



# The Role of the Baldwin Effect in the Evolution of Human Musicality

Piotr Podlipniak\*

*Institute of Musicology, Adam Mickiewicz University in Poznań, Poznań, Poland*

## OPEN ACCESS

### Edited by:

Aleksey Nikolsky,  
Braavo! Enterprises, United States

### Reviewed by:

Nicholas Bannan,  
University of Western Australia,  
Australia

Edward Hagen,  
Washington State University  
Vancouver, United States

### \*Correspondence:

Piotr Podlipniak  
podlip@poczta.onet.pl

### Specialty section:

This article was submitted to  
Auditory Cognitive Neuroscience,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 24 May 2017

**Accepted:** 19 September 2017

**Published:** 06 October 2017

### Citation:

Podlipniak P (2017) The Role of the  
Baldwin Effect in the Evolution of  
Human Musicality.  
*Front. Neurosci.* 11:542.  
doi: 10.3389/fnins.2017.00542

From the biological perspective human musicality is the term referred to as a set of abilities which enable the recognition and production of music. Since music is a complex phenomenon which consists of features that represent different stages of the evolution of human auditory abilities, the question concerning the evolutionary origin of music must focus mainly on music specific properties and their possible biological function or functions. What usually differentiates music from other forms of human sound expressions is a syntactically organized structure based on pitch classes and rhythmic units measured in reference to musical pulse. This structure is an auditory (not acoustical) phenomenon, meaning that it is a human-specific interpretation of sounds achieved thanks to certain characteristics of the nervous system. There is historical and cross-cultural diversity of this structure which indicates that learning is an important part of the development of human musicality. However, the fact that there is no culture without music, the syntax of which is implicitly learned and easily recognizable, suggests that human musicality may be an adaptive phenomenon. If the use of syntactically organized structure as a communicative phenomenon were adaptive it would be only in circumstances in which this structure is recognizable by more than one individual. Therefore, there is a problem to explain the adaptive value of an ability to recognize a syntactically organized structure that appeared accidentally as the result of mutation or recombination in an environment without a syntactically organized structure. The possible solution could be explained by the Baldwin effect in which a culturally invented trait is transformed into an instinctive trait by the means of natural selection. It is proposed that in the beginning musical structure was invented and learned thanks to neural plasticity. Because structurally organized music appeared adaptive (phenotypic adaptation) e.g., as a tool of social consolidation, our predecessors started to spend a lot of time and energy on music. In such circumstances, accidentally one individual was born with the genetically controlled development of new neural circuitry which allowed him or her to learn music faster and with less energy use.

**Keywords:** Baldwin effect, human musicality, cortico-subcortical loops, pitch structure, musical rhythm

## INTRODUCTION

Human musicality can be understood as a set of abilities which enable people to recognize and produce music (Fitch, 2015). Although the worldwide diversity of music reveals the cultural flexibility of *Homo sapiens* in musical behavior, the fact that people in all known cultures sing (Nettl, 2000) and can recognize music without explicit learning (Tillmann, 2005) suggests that

musicality is a part of human biological endowment (Blacking, 1973; Fitch, 2006, 2015). In fact, a variety of theories have been proposed which try to explain the possible adaptive value of music (Cross and Morley, 2008). Many of the possible biological values of music have been put forward such as increasing sexual attractiveness (Darwin, 1871; Miller G. F., 2000), facilitating mother-infant bonds (Dissanayake, 2008), enhancing group consolidation (Roederer, 1984; Storr, 1992; Harvey, 2017), reducing cognitive dissonance (Perlovsky, 2010, 2012), and alerting outsiders about a group cohesiveness (Hagen and Bryant, 2003; Hagen and Hammerstein, 2009), to name only the most popular ideas. However, music is a complex communicative phenomenon (Zimmermann et al., 2013) which is composed of many features. Some of these features are shared with other communicative phenomena, which raises the question about music specificity and in consequence, about the biological character of human musicality. For example, the manipulation of sound intensity, stress and tempo is an important part of all human songs but also of speech. Moreover, a similar use of these features is observed among the vocal expressions of many mammalian species (Zimmermann et al., 2013) including our closest relative—the chimpanzee (*Pan troglodytes*) (Slocombe et al., 2009). The patterns of continuously modulated sounds used as a tool of expression and induction of emotions in other individuals are called “expressive dynamics” (Merker, 2003) or “affective prosody” (Zimmermann et al., 2013) and seem to be evolutionarily more ancient than any species of hominins. Of course, music is also usually composed of such elements which are not shared with other forms of human sound expressions. The interpretation of musical stimuli in terms of pitch classes and temporal isochrony are at least two musical elements which are absent in speech (Fitch, 2013a) and other human vocalizations. Although pitch and the duration of vowels can serve as discreet phonological units in tonal and time sensitive languages respectively (Remijsen and Gilley, 2008; Wong et al., 2012; Remijsen, 2014) both pitch and vowel length in speech lack hierarchical ordering based on mental reference points. Such a hierarchy is evident in music where pitch classes are organized in reference to pitch center (Podlipniak, 2016) and rhythm measures in reference to musical pulse (London, 2012). By taking into account that both of these features are not present in the vocal expressions of chimpanzees it is reasonable to assume that they were also absent in the vocal repertoire of the common ancestor of chimpanzees and humans and thus are relatively evolutionarily young innovations.

The coexistence of different evolutionarily aged features in our musical expression encourages us to think of human musicality in an analogous way, as in the faculty of language (Hauser et al., 2002) namely in terms of musicality in the broad and narrow sense. While musicality in the broad sense can encompass a set of abilities which develop at an early stage of human ontogenesis, and which allow the identification of pitch contour, changes in tempo, and dynamics as parts of the affective prosody (Zimmermann et al., 2013), only abilities which constitute musicality in the narrow sense enable the recognition of musical structure. From this perspective the actual question about the origin of music is related to the origin of these abilities which

constitute musicality in the narrow sense. What makes music a phenomenon distinguishable from other experiences of sounds is the interpretation of sounds by the human nervous system not just as unrelated sound events but as a psychologically unified musical structures. Although it has been proposed that musical structure (e.g., melody) was discovered by early humans due to the resemblance of musical sounds to the acoustic characteristics of human vocalizations (Purves, 2017), the real problem is why the nervous system of early humans started to interpret certain acoustic parameters as discrete pitch and rhythm units organized into syntactically based sequences. After all, despite the fact that a sound stimulus is usually a very complex phenomenon composed of an enormous number of spectral and temporal cues, different species recognize their species-specific vocalizations using only the subsets of features which are distinctive solely to their song cognition (Bregman et al., 2016; Shannon, 2016). In other words, every species is sensitive to specific acoustic cues due to the proclivities of its own nervous system. In addition, speech and music, two complex hierarchical cognitive systems of *H. sapiens* (Fitch, 2014), operate with a restricted number of different distinctive features (Patel, 2008). Therefore, there must be something which predisposes humans to focus their attention on particular acoustic features whilst ignoring others. What is more, the vast majority of musical structures, especially songs of tribal communities (Blacking, 1973), are organized according to certain syntactic rules (Koelsch, 2013 but see London, 2011), which suggests that musical syntax is a natural trait of human vocalization. Apart from this, musical syntax based on pitch classes and rhythmic units measured in reference to musical pulse is a music-specific feature (Fitch, 2013a). This raises the question about the evolution of the abilities which allow the recognition of musical syntax, which seems to be the core of human musicality.

## MUSICAL SYNTAX AS A MUSIC SPECIFIC FEATURE

The human ability to organize and interpret stimuli as syntactically complex sequences is often regarded as a milestone in the evolution of human cognition (Hauser et al., 2002; Fitch, 2014). Even though syntax, understood in the broad sense as rules combining discrete elements into sequences (Patel, 2008), can be attributed to certain animal songs (Okanoya, 2013; Suzuki et al., 2016), both musical and language syntaxes seem to be exceptional in terms of their complexity and function (Fitch and Jarvis, 2013). Musical syntax however, is often thought of as a derivative of language syntax (Patel, 2008 but see Jackendoff and Lerdahl, 2006), or as a product of domain-general structural computation (Fitch, 2014; Van de Cavey and Hartsuiker, 2016) rather than a functionally separate phenomenon. In fact, there are good reasons to assume the general role of structural computation in the processing of language and musical syntaxes. There are many neuroimaging studies that show an overlap in the activation of cortical structures (located in the inferior frontal gyrus) during the processing of musical and language syntaxes (Patel et al., 1998; Maess et al., 2001). Moreover, there is also research which reveals that the same structures (e.g., Broca's area)

play a certain role in the performing of other tasks involving sequential ordering (Tettamanti and Weniger, 2006; Friedrich and Friederici, 2009; Higuchi et al., 2009; Wakita, 2014). Also, the cognitive deficits observed after lesions in the Broca's area include not only deficits in production and recognition of language syntax but also action execution and observation as well as musical syntax processing (Fadiga et al., 2009).

However, from a behavioral point of view, language and music are functionally different phenomena. The former communicates referential meaning by the means of intersubjectively understandable concepts (Bickerton, 2010), whereas the latter exchanges information which is at least more ambiguous in terms of its semantic content (Cross, 2005). Since natural selection operates on phenotypes, we can assume that syntactical language evolved because it gave an advantage to our ancestors over those individuals who could not grasp the syntactical rules present in the language of our ancestors. This means that in the case of cognitive abilities natural selection acts directly upon the behavioral effects of the activity of brain structures. Therefore, when searching for the evolutionary origin of a particular cognitive trait it seems more reasonable to look at it from Tinbergen's perspective (Tinbergen, 1963), namely by considering the possible ultimate function of the brain products (e.g., syntactical language), rather than the proximal function which a particular brain structure fulfills in the processing of a certain task or tasks. From this point of view, the circuitries which perform whole task such as the processing of syntactical language should be assumed as the result of natural selection rather than isolated brain structures which take part in the operations of these circuitries (e.g., Broca's area). Since evolution usually optimizes organisms by the means of adjusting the existing traits to new functions (Jacob, 1977), it is not surprising that the same brain structures are often parts of functionally different circuits. Of course, the human brain is characterized by having huge plasticity, thus the general ability to attribute "tree structures" (Fitch, 2014) (complex syntax) to stimuli, is not restricted solely to language and music. It is well known that people are able to implicitly learn the artificial syntaxes of different stimuli (Reber et al., 1999). Nonetheless, in comparison to artificial grammars, both language and music seem to be exceptional in respect to the rate and easiness of the implicit learning of their syntactic rules by children (Jablonka and Lamb, 2005; Tillmann, 2005). The fact that there is no culture without language and music additionally strengthens the point of view that musicality, similar to language abilities, is a natural part of human behavior rather than being a very old cultural invention similar to writing or playing chess.

## Musical Syntax vs. Language Syntax

There are many similarities between language and musical syntaxes. Both are compositional and hierarchical (Merker, 2002) and both generate long-distance dependencies (Bickerton, 2009; Woolhouse et al., 2016). The default mode of language—speech—is like music in the auditory domain, and it has been observed that the processing of music and speech syntactic tasks activates the peri-Sylvian network which connects the inferior frontal gyrus with sensory cortices located in the temporal lobes (Fitch, 2014). However, musical and language syntaxes are also quite

different in many respects. First of all, music is composed of different units to those of language. The basic units of speech are phonemes which are experienced in our internal world as unique qualities hardly comparable with our experience of pitch class. Their discrimination is also based on different spectral cues. Pitch classes are recognized by the fundamental frequency of harmonic sound ( $F_0$ ) (Stainsby and Cross, 2008) whereas phonemes mainly by the spectral and temporal shape of sound (Xu et al., 2005). Although the processing of certain characteristics of spectral shape is important for the discrimination of timbre in music (McAdams and Giordano, 2008), the role of timbre in musical syntax is at least doubtful. Admittedly, timbre can play an important role in the structural organization of music as it is observed in certain musical styles such as in the deep throat singing of Tuva and Mongolia (Levin and Süzükei, 2006), the music of the Jew's harp (Fox, 1988), and tabla music in India (Patel, 2008). There are also musical cultures (e.g., Yakut culture in Siberia) in which the structure-forming function of pitch is extremely reduced whereas timbre seems to be a dominant factor which structures the sound order. Nevertheless, in all these cases timbral structure is hardly comparable to pitch and rhythm structure mainly because of the multidimensional perceptive character of timbre (Lerdahl, 1987). Also, our mental images of timbre in music and phonemes in speech differ, although both are based on the interpretation of the spectral shape of sound. Therefore, even though language grammar (Lerdahl and Jackendoff, 1983), prosodic structure (Heffner and Slevc, 2015) as well as phonotactics and morphonotactics are to some extent comparable to musical syntax (Lerdahl, 2013), the perceptual salience of these phenomena is disparate. The crucial difference between language and musical syntax however is related to the function which these syntaxes fulfill in music and language being different communicative phenomena. In language, syntax is mapped into conceptual meaning (propositional semantics; Hilliard and White, 2009) which allows a concatenation of meaning i.e., putting together two or more units and thereby creating a new meaning in comparison to the meanings of those units alone (Bickerton, 2009). But this process of mapping does not seem to be unidirectional as semantics can influence syntactic rules as in the case of some verbs in which the meaning determines grammatical patterns (Dor and Jablonka, 2000). Thus, the function of language syntax is strictly related to communication of specific conceptual meanings (Dor and Jablonka, 2001). In contrast, the function of syntax in music has nothing to do with such a complex interdependence between syntax and concepts observed in language. Although there is an endless dispute over the existence and character of musical semantics (Patel, 2008; Koelsch, 2013; Reybrouck, 2013; Seifert et al., 2013) even if one admits that music can communicate referential meaning, both the type of this meaning (Dor and Jablonka, 2001) and its relation to musical syntax (Lerdahl, 2013) are definitely different than in language.

## Musical Rhythm and Pitch as the Basis for Musical Syntax

There is currently no agreement about what musical syntax actually is (Patel, 2003; London, 2011; Koelsch, 2013; Lerdahl, 2013; Asano and Boeckx, 2015; Heffner and Slevc, 2015). The

majority of research on musical syntax has been conducted on Western artistic music—especially on the functional relations between chords (Koelsch, 2013). But the case of Western artistic music seems to be an inadequate example of human musical expressions as the manifestation of *H. sapiens* musicality (Jackendoff and Lerdahl, 2006). After all, functional harmony is an exception within the wide variety of world music. Although music based on functional harmony has become more and more widespread in the last century, its history is very young in comparison to the ancient history of music without functional harmony. Nevertheless, the experience of even simple melody without any accompaniment necessitates the recognition of syntactic relations. These relations are hierarchical, meaning that the sequence of sounds is interpreted by the nervous system as being composed of units (sounds perceived as belonging to a particular pitch class and having a particular rhythmic measure) which possess different prominence. This prominence attributed to sounds as elements of metrical (London, 2012; Fitch, 2013b) and pitch (Lerdahl and Jackendoff, 1983; Huron, 2006) patterns is a mental construct as with the prominence of grammatical categories in language. However, in the case of music the metrical or tonal prominence is rather “felt” than “conceptually known” as in language cognition. This preconceptual character of musical hierarchy is especially evident when music is experienced by non-musicians who, in contrast to professional musicians (Burns and Ward, 1978), do not always recognize pitch intervals by the means of categorical perception (Smith et al., 1994). However, even musically trained listeners experience musical hierarchy in such a preconceptual way despite the fact that they are additionally able to recognize musical structure in terms of precise mental categories. What seems to be a source of the preconceptual experience of musical hierarchy is somehow related to motor and emotional brain processes. The recognition of meter in music can be understood as a kind of entrainment which exists in the connection between our auditory and sensorimotor systems (London, 2012). It has even been proposed that human metrical interpretation of music is based on hidden sensorimotor activity (Repp, 2007). This sensorimotor activity during listening to music leads in turns to emotional reactions (Sievers et al., 2013). Also, the recognition of pitch hierarchy causes measurable emotional reactions (Steinbeis et al., 2006; Koelsch et al., 2008; Mikutta et al., 2015), and perception of pitch changes (often described as “leaps” and “steps”) can lead to sensorimotor interpretation (Nikolsky, 2015). Since emotion is an evolutionarily old motivational mechanism (Toates, 1988), the function of which is to assess a potential danger or attractiveness of perceived stimuli (Panksepp, 1998), the tight connection between musical structure and emotions suggests the biological importance of this human specific interpretation of sound in terms of pitch and metric hierarchy.

## Musical Syntax and Emotions

During the processing of musical syntax, people experience a set of subtle emotional reactions (emotional qualia) dependent on the position of a note in each syntactic context (Huron, 2006; Margulis, 2014). This observation is supported by the fact that the bilateral Amygdalae and the orbitofrontal cortex are differently

activated during listening to music depending on the syntactic relations of musical sequences (Mikutta et al., 2015). In spite of the coexistence of an affective experience during listening to musical syntax, the emotional reaction to musical syntax is often suggested as the mere result of cognitive recognition of syntactic structure (Koelsch et al., 2008) rather than an integral part of this recognition. For example, Huron (Huron, 2006) has proposed that the association of emotions with particular pitch classes is the result of the so called “misattribution effect.” In the case of music it is the misattribution of limbic reward or punishment (caused by fulfilling or not fulfilling predictions about which pitch class will be next) to the pitch classes themselves depending on the general mechanism of prediction. However, in the original misattribution effect (Dutton and Aron, 1974) the feeling experienced in response to a stimulus (e.g., the instability of a bridge perceived by the vestibular and visual systems) is misattributed to another stimulus (e.g., a woman perceived by sight). But in Huron’s example there is only one stimulus—sound. Because prediction is the ultimate function of the nervous system (Llinás, 2001) one can assume that every perception is based on prediction. The limbic reward (or punishment) in response to well predicted (or falsely predicted) stimuli is an evolutionarily old mechanism of the assessment of stimuli (Panksepp, 1998), inseparable from every perception. In other words, both the emotional reaction to a particular stimulus and the prediction of stimuli are parts of cognition. Therefore, the prediction processes of the nervous system cannot be treated similarly to external stimuli as a source of emotions. The actual source of emotion is an external stimulus and a prediction process is one of the mechanisms, the function of which is to deliver information about the external world that is assessed by emotions. Moreover, if the emotional effect was solely the result of prediction then the emotional reactions to equally predicted stimuli should be the same, independent of whether they are parts of e.g., musical syntax, speech phonotactics or the sequence of timbres in music (Gorzelańczyk et al., 2017). Yet the emotional experience of musical syntactic relations seems qualitatively different from the experience of phonotactics and other syntactically organized sequences. This difference is evident if we compare singing with speech. Although both of these vocal expressions are composed of syntactically organized sounds, the variations of pitch in time occur much slower in singing than in speech (Zatorre and Baum, 2012) and the emotional impact of singing on listeners seems to be greater in comparison with speech from our childhood (Nakata and Trehub, 2004) and lasting throughout our lives.

Taking into account the behavioral specificity of singing, both prediction and emotional reactions to successive sounds are in this case rather the integral parts of a mental tool dedicated to processing musical structure. From this perspective, different subtle emotional reactions to a variety of possible syntactic relations are the elements of a functionally specific form of communication, similar to how semantics is strictly connected with grammar in natural language. In other words, music can be understood as a mapping system in which syntactical relations are mapped into preconceptual emotional cues. For the majority of the human population, the recognition of musical syntax is a solely preconceptual experience. Of course, the

conceptual level of musical syntax recognition is achievable by professional musicians. However, in contrast to implicitly learned tacit syntactic knowledge by non-musicians, this additional level of competence necessitates strenuous explicit learning. Changes in the neural architecture of musicians in response to environmental influences (explicit learning) could perhaps represent a kind of phenotypic adaptation within the cognitive domain which is possible due to the brain's plasticity (Moreno and Bidelman, 2014; Strait and Kraus, 2014). In fact, it has been observed that music performance affects the transcription of genes that are related to dopaminergic neurotransmission, motor behavior, neuronal plasticity, and neurocognitive functions such as learning and memory (Kanduri et al., 2015) which can be responsible for the observed differences between musicians and non-musicians. Therefore, the fact that musical syntax can be recognized by musically trained individuals at a conceptual level shows how much cultural influence can extend cognitive skills rather than saying something about its primordial nature. Since for average humans the syntactic relations in music, both metrical prominence (especially evident when they dance or tap), and pitch prominence (when they sing), are somehow felt but are difficult to express conceptually, it is reasonable to assume that motor, emotional and cognitive processing are an integral part of the ability to recognize and produce syntactically organized music.

## MUSIC AND THE BALDWIN EFFECT

The main problem concerning the evolution of the abilities to use syntax as a part of any communicative system is related to the question about the cause of “syntax genes” proliferation (Dor and Jablonka, 2001). If musical or language syntax is adaptive due to extending communicative capabilities, then this means that it must be used by at least two individuals. After all, as long as one individual cannot use syntax in communication with another individual, the use of syntax by the latter is useless. Even with the use of music in the form of self-communication, as in the case of “personal song” in the musical cultures of Siberia, Far East, and Amerindian tribes (Nikolsky, 2015), the development of the abilities to organize music syntactically necessitates learning from another individual. However, the appearance of a new genetic trait in the population is usually a result of accidental mutation or recombination. Because the probability of the coincidence of identical mutations of the same allele in the same generation is very low the appearance of a genetically based predisposition to organize vocal expression syntactically seems puzzling. In other words, all advantages of syntax are useless in a population in which only one individual is able to produce and recognize syntactically organized sequences. A possible solution to this problem can be the Baldwin effect (Baldwin, 1896a; Simpson, 1953).

### The Baldwin Effect

The Baldwin effect is an evolutionary mechanism which transforms a culturally invented and acquired trait into an instinctive trait by the means of natural selection (Baldwin, 1896a; Simpson, 1953; Hall, 2001). Although this mechanism was

independently proposed at the end of the Nineteenth Century by at least three different people (Baldwin, 1896a,b; Morgan, 1896; Osborn, 1896) it was forgotten with some exceptions (Simpson, 1953; Waddington, 1953a,b), for more than half of the next century. Only in the last few decades of the twentieth century has the Baldwinian idea started to inspire scientists and philosophers and has regained popularity (Godfrey-Smith, 2003). The core of this concept is very simple. Some animals due to their cognitive flexibility learn new adaptive behaviors in response to environmental changes. If a particular behavior is adaptive, lasts many generations, and its learning is strenuous and time-consuming, sooner or later a genetically based predisposition appears and starts to be favored by natural selection (Dor and Jablonka, 2000, 2010; Godfrey-Smith, 2003; Jablonka and Lamb, 2005). Therefore, the Baldwin effect is a combination of learning and the genetic assimilation of a learned trait (Dor and Jablonka, 2000, 2001; Godfrey-Smith, 2003). The process of Baldwinian evolution occurs in three stages: (i) the appearance of a new environmental challenge, (ii) the invention of a new behavior as a response to the new environmental challenge and its proliferation by the means of learning—at this stage natural selection favors cognitive plasticity, (iii) the appearance of a new genetically based predisposition (canalization, Jablonka and Lamb, 1995; Dor and Jablonka, 2010)—at this stage natural selection favors less flexible individuals but faster at exhibiting a particular adaptive behavior (Godfrey-Smith, 2003).

Since an important factor in the Baldwinian evolution of human behavior is the social character of our species, the aforesaid environmental challenge can be a part of the socio-cultural products of hominins. This specific socio-cultural environment often described as a “cultural niche” (Godfrey-Smith, 2003) has been proposed as a crucial element in the evolution of natural language (Bickerton, 2010; Deacon, 2010). In some regards, the evolution of language can represent a niche construction (Deacon, 1997, 2003)—a kind of niche extension or elaboration. The importance of this “cultural niche” seems evident when taking into account that at least during the last 2 million years hominins have become more and more socially complex animals in comparison to other primates (Dunbar, 2014). Living in a complex social group definitely causes new challenges which influence both natural selection and niche construction. In the process of niche construction hominin brains and hominin culture can be considered as a specific environment in which language evolved (Deacon, 2003). From this perspective the Baldwinian process is a part of the gene-culture coevolution (Lumsden and Wilson, 1982; Richerson et al., 2010; Gintis, 2011). Since speech (the “default mode” of language) is similar to music in many respects (both are transmitted by the acoustic domain, both are syntactical systems composed of discrete elements etc.), and language is assumed to be a crucial factor in the cultural evolution of *H. sapiens*, it seems reasonable to assume that the evolution of human musicality is somehow related to the evolution of language. However, language is a very elaborate signal related to the exchange of conceptual meaning. The presence of non-symbolic and non-conceptual culture in many other species (Cantor and Whitehead, 2013; van de Waal et al., 2013; Fehér et al., 2016) indicates that the beginning of

cultural niche construction can be based on the exchange of preconceptual meaning. Therefore, music as an example of a communication system operating on preconceptual meaning is a good candidate to be a part of more ancient communicative tool other than language and so the proposed Baldwinian scenarios of language origin must differ from the possible Baldwinian processes that led to the emergence of music.

## The Baldwinian Evolution of Music

Human musicality seems to be a very good example of the potential effects of Baldwinian evolution (Podlipniak, 2015). Huron has suggested that apart from certain reflexes, the human auditory experience is mainly influenced by learning (Huron, 2006). The predominance of learning in shaping human auditory cognition implies, according to Huron, that the auditory environment of hominins must have been very semiotically unstable. Following Huron's reasoning, such a semiotic instability led to the great variety of music found around the world. However, music as a product of human musicality is characterized not only by culture-specific features but also by universals (Nettl, 2000; Bispham, 2009; Brown and Jordania, 2011; Savage et al., 2015) which suggests that apart from the cultural (environmental) influence, the music-specific genetic constraints also shape the musical mind of every human. This means that on the one hand, musicality develops spontaneously and effortlessly but on the other hand, the learning of more sophisticated musical skills, as in the case of professional musicians, is time-consuming and necessitates a lot of effort. The transmission of musical information has its roots in the human ability of vocal learning which is exceptional among primates (Janik and Slater, 1997; Fitch and Jarvis, 2013). However, human vocal learning is canalized into an imitation of selected acoustic characteristics rather than the literal copying of every heard sound (Jackendoff and Lerdahl, 2006). People are very skillful at imitating the distinctive features of phonemes, the temporal order of sound sequences, and fundamental frequency of harmonic sounds but not very talented when they try to simulate the barking of a dog or environmental sounds such as the noise of a refrigerator which seem a very simple task for many parrots. This canalization suggests that apart from the aforementioned environmental instability, certain circumstances related to hominin vocal expressions must have been stabilized long enough during numerous generations to cause natural selection to have promoted an instinct to learn only selected sound features (Briscoe, 2000; Gibson and Tallerman, 2011). As a result, speech and music, similar to many songbirds' songs, are examples of so called ritual culture (Merker, 2009). The most important characteristic of ritual culture is its transmission by the means of imitative social learning (Merker, 2005, 2012). In contrast to non-imitative social learning, learning by imitation consists of copying the behavior of other individuals (Jablonka and Lamb, 2005). Therefore, what is important in the transmission of ritual is not the result of a particular action but the action itself (Merker, 2009). In the case of music, transmitted units are pitch classes and rhythm measures (Merker, 2002, 2003). After all, a melody is recognized independent of whether it is played slower or faster on the flute, piano, or when sung. What

is important for the recognition of melodic pattern is its pitch and rhythm structure, not timbre or dynamics. In this respect music seems to be an even more striking example of ritual culture than speech in which, apart from poetry, what is crucial is the transmission of the semantic content of utterance and not its literal form. However, the thing that makes melody easier to remember is musical syntax, the evolution of which is most easily explained by the Baldwin effect.

The Baldwin effect may promote a particular trait due to different adaptive functions. For example, Morgan proposed that organic evolution (the term which Morgan used to describe the process known today as the Baldwin effect) can explain the origin of bird songs which evolved as a result of sexual selection (Morgan, 1891, 1920). Since bird songs are similar to human music in many respects, it is tempting to explain the origin of musical syntax by the means of the Baldwin effect in which the adaptive function of music is to attract sexual partners. If this is true, musicality could have been used by hominins as a mating handicap (a mark of the quality of a mate, Zahavi, 1975; Zahavi and Zahavi, 1997; Miller G. F., 2000) since the production and recognition of musical syntax is costly in terms of energy (necessary to process the perceived sounds and to control the vocal production of songs) and time spent on singing (which for example can be used for foraging instead). In such a scenario the syntactical complexity of a hominin song should attract females more than a song that lacks such a complexity due to the costliness of the song's complex structure being an indicator of fitness (Miller G., 2000). If these female preferences had been stable enough throughout many generations, the Baldwinian mechanism should have transformed the learning of culturally invented rules of musical structure into an instinct to learn and organize musical sounds in a syntactic way. This emerged instinct to learn the distribution of music-specific discrete elements based on intuitive recognition of their probability of occurrence would have left space for idiosyncratic song modifications similar to those observed in songbirds' behavior. Such a leeway in creating new songs allows the sustaining of the process of sexual selection based on the songs' complexity. However, a study of female preferences toward musical complexity showed that women do not have a tendency to prefer more complex music during and around ovulation (Charlton et al., 2012), which does not support the Baldwinian scenario of music origin based on sexual selection. Similarly, research shows that musical aptitude and achievements are not a predictor of mating success (Mosing et al., 2014). Of course these studies are not conclusive and more studies are necessary to test a possible role of sexual selection in music evolution. Nevertheless, so far the Baldwinian scenario in the sexual selection of human musicality needs more empirical support to be convincing.

Another possible scenario of the Baldwinian origin of music is related to the idea that music can serve as a tool of social consolidation. This scenario started the moment a new social challenge first appeared. The increasing size of the hominin population caused an increase in inter-individual and inter-group competition for food and other resources (Dunbar, 2014). One way to cope with this problem was to form alliances between individuals belonging to a group. This strategy has been

observed in other primates, including our closest relatives—the chimpanzee (Mitani, 2009; Gilby et al., 2013), which suggests that hominins could use a similar strategy. Dunbar has proposed that as group size increased, grooming as the main tool to sustain social alliances became insufficient. Instead, hominin vocalizations started to serve as a tool of social consolidation (Dunbar, 1996). While this idea seems to be unconvincing as far as the origin of language is concerned (Dunbar and Lehmann, 2013; Grueter et al., 2013) its validity as an explanation of the origin of music still remains an open question. An increasing number of studies suggest that communal singing can facilitate social bonds (Dunbar et al., 2012; Tarr et al., 2014; Pearce et al., 2015, 2016, 2017), which supports Dunbar's idea. However, the precise mechanism of how music could act in this way remains a puzzle. Although the obtained results of the aforementioned studies must not necessarily be the effects of the adaptive value of social bonding, being for example a byproduct of sexually selected behavior, certain characteristics of musical syntax seem to bespeak the hypothesis of music as a tool of social consolidation. A big problem which afflicts individuals living in groups is the situation in which some individuals use resources obtained by other individuals—the so called “free-riding problem.” Communal song rituals demand that all participants must know the musical structure of the song. The learning of a particular song's structure is time-consuming and necessitates strenuous imitation of the vocal behavior of others especially in the case when hominins did not possess an instinct to learn musical syntax. Therefore, by devoting equal effort in order to learn a ritualized song, communal singing can serve as a good test of being prone to act together with others. After all, poor singers can be easily recognized which can lead to ostracism. In other words, the consolidation effect observed after communal singing can be a product of the unconscious assessment of other individuals in terms of their proclivity to being free-riders. From this perspective, a lack of synchronization hinders consolidation which can be a result of detecting potential free-riders and can lead to looking for new allies.

Also a proximal explanation of the mechanism responsible for the consolidating power of music can be related to observed characteristics of music processing by the human nervous system. It is possible that music consolidates individuals by the means of temporal and spectral synchronization between the brain states of co-performers (Bharucha et al., 2011). If this is true, our predecessors had to simultaneously imitate their vocalizations in order to sustain social trust (Podlipniak, 2016). This collective imitation became the beginning of a consolidating vocal ritual. Without any predisposition which canalized vocal learning so that hominins would have been sensitive to certain acoustic features, the process of the learning of vocal rituals would have been very strenuous and time-consuming. During this time, the second stage of Baldwinian evolution began in which natural selection preferred individuals who were characterized by the most flexible learning. In order to learn new melodies and sing them together then the appropriate predictions of what (which particular pitch class) and when (the position of a particular rhythm measure in relation to musical pulse) would happen in the near future was necessary. Syntax is exactly what makes the

successful predictions of sound events during singing easier and, as a consequence, facilitates collective singing. At this stage, the learning of simple syntactic rules would have been accessible to hominins in a similar way to people that learn artificial syntaxes today. Because the costs of ritual learning were high an individual who was accidentally endowed with proclivities to predict the melody better than others gained an advantage over the rest of a group. In the long run, the progeny of this individual has dominated the whole population.

In a similar vein the Baldwin effect could have contributed to the origin of music if its adaptive function advertised the defending skills of a group (Hagen and Bryant, 2003; Hagen and Hammerstein, 2009; Jordania, 2014). However, in case an acoustic aposematism (instrumental music, singing) had been directed against predators (Jordania, 2014) a possible role of the Baldwinian mechanism in the origins of human musicality would have been restricted solely to the canalization of the elements of musical display which are recognizable by predators. These elements are a part of musicality in a broad sense such as pitch contour, changes in tempo, and dynamics rather than the syntactic relations specific to the discrete structure of human music. In the Baldwinian scenario, the initially invented complexity of musical syntax which became a part of the hominin cultural niche was most probably accessible (comprehensible) only to the hominin species. Therefore, predator reactions did not depend on the subtleties of musical structure and had not been a selective factor which could have influenced the process of canalization of musical-syntactical abilities which actually define musicality in a narrow sense. In contrast, if music was a coalition signaling display directed toward conspecifics (Hagen and Bryant, 2003), the Baldwinian scenario could have been very similar to that presented above in the case of consolidation as the adaptive function of music. The only difference is that this time the selective pressure which would have been responsible for the canalization of human musicality had been induced by the reactions of enemies. However, while rhythm syntax seems to contribute to the signaling/deteriorating function of music (despite the fact that people from different cultures can recognize different units of musical pulse in the same piece of music (London, 2012), a well synchronized rhythm is perceivable independent of whether the rules of rhythm syntax are familiar to us or not), the origin of pitch syntax is more problematic as a result of its possible adaptive signaling/deteriorating function.

First of all, even today when people are most probably endowed with an instinct to learn pitch system and pitch syntax, well spectrally synchronized music which is based on an unfamiliar pitch system can be perceived as being out-of-tune (Ellis, 1885), which can be a sign of a poor performance rather than a signal of a performers' coalition or consolidation. Additionally, the recognition of pitch syntax by contemporary humans depends on tacit knowledge about the statistical distribution of pitch classes in a particular music (Tillmann et al., 2000; Tillmann, 2005; Huron, 2006). Foreigners who listen to unfamiliar music usually experience different tonal qualia than people familiar with this music (Castellano et al., 1984; Kessler et al., 1984; Stevens, 2004, 2012; Curtis and Bharucha, 2009). Although this situation looks similar to the

aforesaid difference in musical pulse perception, it differs in one important aspect. Pitch syntax is based on pitch hierarchy in which the most prominent place is pitch center, the experience of which is accompanied by the emotional qualia of completeness, resolution etc. There is nothing resembling pitch center in rhythm hierarchy. The misrecognition of actual tonal relations in reference to pitch center by foreigners can lead to divergence between the observed emotional expression of performers (also by the means of expressive dynamics) and tonal qualia felt by those foreigners. Such a divergence can also be a signal of poor performance. Importantly, without the canalized strategy to learn pitch syntax the differences in musical dialects between hominin groups would have been even greater than between modern geographically distant musical cultures causing the aforesaid divergence to be even greater. Therefore, while in the scenario of music origin in which music is a coalition signaling system the Baldwinian mechanism can explain the origin of musical rhythm so it is difficult to imagine a similar role of the Baldwin effect in the origin of pitch syntax as a result of its coalition signaling function.

It is worth mentioning however, that the possible different adaptive functions of music are not mutually exclusive and the Baldwin effect could have played an important role at different stages of the gradual process of the evolution of human musicality. Nevertheless, it seems that Baldwinian evolution was necessary at least in the process that led to the emergence of the complex musical syntax as a part of hominins' singing behavior.

## THE EVOLUTION OF NEW CIRCUITRY

The appearance of ritualized singing behavior among hominins required the development of new abilities and recruitment of existing skills. An important ability which had to be a necessary condition of the development of culturally variable vocal communication is the aforesaid vocal learning (Janik and Slater, 1997). This ability had to be present at the first stage of the Baldwinian evolution of human musicality. The similarity of vocal pathways in vocal learning birds to cortical–basal ganglia–thalamic–cortical loops in humans suggests the role of the latter in the processing of speech (Jarvis, 2007). In fact, there is an increasing number of studies which emphasize the role of the basal ganglia in the processing of language (Booth et al., 2005) especially in the learning of language during childhood (Krishnan et al., 2016). The fact that language impairment can be the result of neurodevelopmental deficits of the corticostriatal loops (Krishnan et al., 2016) shows that corticostriatal connectivity might have been an important element of evolutionary change leading to the evolution of vocal communication among our predecessors. It is not surprising that the basal ganglia contributes to the processing of sound sequences specific to vocal communication. It is known that the basal ganglia is important in reinforcement learning (Bar-Gad et al., 2003) which is necessary to acquire the majority of culturally transmitted information. This ability is strictly connected to predictive functions which are related to internal timing (Dreher and Grafman, 2002). It is reasonable to assume that in the beginning hominin vocal

communication was composed of simple vocal expressions which must have been vocally learned.

In the process of vocal learning the recognition and prediction of distinctive acoustic features is necessary and so the specific connections between the basal ganglia and auditory cortices had to be favored by natural selection during the second stage of the Baldwinian evolution of human musicality. Because contemporary humans are characterized by the ability to sustain and reproduce  $F_0$  which is crucial for singing (Bannan, 2012) but not for speaking, the sequencing of sounds based on their  $F_0$  characteristics must have become an important trait of hominin sound rituals. The consolidating function of them did not necessitate any referential conceptual meaning. Instead simple emotional cues were enough to establish close social relations. The emotional reinforcement of a well predicted sound i.e., the sound which is perceived as possessing a particular pitch and which happens at an exactly predicted point of time, started to act as a cue for social acceptance. The positive emotional reaction in response to music is in fact an evolutionarily old clue of social acceptance. It has been observed that this emotional reward occurring in response to music is related to the corticostriatal interactions involving auditory cortices and the nucleus accumbens (Salimpoor et al., 2013). Also the perception of beat in music is based on corticostriatal interactions (Grahn and Rowe, 2009, 2013). It is suspected that the role of the cortico-basal ganglia circuits in speech evolution is related to the positive selection of the FOXP2 gene variant (Enard, 2011). However, because the mutation of FOXP2 also impairs rhythm processing in music leaving intact pitch processing (Alcock et al., 2000; Tan et al., 2014) the evolution of the abilities to recognize pitch syntax must have been related to other genetic factors. Independent of what particular genetic factors influence pitch and rhythm processing in music (Tan et al., 2014), human perceptive preferences to recognize pitch classes and rhythm measures as parts of syntactically organized sequences suggest that at the last stage of the Baldwinian scenarios natural selection started to prefer individuals endowed with these canalized perceptive proclivities.

## CONCLUSION

The proposed possible Baldwinian scenarios of the evolution of human musicality solve the problem of the specificity of musical syntax which is functionally and structurally different from language syntax. This specificity suggests that all theories which explain the syntactic characteristic of music as a byproduct seem unconvincing. After all, music seems to be the only spontaneously emerging syntactic system apart from speech, dance (Opacic et al., 2009), and some drum (Winter, 2014) and whistled languages (Carreiras et al., 2005; Güntürkün et al., 2015; Meyer and Busnel, 2015). Although the details of these proposed scenarios are speculative, future research can elucidate which particular elements of these scenarios are more probable. The most promising would be studies which concentrate on the comparison between the functions of cortico-striatal loops (Gorzelańczyk, 2011) in speech and music processing. In order

to investigate which particular loop is mostly involved in the processing of musical syntax, neuroimaging studies could be conducted in which the activity of limbic and dorsolateral-prefrontal loops can be compared during performing tasks related to the recognition of musical and language syntaxes. Although the complex interactions between genetic, epigenetic, and cultural information which occur in evolution have not so far been explained in detail (Jablonka and Lamb, 2005) it seems important to consider them in future models of the evolution of human musicality. The rapidly advancing development of genomics and transcriptomics allows us to expect that the details of these complex interactions will be better understood in the near future.

Another possible study which could be conducted to test the proposed Baldwinian origin of musicality is to compare the results of singing syntactically simple with syntactically complex (more demanding in terms of explicit learning) tonal melodies. If singing syntactically complex melodies leads to a greater consolidation of singers or if the singing group is assessed by others as being more consolidated than singers of the syntactically simple melodies then it would suggest that the tendency which was proposed as the main source

of the Baldwinian evolution of musicality is still present in the human population. Also the comparison of singing tonal melodies by a group of people with other collective sound expressions such as simultaneously reading prose, reciting poetry, drumming, and singing atonal melodies in free rhythm (without musical syntax) would be informative as far as the question of what particular features of human sound expressions are responsible for the observed effects. If the proposed consolidating function of human musicality is tenable, then the consolidating effect of singing tonal melodies and drumming should be greater than other collective behaviors. Nevertheless, in order to better understand the origin of music the broad holistic view of human musicality is necessary. This view should be based on integrated knowledge taken from such disciplines as genetics, evolutionary biology, paleoanthropology, neuroscience, psychology, archeology, ethnomusicology and cognitive musicology.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## REFERENCES

- Alcock, K. J., Passingham, R. E., Watkins, K., and Vargha-Khadem, F. (2000). Pitch and timing abilities in inherited speech and language impairment. *Brain Lang.* 75, 34–46. doi: 10.1006/brln.2000.2323
- Asano, R., and Boeckx, C. (2015). Syntax in language and music: what is the right level of comparison? *Front. Psychol.* 6:942. doi: 10.3389/fpsyg.2015.00942
- Baldwin, J. M. (1896a). A new factor in evolution. *Am. Nat.* 30, 441–451. doi: 10.1086/276408
- Baldwin, J. M. (1896b). A new factor in evolution (Continued). *Am. Nat.* 30, 536–553. doi: 10.1086/276428
- Bannan, N. (2012). “Harmony and its role in human evolution,” in *Music, Language, and Human Evolution*, ed N. Bannan (Oxford: Oxford University Press), 288–340.
- Bar-Gad, I., Morris, G., and Bergman, H. (2003). Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Prog. Neurobiol.* 71, 439–473. doi: 10.1016/j.pneurobio.2003.12.001
- Bharucha, J., Curtis, M., and Paroo, K. (2011). “Musical communication as alignment of brain states,” in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. A. Hawkins, and I. Cross (Oxford; New York, NY: Oxford University Press), 139–155.
- Bickerton, D. (2009). “Syntax for Non-syntacticians a brief primer,” in *Biological Foundations and Origin of Syntax*, eds D. Bickerton and E. Szathmáry (Cambridge; London: The MIT Press), 3–13.
- Bickerton, D. (2010). *Adam’s Tongue: How Humans Made Language, How Language Made Humans*. New York, NY: Hill and Wang.
- Bispham, J. C. (2009). Music’s “design features”: musical motivation, musical pulse, and musical pitch. *Music. Sci.* 13, 41–61. doi: 10.1177/1029864909013002041
- Blacking, J. (1973). *How Musical is Man?* Seattle, WC; London: University of Washington Press.
- Booth, J. R., Wood, L., Lu, D., Houk, J. C., Bitan, T., Kotz, S. A., et al. (2005). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Brain* 128, 392–399. doi: 10.1016/j.tics.2010.06.005
- Bregman, M. R., Patel, A. D., and Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proc. Natl. Acad. Sci. U.S.A.* 113, 1666–1671. doi: 10.1073/pnas.1515380113
- Briscoe, T. (2000). Grammatical acquisition: inductive bias and coevolution of language and the language acquisition device. *Language* 76, 245–296. doi: 10.2307/417657
- Brown, S., and Jordania, J. (2011). Universals in the world’s musics. *Psychol. Music* 41, 229–248. doi: 10.1177/0305735611425896
- Burns, E. M., and Ward, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *J. Acoust. Soc. Am.* 63, 456–468. doi: 10.1121/1.381737
- Cantor, M., and Whitehead, H. (2013). The interplay between social networks and culture: theoretically and among whales and dolphins. The interplay between social networks and culture: theoretically and among whales and dolphins. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368, 1–10. doi: 10.1098/rstb.2012.0340
- Carreiras, M., Lopez, J., Rivero, F., and Corina, D. (2005). Linguistic perception: neural processing of a whistled language. *Nature* 433, 31–32. doi: 10.1038/433031a
- Castellano, M. A., Bharucha, J. J., and Krumhansl, C. L. (1984). Tonal hierarchies in the music of north India. *J. Exp. Psychol. Gen.* 113, 394–412. doi: 10.1037/0096-3445.113.3.394
- Charlton, B. D., Filippi, P., Fitch, W. T., Hamilton, C., and Perrett, D. (2012). Do women prefer more complex music around ovulation? *PLoS ONE* 7:e35626. doi: 10.1371/journal.pone.0035626
- Cross, I. (2005). “Music and meaning, ambiguity and evolution,” in *Musical Communication*, eds D. Miell, R. MacDonald, and D. J. Hargreaves (Oxford; New York, NY: Oxford University Press), 27–44.
- Cross, I., and Morley, I. (2008). “The evolution of music: theories, definitions and the nature of the evidence,” in *Communicative Musicality*, eds S. Malloch and C. Trevarthen (Oxford; New York, NY: Oxford University Press), 61–82.
- Curtis, M. E., and Bharucha, J. J. (2009). Memory and musical expectation for tones in cultural context. *Music Percept.* 26, 365–375. doi: 10.1525/mp.2009.26.4.365
- Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex, 1st Edn.* London: John Murray.
- Deacon, T. W. (1997). *The Symbolic Species: The Evolution of Language and the Brain*. New York, NY; London: W.W. Norton.
- Deacon, T. W. (2003). “Multilevel selection in a complex adaptive system: The problem of language origins,” in *Evolution and Learning: The Baldwin Effect Reconsidered. Life and Mind*, eds B. H. Weber and D. J. Depew (Cambridge; London: Bradford Book; MIT Press), 81–106.

- Deacon, T. W. (2010). A role for relaxed selection in the evolution of the language capacity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9000–9006. doi: 10.1073/pnas.0914624107
- Dissanayake, E. (2008). If music is the food of love, what about survival and reproductive success? *Mus. Sci.* 12, 169–195. doi: 10.1177/1029864908012001081
- Dor, D., and Jablonka, E. (2000). From cultural selection to genetic selection: a framework for the evolution of language. *Selection* 1, 33–56. doi: 10.1556/Select.1.2000.1-3.5
- Dor, D., and Jablonka, E. (2001). “How language changed the genes: toward an explicit account of the evolution of language,” in *New Essays on the Origin of Language* eds J. Trabant and S. Ward (Berlin; New York, NY: De Gruyter Mouton), 149–176.
- Dor, D., and Jablonka, E. (2010). “Plasticity and canalization in the evolution of linguistic communication: an evolutionary developmental approach,” in *The Evolution of Human Language*, eds R. K. Larson, V. DePrez, and H. Yamakido (Cambridge: Cambridge University Press), 135–147.
- Dreher, J.-C., and Grafman, J. (2002). The roles of the cerebellum and basal ganglia in timing and error prediction. *Eur. J. Neurosci.* 16, 1609–1619. doi: 10.1046/j.1460-9568.2002.02212.x
- Dunbar, R. I. M. (1996). *Grooming, Gossip, and the Evolution of Language*. London: Harvard University Press.
- Dunbar, R. I. M. (2014). *Human Evolution*. London: Penguin books Ltd.
- Dunbar, R. I. M., and Lehmann, J. (2013). Grooming and social cohesion in primates: a comment on Grueter et al. *Evol. Hum. Behav.* 34, 453–455. doi: 10.1016/j.evolhumbehav.2013.08.003
- Dunbar, R., Kaskatis, K., MacDonald, I., and Barra, V. (2012). Performance of music elevates pain threshold and positive affect: implications for the evolutionary function of music. *Evol. Psychol.* 10, 688–702. doi: 10.1177/147470491201000403
- Dutton, D. G., and Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *J. Pers. Soc. Psychol.* 30, 510–517. doi: 10.1037/h0037031
- Ellis, A. (1885). On the musical scales of various nations. *J. Soc. Arts* 33, 485–527.
- Enard, W. (2011). FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. *Curr. Opin. Neurobiol.* 21, 415–424. doi: 10.1016/j.conb.2011.04.008
- Fadiga, L., Craighero, L., and D’Ausilio, A. (2009). Broca’s area in language, action, and music. *Ann. N.Y. Acad. Sci.* 1169, 448–458. doi: 10.1111/j.1749-6632.2009.04582.x
- Fehér, O., Ljubičić, I., Suzuki, K., Okanoya, K., and Tchernichovski, O. (2016). Statistical learning in songbirds: from self-tutoring to song culture. *Philos. Trans. R. Soc. London B Biol. Sci.* 372:20160053. doi: 10.1098/rstb.2016.0053
- Fitch, W. T. (2006). On the biology and evolution of music. *Music Percept.* 24, 85–88. doi: 10.1525/mp.2006.24.1.85
- Fitch, W. T. (2013a). “Musical Protolanguage: Darwin’s theory of language evolution revisited,” in *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain*, eds J. J. Bolhuis and M. Everaert (Cambridge; London: The MIT Press), 489–503.
- Fitch, W. T. (2013b). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Front. Syst. Neurosci.* 7:68. doi: 10.3389/fnsys.2013.00068
- Fitch, W. T. (2014). Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Phys. Life Rev.* 11, 329–364. doi: 10.1016/j.plrev.2014.04.005
- Fitch, W. T. (2015). Four principles of bio-musicology. *Philos. Trans. R. Soc. London B Biol. Sci.* 370:20140091. doi: 10.1098/rstb.2014.0091
- Fitch, W. T., and Jarvis, E. D. (2013). “Birdsong and other animal models for human speech, song, and vocal learning,” in *Language, Music and the Brain*, ed M. A. Arbib (Cambridge; London: The MIT Press), 499–539.
- Fox, L. (1988). *The Jew’s Harp: a Comprehensive Anthology*. London: Associated University Presses.
- Friedrich, R., and Friederici, A. D. (2009). Mathematical logic in the human brain: syntax. *PLoS ONE* 4:e5599. doi: 10.1371/journal.pone.0005599
- Gibson, K. R., and Tallerman, M. (2011). “Introduction to Part II: the biology of language evolution: anatomy, genetics and neurology,” in *The Oxford Handbook of Language Evolution*, eds K. Gibson and M. Tallerman (Oxford; New York, NY: Oxford University Press), 133–142.
- Gilby, I. C., Brent, L. J. N., Wroblewski, E. E., Rudicell, R. S., Hahn, B. H., Goodall, J., et al. (2013). Fitness benefits of coalitionary aggression in male chimpanzees. *Behav. Ecol. Sociobiol.* 67, 373–381. doi: 10.1007/s00265-012-1457-6
- Gintis, H. (2011). Gene-culture coevolution and the nature of human sociality. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 366, 878–888. doi: 10.1098/rstb.2010.0310
- Godfrey-Smith, P. (2003). “Between Baldwin Skepticism and Baldwin Boosterism,” in *Evolution and Learning: The Baldwin Effect Reconsidered*, eds B. H. Weber and D. J. Depew (Cambridge: The MIT Press), 53–67.
- Gozelańczyk, E. J. (2011). “Functional Anatomy, Physiology and Clinical Aspects of Basal Ganglia,” in *Neuroimaging for Clinicians - Combining Research and Practice*, ed J. F. P. Peres (Rijeka: InTech), 89–106.
- Gozelańczyk, E. J., Podlipniak, P., Walecki, P., Karpiński, M., and Tarnowska, E. (2017). Pitch Syntax violations are linked to greater skin conductance changes, relative to timbral violations – the predictive role of the reward system in perspective of cortico-subcortical loops. *Front. Psychol.* 8:586. doi: 10.3389/fpsyg.2017.00586
- Grahn, J. A., and Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J. Neurosci.* 29, 7540–7548. doi: 10.1523/JNEUROSCI.2018-08.2009
- Grahn, J. A., and Rowe, J. B. (2013). Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cereb. Cortex* 23, 913–921. doi: 10.1093/cercor/bhs083
- Grueter, C. C., Bissonnette, A., Isler, K., and van Schaik, C. P. (2013). Grooming and group cohesion in primates: implications for the evolution of language. *Evol. Hum. Behav.* 34, 61–68. doi: 10.1016/j.evolhumbehav.2012.09.004
- Güntürkün, O., Güntürkün, M., and Hahn, C. (2015). Whistled Turkish alters language asymmetries. *Curr. Biol.* 25, R706–R708. doi: 10.1016/j.cub.2015.06.067
- Hagen, E. H., and Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Hum. Nat.* 14, 21–51. doi: 10.1007/s12110-003-1015-z
- Hagen, E. H., and Hammerstein, P. (2009). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. *Mus. Sci.* 13, 291–320. doi: 10.1177/1029864909013002131
- Hall, B. K. (2001). Organic selection: proximate environmental effects on the evolution of morphology and behaviour. *Biol. Philos.* 16, 215–237. doi: 10.1023/A:1006773408919
- Harvey, A. R. (2017). *Music, Evolution, and the Harmony of Souls*. Oxford: Oxford University Press.
- Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569
- Heffner, C. C., and Slevc, L. R. (2015). Prosodic structure as a parallel to musical structure. *Front. Psychol.* 6:1962. doi: 10.3389/fpsyg.2015.01962
- Higuchi, S., Chaminade, T., Imamizu, H., and Kawato, M. (2009). Shared neural correlates for language and tool use in Broca’s area. *Neuroreport* 20, 1376–1381. doi: 10.1097/WNR.0b013e3283315570
- Hilliard, A. T., and White, S. A. (2009). “Possible Precursors of Syntactic Components in Other Species,” in *Biological Foundations and Origin of Syntax*, eds D. Bickerton and E. Szathmáry (Cambridge; London: The MIT Press), 161–183.
- Huron, D. B. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge; London: The MIT Press.
- Jablonka, E., and Lamb, M. J. (1995). *Epigenetic Inheritance and Evolution: the Lamarckian Dimension*. Oxford; New York, NY; Tokyo: Oxford University Press.
- Jablonka, E., and Lamb, M. J. (2005). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge; London: MIT Press.
- Jackendoff, R., and Lerdahl, F. (2006). The capacity for music: what is it, and what’s special about it? *Cognition* 100, 33–72. doi: 10.1016/j.cognition.2005.11.005
- Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166. doi: 10.1126/science.860134
- Janik, V. M., and Slater, P. J. B. (1997). Vocal learning in mammals. *Adv. Study Behav.* 26, 59–99. doi: 10.1016/S0065-3454(08)60377-0
- Jarvis, E. D. (2007). Neural systems for vocal learning in birds and humans: a synopsis. *J. Ornithol.* 148, 35–44. doi: 10.1007/s10336-007-0243-0

- Jordania, J. (2014). *Tigers, Lions and Humans: History of Rivalry, Conflict, Reverence and Love*. Tbilisi: Logos.
- Kanduri, C., Kuusi, T., Ahvenainen, M., Philips, A. K., Lähdesmäki, H., and Järvelä, I. (2015). The effect of music performance on the transcriptome of professional musicians. *Sci. Rep.* 5:9506. doi: 10.1038/srep09506
- Kessler, E. J., Hansen, C., and Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and in the west. *Music Percept.* 2, 131–165. doi: 10.2307/40285289
- Koelsch, S. (2013). *Brain and Music*. Oxford: Wiley-Blackwell.
- Koelsch, S., Kilches, S., Steinbeis, N., and Schelinski, S. (2008). Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLoS ONE* 3:e2631. doi: 10.1371/journal.pone.0002631
- Krishnan, S., Watkins, K. E., and Bishop, D. V. M. (2016). Neurobiological basis of language learning difficulties. *Trends Cogn. Sci.* 20, 701–714. doi: 10.1016/j.tics.2016.06.012
- Lerdahl, F. (1987). Timbral hierarchies. *Contemp. Music Rev.* 2, 135–160. doi: 10.1080/07494468708567056
- Lerdahl, F. (2013). "Musical syntax and its relation to linguistic syntax," in *Language, Music, and the Brain*, ed M. A. Arbib (Cambridge; London: The MIT Press), 257–272.
- Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge; London: The MIT Press.
- Levin, T. C., and Süzükei, V. (2006). *Where Rivers and Mountains Sing: Sound, music, and Nomadism in Tuva and Beyond*. Bloomington: Indiana University Press.
- Llinás, R. R. (2001). *I of the Vortex: from Neurons to Self*. Cambridge; London: MIT Press.
- London, J. (2011). "Schemas, not syntax: a reply to Patel," in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. A. Hawkins, and I. Cross (Oxford; New York, NY: Oxford University Press), 242–247.
- London, J. (2012). Three things linguists need to know about rhythm and time in music. *Empir. Musicol. Rev.* 7, 5–11. doi: 10.18061/1811/52973
- Lumsden, C. J., and Wilson, E. O. (1982). Précis of genes, mind, and culture. *Behav. Brain Sci.* 5, 1–37. doi: 10.1017/S0140525X00010128
- Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.* 4, 540–545. doi: 10.1038/87502
- Margulis, E. H. (2014). *On Repeat: how Music Plays the Mind*. Oxford; New York, NY: Oxford University Press.
- McAdams, S., and Giordano, B. L. (2008). "The perception of musical timbre," in *The Oxford Handbook of Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford; New York, NY: Oxford University Press), 72–80.
- Merker, B. (2002). Music: the missing humboldt system. *Music. Sci.* 6, 3–21. doi: 10.1177/102986490200600101
- Merker, B. (2003). "Is there a biology of music? And why does it matter?" in *Proceedings of the 5th Triennial ESCOM Conference*, eds C. Kopiez, R. Lehmann, A. C. Wolther, and I. Wolf (Hanover: Hanover University of Music and Drama), 402–405.
- Merker, B. (2005). The conformal motive in birdsong, music, and language: an introduction. *Ann. N.Y. Acad. Sci.* 1060, 17–28. doi: 10.1196/annals.1360.003
- Merker, B. (2009). "Ritual foundations of human uniqueness," in *Communicative Musicality. Exploring the Basis of Human Companionship*, eds C. Malloch and S. Trevarthen (Oxford; New York, NY: Oxford University Press), 45–59.
- Merker, B. (2012). "The Vocal Learning Constellation," in *Music, Language, and Human Evolution*, ed N. Bannan (London: Oxford University Press), 215–260.
- Meyer, J., and Busnel, R. G. (2015). *Whistled Languages: A Worldwide Inquiry on Human Whistled Speech*. New York, NY; London: Springer-Verlag.
- Mikutta, C. A., Dürschmid, S., Bean, N., Lehne, M., Lubell, J., Altorfer, A., et al. (2015). Amygdala and orbitofrontal engagement in breach and resolution of expectancy: a case study. *Psychomusicol. Music Mind Brain* 25, 357–365. doi: 10.1037/pmu0000121
- Miller, G. (2000). Mental traits as fitness indicators. Expanding evolutionary psychology's adaptationism. *Ann. N.Y. Acad. Sci.* 907, 62–74. doi: 10.1111/j.1749-6632.2000.tb06616.x
- Miller, G. F. (2000). "Evolution of human music through sexual selection," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge: The MIT Press), 329–360.
- Mitani, J. C. (2009). Cooperation and competition in chimpanzees: current understanding and future challenges. *Evol. Anthropol. Issues News Rev.* 18, 215–227. doi: 10.1002/evan.20229
- Moreno, S., and Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hear. Res.* 308, 84–97. doi: 10.1016/j.heares.2013.09.012
- Morgan, C. L. (1891). *Animal Life and Intelligence*. Boston, MA: Ginn & Company.
- Morgan, C. L. (1896). *Habitat and Instinct*. London: Edward Arnold Press.
- Morgan, C. L. (1920). *Animal Behaviour*. London: Edward Arnold Press.
- Mosing, M. A., Verweij, K. J. H., Madison, G., Pedersen, N. L., Zietsch, B. P., and Ullen, F. (2014). Did sexual selection shape human music? Testing predictions from the sexual selection hypothesis of music evolution using a large genetically informative sample of over 10,000 twins. *Evol. Hum. Behav.* 36, 359–366. doi: 10.1016/j.evolhumbehav.2015.02.004
- Nakata, T., and Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behav. Dev.* 27, 455–464. doi: 10.1016/j.infbeh.2004.03.002
- Nettl, B. (2000). "An ethnomusicologist contemplates universals in musical sound and musical culture," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, UK: The MIT Press), 463–472.
- Nikolsky, A. (2015). Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. Part-1: Prehistoric. *Front. Psychol.* 6:1405. doi: 10.3389/fpsyg.2015.0140
- Okanoya, K. (2013). "Finite-state song syntax in bengalese finches: sensorimotor evidence, developmental processes, and formal procedures for syntax extraction," in *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain*, eds J. J. Bolhuis and M. Everaert (Cambridge, UK: MIT Press), 229–242.
- Opacic, T., Stevens, C., and Tillmann, B. (2009). Unspoken knowledge: implicit learning of structured human dance movement. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1570–1577. doi: 10.1037/a0017244
- Osborn, H. F. (1896). A mode of evolution requiring neither natural selection nor the inheritance of acquired characters. *Trans. N.Y. Acad. Sci.* 15, 141–142.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York, NY: Oxford University Press.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681. doi: 10.1038/nm1082
- Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford; New York, NY: Oxford University Press.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (1998). Processing syntactic relations in language and music: an event-related potential study. *J. Cogn. Neurosci.* 10, 717–733. doi: 10.1162/089892998563121
- Pearce, E., Launay, J., and Dunbar, R. I. M. (2015). The ice-breaker effect: singing mediates fast social bonding. *R. Soc. Open Sci.* 2, 1–9. doi: 10.1098/rsos.150221
- Pearce, E., Launay, J., van Duijn, M., Rotkirch, A., David-Barrett, T., and Dunbar, R. I. M. (2016). Singing together or apart: the effect of competitive and cooperative singing on social bonding within and between sub-groups of a university Fraternity. *Psychol. Music* 44, 1255–1273. doi: 10.1177/03057356166636208
- Pearce, E., MacCarron, P., Launay, J., and Dunbar, R. I. M. (2017). Tuning in to others: exploring relational and collective bonding in singing and non-singing groups over time. *Psychol. Music* 45, 496–512. doi: 10.1177/0305735616667543
- Perlovsky, L. (2010). Musical emotions: functions, origins, evolution. *Phys. Life Rev.* 7, 2–27. doi: 10.1016/j.plrev.2009.11.001
- Perlovsky, L. (2012). Cognitive function, origin, and evolution of musical emotions. *Music. Sci.* 16, 185–199. doi: 10.1177/1029864912448327
- Podlipniak, P. (2015). "The origin of music and the Baldwin effect," in *Proceedings of Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music*, eds J. Ginsborg, A. Lamont, and S. Bramley (Manchester: Royal Northern College of Music), 671–677.
- Podlipniak, P. (2016). The evolutionary origin of pitch centre recognition. *Psychol. Music* 44, 527–543. doi: 10.1177/0305735615577249
- Purves, D. (2017). *Music as Biology: The Tones We Like and Why*. Cambridge: Harvard University Press.
- Reber, A. S., Allen, R., and Reber, P. J. (1999). "Implicit versus Explicit Learning," in *The Nature of Cognition*, ed R. J. Sternberg (Cambridge: MIT Press), 475–514.
- Remijns, B. (2014). The study of tone in languages with a quantity contrast. *Lang. Doc. Conserv.* 8, 672–689. Available online at: <http://hdl.handle.net/10125/24620>

- Remijsen, B., and Gilley, L. (2008). Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *J. Phon.* 36, 318–344. doi: 10.1016/j.wocn.2007.09.002
- Repp, B. H. (2007). Hearing a melody in different ways: multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization. *Cognition* 102, 434–454. doi: 10.1016/j.cognition.2006.02.003
- Reybrouck, M. (2013). From sound to music: an evolutionary approach to musical semantics. *Biosemiotics* 6, 585–606. doi: 10.1007/s12304-013-9192-6
- Richerson, P. J., Boyd, R., and Henrich, J. (2010). Gene-culture coevolution in the age of genomics. *Proc. Natl. Acad. Sci. U.S.A.* 107(Suppl.), 8985–8992. doi: 10.1073/pnas.0914631107
- Roederer, J. G. (1984). The search for a survival value of music. *Music Percept. Interdiscip. J.* 1, 350–356. doi: 10.2307/40285265
- Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., and Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340, 216–219. doi: 10.1126/science.1231059
- Savage, P. E., Brown, S., Sakai, E., and Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8987–8992. doi: 10.1073/pnas.1414495112
- Seifert, U., Verschure, P. F. M. J., Arbib, M. A., Cohen, A. J., Fogassi, L., Fritz, T., et al. (2013). “Semantics of internal and external worlds,” in *Language, Music, and the Brain* (Cambridge; London: The MIT Press), 203–230. doi: 10.7551/mitpress/9780262018104.003.0008
- Shannon, R. V. (2016). Is birdsong more like speech or music? *Trends Cogn. Sci.* 20, 245–247. doi: 10.1016/j.tics.2016.02.004
- Sievers, B., Polansky, L., Casey, M., and Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proc. Natl. Acad. Sci. U.S.A.* 110, 70–75. doi: 10.1073/pnas.1209023110
- Simpson, G. G. (1953). The Baldwin Effect. *Evolution* 7, 110–117.
- Slocombe, K. E., Townsend, S. W., and Zuberbühler, K. (2009). Wild chimpanzees (*Pan troglodytes schweinfurthii*) distinguish between different scream types: evidence from a playback study. *Anim. Cogn.* 12, 441–449. doi: 10.1007/s10071-008-0204-x
- Smith, J. D., Nelson, D. G., Grohskopf, L. A., and Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition* 52, 23–54. doi: 10.1016/0010-0277(94)90003-5
- Stainsby, T., and Cross, I. (2008). “The perception of pitch,” in *The Oxford Handbook of Music Psychology*, eds. S. Hallam, I. Cross, and M. Thaut (Oxford; New York, NY: Oxford University Press), 47–58.
- Steinbeis, N., Koelsch, S., and Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses. *J. Cogn. Neurosci.* 18, 1380–1393. doi: 10.1162/jocn.2006.18.8.1380
- Stevens, C. (2004). Cross-cultural studies of musical pitch and time. *Acoust. Sci. Technol.* 25, 433–438. doi: 10.1250/ast.25.433
- Stevens, C. J. (2012). Music perception and cognition: a review of recent cross-cultural research. *Top. Cogn. Sci.* 4, 653–667. doi: 10.1111/j.1756-8765.2012.01215.x
- Storr, A. (1992). *Music and the Mind*. New York, NY: Ballantine Books.
- Strait, D. L., and Kraus, N. (2014). Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hear. Res.* 308, 109–121. doi: 10.1016/j.heares.2013.08.004
- Suzuki, T. N., Wheatcroft, D., and Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nat. Commun.* 7:10986. doi: 10.1038/ncomms10986
- Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., and Wilson, S. J. (2014). The genetic basis of music ability. *Front. Psychol.* 5:658. doi: 10.3389/fpsyg.2014.00658
- Tarr, B., Launay, J., and Dunbar, R. I. M. (2014). Music and social bonding: “Self-other” merging and neurohormonal mechanisms. *Front. Psychol.* 5:1096. doi: 10.3389/fpsyg.2014.01096
- Tettamanti, M., and Weniger, D. (2006). Broca’s area: a supramodal hierarchical processor? *Cortex* 42, 491–494. doi: 10.1016/S0010-9452(08)70384-8
- Tillmann, B. (2005). Implicit investigations of tonal knowledge in nonmusical listeners. *Ann. N.Y. Acad. Sci.* 1060, 100–110. doi: 10.1196/annals.1360.007
- Tillmann, B., Bharucha, J. J., and Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychol. Rev.* 107, 885–913. doi: 10.1037/0033-295X.107.4.885
- Tinbergen, N. (1963). On aims and methods of Ethology. *Z. Tierpsychol.* 20, 410–433. doi: 10.1111/j.1439-0310.1963.tb01161.x
- Toates, F. M. (1988). “Motivation and emotion from a biological perspective,” in *Cognitive Perspectives on Emotion and Motivation*, eds V. Hamilton, G. H. Bower, and N. H. Frijda (Dordrecht: Springer), 3–35.
- Van de Cavey, J., and Hartsuiker, R. J. (2016). Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions, and language. *Cognition* 146, 172–184. doi: 10.1016/j.cognition.2015.09.013
- van de Waal, E., Borgeaud, C., and Whiten, A. (2013). Potent social learning and conformity shape a wild primate’s foraging decisions. *Science* 340, 483–485. doi: 10.1126/science.1232769
- Waddington, C. H. (1953a). Genetic assimilation of an acquired character. *Evolution* 7, 118–126.
- Waddington, C. H. (1953b). The “Baldwin effect”, “Genetic Assimilation”, and “Homeostasis.” *Evolution* 7, 386–387.
- Wakita, M. (2014). Broca’s area processes the hierarchical organization of observed action. *Front. Hum. Neurosci.* 7:937. doi: 10.3389/fnhum.2013.00937
- Winter, Y. (2014). On the grammar of a Senegalese drum language. *Language* 90, 644–668. doi: 10.1353/lan.2014.0061
- Wong, A. W.-K., Huang, J., Chen, H.-C., Chen, H., and Hills, M. (2012). Phonological units in spoken word production: insights from cantonese. *PLoS ONE* 7:e48776. doi: 10.1371/journal.pone.0048776
- Woolhous, M., Cross, I., and Horton, T. (2016). Perception of nonadjacent tonic-key relationships. *Psychol. Music* 44, 802–815. doi: 10.1177/0305735615593409
- Xu, L., Thompson, C. S., and Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *J. Acoust. Soc. Am.* 117, 3255–3267. doi: 10.1121/1.1886405
- Zahavi, A. (1975). Mate selection—A selection for a handicap. *J. Theor. Biol.* 53, 205–214. doi: 10.1016/0022-5193(75)90111-3
- Zahavi, A., and Zahavi, A. (1997). *The Handicap Principle: A Missing Piece of Darwin’s Puzzle*. New York, NY; Oxford: Oxford University Press.
- Zatorre, R. J., and Baum, S. R. (2012). Musical melody and speech intonation: singing a different tune. *PLoS Biol.* 10:e1001372. doi: 10.1371/journal.pbio.1001372
- Zimmermann, E., Leliveld, L., and Schehka, S. (2013). “Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices,” in *Evolution of Emotional Communication: from Sounds in Nonhuman Mammals to Speech and Music in Man*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford; New York, NY: Oxford University Press), 116–132.

**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 © 2017 Podlipniak. 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) or licensor 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.