, L., Wang, J., Guo, Y., Lu, M., Dong, Y., and Wei, X. (2020). Altered inter-frequency dynamics of brain networks in disorder of consciousness. *J. Neural Eng.* 17, 036006. doi: 10.1088/1741-2552/ab8b2c
- Candia-Rivera, D., Annen, J., Gosseries, O., Martial, C., Thibaut, A., Laureys, S., et al. (2021). Neural responses to heartbeats detect residual signs of consciousness during resting state in postcomatose patients. *J. Neurosci.* 41, 5251–5262. doi: 10.1523/JNEUROSCI.1740-20.2021
- Cao, J., Zhao, Y., Shan, X., Wei, H.-J., Guo, Y., Chen, L., et al. (2021). Brain functional and effective connectivity based on electroencephalography recordings: a review. *Hum. Brain Mapp.* 43, 860–879. doi: 10.1002/hbm.25683
- Cavinato, M., Genna, C., Manganotti, P., Formaggio, E., Storti, S. F., Campostrini, S., et al. (2014). Coherence and consciousness: study of fronto-parietal gamma synchrony in patients with disorders of consciousness. *Brain Topogr.* 28, 570–579. doi: 10.1007/s10548-014-0383-5
- Chennu, S., Annen, J., Wannez, S., Thibaut, A., Chatelle, C., Cassol, H., et al. (2017). Brain networks predict metabolism, diagnosis and prognosis at the bedside in disorders of consciousness. *Brain* 140, 2120–2132. doi: 10.1093/brain/awx163
- Chennu, S., Finoia, P., Kamau, E., Allanson, J., Williams, G. B., Monti, M. M., et al. (2014). Spectral signatures of reorganised brain networks in disorders of consciousness. *PLoS Comput. Biol.* 10, e1003887. doi: 10.1371/journal.pcbi.1003887
- Coleman, M., Menon, D., Fryer, T., and Pickard, J. (2005). Neurometabolic coupling in the vegetative and minimally conscious states: preliminary findings. *J. Neurol. Neurosurg. Psychiatry* 76, 432–434. doi: 10.1136/jnnp.2004.045930
- Corchs, S., Chioma, G., Dondi, R., Gasparini, F., Manzoni, S., Markowska-Kaczmar, U., et al. (2019). Computational methods for resting-state eeg of patients with disorders of consciousness. *Front. Neurosci.* 13, 1323. doi: 10.3389/fnins.2019.01323
- Damoiseaux, J. S., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., et al. (2006). Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13848–13853. doi: 10.1073/pnas.0601417103
- Demertzi, A., Antonopoulos, G., Heine, L., Voss, H. U., Crone, J. S., de Los Angeles, C., et al. (2015). Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. *Brain* 138, 2619–2631. doi: 10.1093/brain/awv169
- Engemann, D. A., Raimondo, F., King, J.-R., Rohaut, B., Louppe, G., Faugeras, F., et al. (2018). Robust eeg-based cross-site and cross-protocol classification of states of consciousness. *Brain* 141, 3179–3192. doi: 10.1093/brain/awy251
- Englot, D. J., Yang, L., Hamid, H., Danielson, N., Bai, X., Marfeo, A., et al. (2010). Impaired consciousness in temporal lobe seizures: role of cortical slow activity. *Brain* 133, 3764–3777. doi: 10.1093/brain/awq316
- Ferenets, R., Lipping, T., Anier, A., Jantti, V., Melto, S., and Hovilehto, S. (2006). Comparison of entropy and complexity measures for the assessment of depth of sedation. *IEEE Trans. Biomed. Eng.* 53, 1067–1077. doi: 10.1109/TBME.2006.873543
- Fingelkurts, A. A., Fingelkurts, A. A., Bagnato, S., Boccagni, C., and Galardi, G. (2011). Life or death: prognostic value of a resting eeg with regards to survival in patients in vegetative and minimally conscious states. *PLoS ONE* 6, e25967. doi: 10.1371/journal.pone.0025967
- Fingelkurts, A. A., Fingelkurts, A. A., Bagnato, S., Boccagni, C., and Galardi, G. (2012). Dmn operational synchrony relates to self-consciousness: evidence from patients in vegetative and minimally conscious states. *Open Neuroimag J.* 6, 55. doi: 10.2174/1874440001206010055
- Forgacs, P. B., Devinsky, O., and Schiff, N. D. (2020). Independent functional outcomes after prolonged coma following cardiac arrest: a mechanistic hypothesis. *Ann. Neurol.* 87, 618–632. doi: 10.1002/ana.25690
- Forgacs, P. B., Frey, H.-P., Velazquez, A., Thompson, S., Brodie, D., Moitra, V., et al. (2017). Dynamic regimes of neocortical activity linked to corticothalamic integrity correlate with outcomes in acute anoxic brain injury after cardiac arrest. *Ann. Clin. Transl. Neurol.* 4, 119–129. doi: 10.1002/acn3.385
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102
- Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008
- Giacino, J. T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., Katz, D. I., et al. (2002). The minimally conscious state definition and diagnostic criteria. *Neurology* 58, 349–353. doi: 10.1212/WNL.58.3.349
- Giacino, J. T., Kalmar, K., and Whyte, J. (2004). The JFK coma recovery scale-revised: measurement characteristics and diagnostic utility. *Arch. Phys. Med. Rehabil.* 85, 2020–2029. doi: 10.1016/j.apmr.2004.02.033
- Gloor, P., Ball, G., and Schaul, N. (1977). Brain lesions that produce delta waves in the eeg. *Neurology* 27, 326–326. doi: 10.1212/WNL.27.4.326
- Gosseries, O., Schnakers, C., Ledoux, D., Vanhaudenhuyse, A., Bruno, M.-A., Demertzi, A., et al. (2011). Automated EEG entropy measurements in coma, vegetative state/unresponsive wakefulness syndrome and minimally conscious state. *Funct. Neurol.* 26, 25.
- Gosseries, O., Zasler, N. D., and Laureys, S. (2014). Recent advances in disorders of consciousness: focus on the diagnosis. *Brain Injury* 28, 1141–1150. doi: 10.3109/02699052.2014.920522
- Gugino, L., Chabot, R., Prichep, L., John, E., Formanek, V., and Aglio, L. (2001). Quantitative eeg changes associated with loss and return of consciousness in healthy adult volunteers anaesthetized with propofol or sevoflurane. *Br. J. Anaesth.* 87, 421–428. doi: 10.1093/bja/87.3.421
- Hagihira, S. (2015). Changes in the electroencephalogram during anaesthesia and their physiological basis. *Br. J. Anaesth.* 115(suppl\_1), i27–i31. doi: 10.1093/bja/aev212
- Höller, Y., Thomschewski, A., Bergmann, J., Kronbichler, M., Crone, J. S., Schmid, E. V., et al. (2014). Connectivity biomarkers can differentiate patients with different levels of consciousness. *Clin. Neurophysiol.* 125, 1545–1555. doi: 10.1016/j.clinph.2013.12.095
- King, J.-R., Sitt, J. D., Faugeras, F., Rohaut, B., El Karoui, L., Cohen, L., et al. (2013). Information sharing in the brain indexes consciousness in noncommunicative patients. *Curr. Biol.* 23, 1914–1919. doi: 10.1016/j.cub.2013.07.075
- Lechinger, J., Bothe, K., Pichler, G., Michitsch, G., Donis, J., Klimesch, W., et al. (2013). Crs-r score in disorders of consciousness is strongly related to spectral eeg at rest. *J. Neurol.* 260, 2348–2356. doi: 10.1007/s00415-013-6982-3
- Lee, H., Golkowski, D., Jordan, D., Berger, S., Ilg, R., Lee, J., et al. (2019a). Relationship of critical dynamics, functional connectivity, and states of consciousness in large-scale human brain networks. *Neuroimage* 188, 228–238. doi: 10.1016/j.neuroimage.2018.12.011
- Lee, M., Baird, B., Gosseries, O., Nieminen, J. O., Boly, M., Postle, B. R., et al. (2019b). Connectivity differences between consciousness and unconsciousness in non-rapid eye movement sleep: a tms-eeg study. *Sci. Rep.* 9, 1–9. doi: 10.1038/s41598-019-41274-2
- Lee, M., Sanders, R. D., Yeom, S.-K., Won, D.-O., Seo, K.-S., Kim, H. J., et al. (2017). Network properties in transitions of consciousness during propofol-induced sedation. *Sci. Rep.* 7, 1–13. doi: 10.1038/s41598-017-15082-5
- Lee, M., Sanz, L. R., Barra, A., Wolff, A., Nieminen, J. O., Boly, M., et al. (2022). Quantifying arousal and awareness in altered states of consciousness using interpretable deep learning. *Nat. Commun.* 13, 1–14. doi: 10.1038/s41467-022-28451-0
- Lehembre, R., Bruno, M.-A., Vanhaudenhuyse, A., Chatelle, C., Cologan, V., Leclercq, Y., et al. (2012a). Resting-state eeg study of comatose patients: a connectivity and frequency analysis to find differences between vegetative and minimally conscious states. *Funct. Neurol.* 27, 41.
- Lehembre, R., Gosseries, O., Lugo, Z., Jedidi, Z., Chatelle, C., Sadzot, B., et al. (2012b). Electrophysiological investigations of brain function in coma, vegetative and minimally conscious patients. *Arch. Ital. Biol.* 150, 122–139. doi: 10.4449/aib.v150i2.1374
- Lei, L., Liu, K., Yang, Y., Doubliez, A., Hu, X., Xu, Y., et al. (2022). Spatio-temporal analysis of eeg features during consciousness recovery in patients with disorders of consciousness. *Clin. Neurophysiol.* 133, 135–144. doi: 10.1016/j.clinph.2021.08.027
- Leon-Carrion, J., Leon-Dominguez, U., Pollonini, L., Wu, M.-H., Frye, R. E., Dominguez-Morales, M. R., et al. (2012). Synchronization between the anterior and posterior cortex determines consciousness level in patients with traumatic brain injury (tbi). *Brain Res.* 1476, 22–30. doi: 10.1016/j.brainres.2012.03.055
- Leon-Carrion, J., Martin-Rodriguez, J., Damas-Lopez, J., y Martin, J. B., and Dominguez-Morales, M. (2008). Brain function in the minimally conscious state: a quantitative neurophysiological study. *Clin. Neurophysiol.* 119, 1506–1514. doi: 10.1016/j.clinph.2008.03.030
- Liang, Z., Wang, Y., Sun, X., Li, D., Voss, L. J., Sleight, J. W., et al. (2015). Eeg entropy measures in anesthesia. *Front. Comput. Neurosci.* 9, 16. doi: 10.3389/fncom.2015.00016
- Lutkenhoff, E. S., Nigri, A., Sebastiano, D. R., Sattin, D., Visani, E., Rosazza, C., et al. (2019). Eeg power spectra and subcortical pathology in chronic disorders of consciousness. *bioRxiv* 695288. doi: 10.1101/695288

- Marinazzo, D., Gosseries, O., Boly, M., Ledoux, D., Rosanova, M., Massimini, M., et al. (2014). Directed information transfer in scalp electroencephalographic recordings: insights on disorders of consciousness. *Clin. EEG Neurosci.* 45, 33–39. doi: 10.1177/1550059413510703
- Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., and Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science* 315, 393–395. doi: 10.1126/science.1131295
- Naro, A., Bramanti, A., Leo, A., Cacciola, A., Manuli, A., Bramanti, P., et al. (2018). Shedding new light on disorders of consciousness diagnosis: the dynamic functional connectivity. *Cortex* 103, 316–328. doi: 10.1016/j.cortex.2018.03.029
- Naro, A., Bramanti, P., Leo, A., Cacciola, A., Bramanti, A., Manuli, A., et al. (2016). Towards a method to differentiate chronic disorder of consciousness patients' awareness: The low-resolution brain electromagnetic tomography analysis. *J. Neurol. Sci.* 368, 178–183. doi: 10.1016/j.jns.2016.07.016
- Naro, A., Maggio, M. G., Leo, A., and Calabro, R. S. (2021). Multiplex and multilayer network eeg analyses: a novel strategy in the differential diagnosis of patients with chronic disorders of consciousness. *Int. J. Neural Syst.* 31, 2050052. doi: 10.1142/S0129065720500525
- O'Donnell, A., Pauli, R., Banellis, L., Sokoliuk, R., Hayton, T., Sturman, S., et al. (2021). The prognostic value of resting-state eeg in acute post-traumatic unresponsive states. *Brain Commun.* 3, fcab017. doi: 10.1093/braincomms/fcab017
- Piarulli, A., Bergamasco, M., Thibaut, A., Cologan, V., Gosseries, O., and Laureys, S. (2016). Eeg ultradian rhythmicity differences in disorders of consciousness during wakefulness. *J. Neurol.* 263, 1746–1760. doi: 10.1007/s00415-016-8196-y
- Pollonini, L., Pophale, S., Situ, N., Wu, M.-H., Frye, R. E., Leon-Carrion, J., et al. (2010). Information communication networks in severe traumatic brain injury. *Brain Topogr.* 23, 221–226. doi: 10.1007/s10548-010-0139-9
- Purdon, P. L., Pierce, E. T., Mukamel, E. A., Prerau, M. J., Walsh, J. L., Wong, K. F. K., et al. (2013). Electroencephalogram signatures of loss and recovery of consciousness from propofol. *Proc. Natl. Acad. Sci. U.S.A.* 110, E1142–E1151. doi: 10.1073/pnas.1221180110
- Raichle, M. E. (2015). The brain's default mode network. *Annu. Rev. Neurosci.* 38, 433–447. doi: 10.1146/annurev-neuro-071013-014030
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682. doi: 10.1073/pnas.98.2.676
- Rizkallah, J., Annen, J., Modolo, J., Gosseries, O., Benquet, P., Mortaheb, S., et al. (2019). Decreased integration of EEG source-space networks in disorders of consciousness. *NeuroImage* 23, 101841. doi: 10.1016/j.nicl.2019.101841
- Sarà, M., and Pistoia, F. (2010). Complexity loss in physiological time series of patients in a vegetative state. *Nonlinear Dyn. Psychol. Life Sci.* 14, 1.
- Sarà, M., Pistoia, F., Pasqualetti, P., Sebastiano, F., Onorati, P., and Rossini, P. M. (2011). Functional isolation within the cerebral cortex in the vegetative state: a nonlinear method to predict clinical outcomes. *Neurorehabil. Neural Repair.* 25, 35–42. doi: 10.1177/1545968310378508
- Schiff, N. D. (2016). "Mesocircuit mechanisms underlying recovery of consciousness following severe brain injuries: Model and predictions," in *Brain Function and Responsiveness in Disorders of Consciousness*, eds M. Monti, and W. Sannita (Cham: Springer). doi: 10.1007/978-3-319-21425-2\_15
- Schiff, N. D., Nauvel, T., and Victor, J. D. (2014). Large-scale brain dynamics in disorders of consciousness. *Curr. Opin. Neurobiol.* 25, 7–14. doi: 10.1016/j.conb.2013.10.007
- Schnakers, C., Ledoux, D., Majerus, S., Damas, P., Damas, F., Lambermont, B., et al. (2008). Diagnostic and prognostic use of bispectral index in coma, vegetative state and related disorders. *Brain Injury* 22, 926–931. doi: 10.1080/02699050802530565
- Schorr, B., Schlee, W., Arndt, M., and Bender, A. (2016). Coherence in resting-state eeg as a predictor for the recovery from unresponsive wakefulness syndrome. *J. Neurol.* 263, 937–953. doi: 10.1007/s00415-016-8084-5
- Sebastiano, D. R., Varotto, G., Sattin, D., and Franceschetti, S. (2021). Eeg assessment in patients with disorders of consciousness: Aims, advantages, limits, and pitfalls. *Front. Neurol.* 12, 649849. doi: 10.3389/fneur.2021.649849
- Shen, H. H. (2015). Core concept: resting-state connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14115–14116. doi: 10.1073/pnas.1518785112
- Shin, H. W., Kim, H. J., Jang, Y. K., You, H. S., Huh, H., Choi, Y. J., et al. (2020). Monitoring of anesthetic depth and eeg band power using phase lag entropy during propofol anesthesia. *BMC Anesthesiol.* 20, 1–10. doi: 10.1186/s12871-020-00964-5
- Sitt, J. D., King, J.-R., El Karoui, I., Rohaut, B., Faugeras, F., Gramfort, A., et al. (2014). Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* 137, 2258–2270. doi: 10.1093/brain/awu141
- Stefan, S., Schorr, B., Lopez-Rolon, A., Kolassa, I.-T., Shock, J. P., Rosenfelder, M., et al. (2018). Consciousness indexing and outcome prediction with resting-state eeg in severe disorders of consciousness. *Brain Topogr.* 31, 848–862. doi: 10.1007/s10548-018-0643-x
- Steriade, M., Nunez, A., and Amzica, F. (1993). A novel slow (<1 Hz) oscillation of neocortical neurons *in vivo*: depolarizing and hyperpolarizing components. *J. Neurosci.* 13, 3252–3265. doi: 10.1523/JNEUROSCI.13-08-03252.1993
- Thibaut, A., Panda, R., Annen, J., Sanz, L. R., Naccache, L., Martial, C., et al. (2021). Preservation of brain activity in unresponsive patients identifies mcs star. *Ann. Neurol.* 90, 89–100. doi: 10.1002/ana.26095
- Timofeev, I., Grenier, F., and Steriade, M. (2001). Disfacilitation and active inhibition in the neocortex during the natural sleep-wake cycle: an intracellular study. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1924–1929. doi: 10.1073/pnas.98.4.1924
- van den Brink, R., Nieuwenhuis, S., van Bostel, G., van Luijckelaar, G., Eilander, H., and Wijnen, V. (2018). Task-free spectral eeg dynamics track and predict patient recovery from severe acquired brain injury. *Neuroimage Clin.* 17, 43–52. doi: 10.1016/j.nicl.2017.10.003
- van Putten, M. J., and Hofmeijer, J. (2016). Eeg monitoring in cerebral ischemia: basic concepts and clinical applications. *J. Clin. Neurophysiol.* 33, 203–210. doi: 10.1097/WNP.0000000000000272
- Vanhudenhuysse, A., Noirhomme, Q., Tshibanda, L. J.-F., Bruno, M.-A., Boveroux, P., Schnakers, C., et al. (2009). Default network connectivity reflects the level of consciousness in non-communicative brain-damaged patients. *Brain* 133, 161–171. doi: 10.1093/brain/awp313
- Varotto, G., Fazio, P., Sebastiano, D. R., Duran, D., D'Incerti, L., Parati, E., et al. (2014). Altered resting state effective connectivity in long-standing vegetative state patients: an eeg study. *Clin. Neurophysiol.* 125, 63–68. doi: 10.1016/j.clinph.2013.06.016
- Wang, H. E., Bénar, C. G., Quilichini, P. P., Friston, K. J., Jirsa, V. K., and Bernard, C. (2014). A systematic framework for functional connectivity measures. *Front. Neurosci.* 8, 405. doi: 10.3389/fnins.2014.00405
- Whitman, J. C., Ward, L. M., and Woodward, T. (2013). Patterns of cortical oscillations organize neural activity into whole-brain functional networks evident in the fmri bold signal. *Front. Hum. Neurosci.* 7, 80. doi: 10.3389/fnhum.2013.00080
- Wu, D., Cai, G., Yuan, Y., Liu, L., Li, G.-Q., Song, W.-Q., et al. (2011a). Application of nonlinear dynamics analysis in assessing unconsciousness: a preliminary study. *Clin. Neurophysiol.* 122, 490–498. doi: 10.1016/j.clinph.2010.05.036
- Wu, D., Cai, G., Zorowitz, R. D., Yuan, Y., Wang, J., and Song, W.-Q. (2011b). Measuring interconnection of the residual cortical functional islands in persistent vegetative state and minimal conscious state with eeg nonlinear analysis. *Clin. Neurophysiol.* 122, 1956–1966. doi: 10.1016/j.clinph.2011.03.018
- Wutzl, B., Golaszewski, S. M., Leibnitz, K., Langthaler, P. B., Kunz, A. B., Leis, S., et al. (2021). Narrative review: quantitative eeg in disorders of consciousness. *Brain Sci.* 11, 697. doi: 10.3390/brainsci11060697

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Duszyk-Bogorodzka, Zieleniewska and Jankowiak-Siuda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [frontiersin.org/about/contact](http://frontiersin.org/about/contact)



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership