



# Replication unreliability in psychology: elusive phenomena or “elusive” statistical power?

Patrizio E. Tressoldi\*

Dipartimento di Psicologia Generale, Università di Padova, Padova, Italy

## Edited by:

Jason W. Osborne, Old Dominion University, USA

## Reviewed by:

Fiona Fidler, University of Melbourne, Australia

Darrell Hull, University North Texas, USA

Donald Sharpe, University of Regina, Canada

## \*Correspondence:

Patrizio E. Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Padova, Italy.  
e-mail: patrizio.tressoldi@unipd.it

The focus of this paper is to analyze whether the unreliability of results related to certain controversial psychological phenomena may be a consequence of their low statistical power. Applying the Null Hypothesis Statistical Testing (NHST), still the widest used statistical approach, unreliability derives from the failure to refute the null hypothesis, in particular when exact or quasi-exact replications of experiments are carried out. Taking as example the results of meta-analyses related to four different controversial phenomena, subliminal semantic priming, incubation effect for problem solving, unconscious thought theory, and non-local perception, it was found that, except for semantic priming on categorization, the statistical power to detect the expected effect size (ES) of the typical study, is low or very low. The low power in most studies undermines the use of NHST to study phenomena with moderate or low ESs. We conclude by providing some suggestions on how to increase the statistical power or use different statistical approaches to help discriminate whether the results obtained may or may not be used to support or to refute the reality of a phenomenon with small ES.

**Keywords:** incubation effect, non-local perception, power, subliminal priming, unconscious thought theory

## INTRODUCTION

### ARE THERE ELUSIVE PHENOMENA OR IS THERE AN “ELUSIVE” POWER TO DETECT THEM?

When may a phenomenon be considered to be real or very probable, following the rules of current scientific methodology? Among the many requirements, there is a substantial consensus that replication is one of the more fundamental (Schmidt, 2009). In other words, a phenomenon may be considered real or very probable when it has been observed many times and preferably by different people or research groups. Whereas a failure to replicate is quite expected in the case of conceptual replication, or when the experimental procedure or materials entail relevant modifications, a failure in the case of an exact or quasi-exact replication, give rise to serious concerns about the reality of the phenomenon under investigation. This is the case in the four phenomena used as examples in this paper, namely, semantic subliminal priming, incubation effects on problem solving, unconscious thought, and non-local perception (NLP; e.g., Kennedy, 2001; Pratte and Rouder, 2009; Waroquier et al., 2009).

The focus of this paper is to demonstrate that for all phenomena with a moderate or small effect size (ES), approximately below 0.5 if we refer to standardized differences such as Cohen's *d*, the typical study shows a power level insufficient to detect the phenomenon under investigation.

Given that the majority of statistical analyses are based on the Null Hypothesis Statistical Testing (NHST) frequentist approach, their failure is determined by the rejection of the (nil) null hypothesis  $H_0$ , usually setting  $\alpha < 0.05$ . Even if this procedure is considered incorrect because the frequentist approach only supports  $H_0$

rejection and not  $H_0$  validity<sup>1</sup>, it may be tolerated if there is proof of a high level of statistical power as recommended in the recent APA statistical recommendations [American Psychological Association, APA (2010)]: “Power: When applying inferential statistics, take seriously the statistical power considerations associated with the tests of hypotheses. Such considerations relate to the likelihood of correctly rejecting the tested hypotheses, given a particular alpha level, ES, and sample size. In that regard, routinely provide evidence that the study has sufficient power to detect effects of substantive interest. Be similarly careful in discussing the role played by sample size in cases in which not rejecting the null hypothesis is desirable (i.e., when one wishes to argue that there are no differences), when testing various assumptions underlying the statistical model adopted (e.g., normality, homogeneity of variance, homogeneity of regression), and in model fitting. pag. 30.”

### HOW MUCH POWER?

Statistical power depends on three classes of parameters: (1) the significance level (i.e., the Type I error probability) of the test, (2) the size(s) of the sample(s) used for the test, and (3) an ES parameter defining  $H_1$  and thus indexing the degree of deviation from  $H_0$  in the underlying population.

Power analysis should be used prospectively to calculate the minimum sample size required so that one can reasonably detect an effect of a given size. Power analysis can also be used to calculate

<sup>1</sup>The line of reasoning from “the null hypothesis is false” to “the theory is therefore true” involves the logical fallacy of affirming the consequent: “If the theory is true, the null hypothesis will prove to be false. The null hypothesis proved to be false; therefore, the theory must be true” (Nickerson, 2000).

the minimum ES that is likely to be detected in a study using a given sample size.

In most experimental designs, the accepted probability of making a Type I error is  $\alpha = 0.05$  and the desired power is not less than 0.80. However, in order to define how to obtain such a level of power, it is necessary to know the ES of the phenomena being identified. It is intuitive that the smaller the phenomenon, the greater should be the means to detect it. This analogy is similar to the signal/noise relationship. The smaller the signal, the stronger must be the means to detect it in the noise. In psychological experiments, these means are the number of participants taking part in the study and the number of trials they are requested to perform.

Given that  $\text{Power} = 1 - \beta = \text{ES}^* \sqrt{N/SD}^* \alpha$ , if we know the estimated ES of a phenomenon, after the definition of the desired power and the  $\alpha$  level, the only free parameters is  $N$ , that is the number of participants or trials.

A review of 322 meta-analyses published before 1998, summarizing 25,000 studies referred to 474 social psychological effects, reports that the mean ES reported is  $r = 0.21$  and the mode was less than  $r = 0.10$  (Richard et al., 2003). For this observed mean ES, to obtain a statistical power for independent sample  $t$  tests  $> 0.9$ , the sample size for each group should be at least 90, a number rarely observed in the studies.

Setting aside the strong criticisms of the use of NHST (Cohen, 1994; Kline, 2004), a neglected aspect of this approach, is the control of how much statistical power is necessary to detect what the researcher aims to find<sup>2</sup>.

This problem is not new and has already been raised by Sedlmeier and Gigerenzer (1989), Cohen (1992), Bezeau and Graves (2001), and Maxwell (2004) among others. However the widespread adherence to “The Null Ritual” as discussed by Gigerenzer et al. (2004), which consists in: (a) Set up a statistical null hypothesis of “no mean difference” or “zero correlation.” (b) Don’t specify the predictions of your research hypothesis or of any alternative substantive hypotheses; (c) Use 5% as a convention for rejecting the null; (d) If significant, accept your research hypothesis (e) Always perform this procedure, seems to prevent most researchers taking into account such fundamental statistical parameter. In Gigerenzer et al. (2004) check, covering the years 2000–2002, and encompassing with some 220 empirical articles, only nine researchers who computed the power of their tests were found.

Using the results of four recent meta-analyses related to “controversial” or “elusive” psychological phenomena, we illustrate the importance of using the available ESs to derive the appropriate number of participants to achieve a power  $> 0.90$ . Only if a replication fails with this level of power, it is legitimate to raise doubts about the reality of the phenomena under investigation.

### CAN SUBLIMINALLY PRESENTED INFORMATION INFLUENCE BEHAVIOR?

This question is one of the most controversial questions in psychology and it remains an intriguing and strongly debated issue. Apart from the controversies relating to the controls that information

(priming) was effectively masked and the identification of serious methodological flaws which caused great doubt as to the existence of subliminal processing, the debate is even hotter around the topic of the level of influence of this unconscious (subliminal) information. Hitherto, the debate as to whether subliminal priming reflects genuine semantic processing of the subliminal information or the formation of automatic S–R mappings remains unresolved.

The meta-analysis of Van den Bussche et al. (2009) tried to shed light on these questions by analyzing all the available literature between 1983 and December 2006. Their analysis was carried out separately from the two most studied protocols, subliminal priming for semantic categorization and subliminal priming for lexical decision and naming. If semantic subliminal priming facilitated the subsequent categorization of targets belonging to the same semantic category, it suggests that the primes were unconsciously categorized and processed semantically. The same effect is postulated if lexical decision and naming are faster or more accurate to semantically subliminal related prime–target pairs than to unrelated pairs. A naming task is similar to the lexical decision task except that the targets are all words, and participants are asked to name the targets aloud.

The synthesis of the main results is reported in **Table 1**.

### DOES INCUBATION ENHANCE PROBLEM SOLVING?

The “incubation period” is the temporary shift away from an unsolved problem in order to allow a solution to emerge in the mind of the individual, seemingly with no additional effort, after he or she has put the problem aside for a period of time, having failed in initial attempts to solve it.

Among the questions under debate there is the problem of whether the nature of the discovery of the solution is really unconscious and if it is qualitatively different from that used to tackle problems that do not require such insight.

The meta-analysis of Sio and Ormerod (2009) tried to shed light on this and other related questions, analyzing all the available literature from 1964 to 2007. The main findings of their meta-analysis are reported in **Table 1**.

### UNCONSCIOUS THOUGHT THEORY

The key assumption of Unconscious Thought Theory (UTT, Dijksterhuis et al., 2006) is that unconscious thought and conscious thought are characterized by different processes. That is, “unconscious thought” processes have a relatively large capacity – hence, they allow for an optimal decision strategy in which all attributes of chosen alternatives are weighted according to their importance. These unconscious processes require time, therefore the quality of decisions increases with the duration of unconscious thought. “Conscious thought” processes on the other hand, have a small capacity and therefore only allow for simplified decision making strategies. As summarized by Dijksterhuis et al., 2006, p. 105): “When a decision strategy warrants the careful and strict application of one specific rule, as in a lexicographic strategy, use conscious thought. When matters become more complicated and weighting is called for, as in the weighting strategy, use unconscious thought.”

As expected, among the questions raised by this theory, a critical problem is whether really optimal decisions are obtained after

<sup>2</sup>Even if power estimate is important in meta-analyses (i.e., Valentine et al., 2009), in this paper we focus only on power estimates in single studies.

**Table 1 | Descriptive statistics of the four meta-analyses, related to Unconscious Semantic Priming, Incubation effect, UTT, and NLP with the estimated power of a typical study, the mean and 95% CI power calculated from all studies included in the meta-analysis and the number of participants necessary to obtain a Power = 0.90.**

Phenomena	Protocols	Source	<i>N</i> studies	Averaged <i>N</i> × study (range)	ES* (mean and 95% CI)	<i>p</i>	Estimated power of a typical study	Observed power (mean and 95% CI)	Estimated <i>N</i> to achieve power = 0.90
Semantic priming	Semantic categorization	Van den Bussche et al. (2009)	23	20 (6–80)	0.80 ± 0.2	n.a.	0.96	0.90 (±1.7)	
	Lexical and naming	Van den Bussche et al. (2009)	32	33 (9–132)	0.47 ± 0.11	n.a.	0.84	0.69 (±0.4.6)	40
Incubation effect		Sio and Ormerod (2009)	117	31 (7–278)	0.29 ± 0.09	n.a.	0.21	0.43 (±3.3)	100
Unconscious thought theory		Strick et al. (2011)	92	26 (14–55)	0.22 ± 0.08	1.2 × 10 <sup>-8</sup>	0.20§	0.19 (±0.72)	400§
Non-local perception	Remote Vision	Milton (1997)	78	34 (1–74)	0.16 ± 0.06	6 × 10 <sup>-9</sup>	0.55	n.a.	73
	Ganzfeld	Storm et al. (2010)	108	40 (7–128)	0.13 ± 0.04	8.3 × 10 <sup>-11</sup>	0.46	0.45 (±3.8)	110
	Forced-choice with normal state of consciousness	Storm et al. (in press)	72	128 (12–887)	0.011 ± 0.004	5.3 × 10 <sup>-7</sup>	0.07	0.065 (±0.81)	3450

\*Random effect Cohen's *d*; §two groups comparison.

a period of distraction from deliberate conscious mental activity for the same amount of time as would be the case had the decisions been taken deliberately.

The meta-analysis of Strick et al. (2011), aimed to give an answer to this and other related questions by analyzing all the available evidence up to May 2011. The main findings are reported in **Table 1**.

### NON-LOCAL PERCEPTION

Non-local perception (NLP) is based on the hypothesis that the human mind may have quantum like properties, that is, that some of its functions, such as perceptual abilities, reasoning, etc., may be analyzed using quantum formalism.

The main non-local properties which are studied within the realm of quantum physics and which are supported by “extraordinary evidence” (see Genovese, 2005, 2010), are “entanglement” and “measurement interference.” The first property, entanglement, allows two or more physical objects to behave as one even if they are separated in space and time. This “strange” property allows a form of immediate communication of the objects’ characteristics over distances between or among the entangled objects, as has been observed in teleportation experiments (i.e., Bouwmeester et al., 1997). The possibility that quantum-like properties may be observed not only in physics but also even in biology and psychology has not only been studied theoretically (von Lucadou et al., 2007; Khrennikov, 2010; Walach and von Stillfried, 2011) but also experimentally (see Gutiérrez et al., 2010 for biology and Busemeyer et al., 2011, for psychology).

One of the main concerns about the studies related to different aspects of NLP, is their inconsistency in obtaining results satisfying the statistical cut off criteria to refute the null hypothesis that extra sensory perception does not exist, usually setting  $\alpha < 0.05$ .

This problem is recognized by some researchers involved in this field of research (Kennedy, 2001) and even more by all deniers of NLP evidence (e.g., Alcock, 2003).

The main findings of three meta-analysis related to three NLP different protocols, Remote Vision, that is NLP using a free-choice response (Milton, 1997), NLP in a Ganzfeld state using free-choice response (Storm et al., 2010) and NLP in a normal state of consciousness using a forced-choice response (Storm et al., in press), covering all evidence available from 1964 to 1992, 1974 to 2009, and 1987 to 2010 respectively, are reported in **Table 1**

### POWER ESTIMATION

A synthesis of the descriptive statistics related to the four phenomena described above is presented in **Table 1** in decreasing order of magnitude of ESs. For each meta-analysis, the retrospective statistical power with  $\alpha = 0.05$ , achieved by a typical study using the mean of the number of participants of all studies included in the meta-analysis was estimated.

For all but one meta-analysis, it was also possible to calculate the mean and 95% CI *post hoc* power, using the number of participants of each study included in the meta-analyses and the estimated random ES, setting  $\alpha = 0.05$ .

Furthermore, for each of the four psychological phenomena, the number of participants necessary to obtain a statistical

power = 0.9 with  $\alpha = 0.05$  given the observed random ESs, was estimated.

Statistical power was calculated using the software G\*Power (Faul et al., 2007).

## COMMENT

The results are quite clear: apart from the unconscious semantic priming for semantic categorization, where the number of participants in a typical experiment is sufficient to obtain a statistical power above 0.90, for all remaining phenomena, to achieve this level of power, it is necessary to increase the number of participants in a typical study, from a minimum of seven participants for the unconscious semantic priming for lexical decision and naming to around 3400 to investigate NLP using the forced-choice with normal state of consciousness protocol.

## GENERAL DISCUSSION

The response to the question posed in the introduction, as to whether there are elusive phenomena or an elusive power to detect them, is quite clear. If there are clear estimates of ESs from the evidence of the phenomenon derived from a sufficient number of studies analyzed meta-analytically and their values are moderate or low, it is mandatory to increase the number of participants to achieve a statistical power of 0.90, with the inevitable consequence of investing more time and money into each study before interpreting the results as support for reality or unreality of a phenomenon.

Are there alternatives to this obligation? Yes, and we briefly illustrate some of these, also providing references for those interested in using them.

## CONFIDENCE INTERVALS

In line with the statistical reform movement (i.e., Cumming, 2012), in the APA manual (American Psychological Association, APA, 2010), there are the following statistical recommendations “Alternatively, (to the use of NHST) use calculations based on a chosen target precision (confidence interval width) to determine sample sizes. Use the resulting confidence intervals to justify conclusions concerning ESs (e.g., that some effect is negligibly small) p. 30.”

## EQUIVALENCE TESTING

Equivalence tests are inferential statistics designed to provide evidence for a null hypothesis. Like effect tests, the nil-null is eschewed in equivalence testing. However unlike standard NHST, equivalence tests provide evidence that there is little difference or effect. A significant result in an equivalence test means that

## REFERENCES

- Alcock, J. E. (2003). Give the null hypothesis a chance: reasons to remain doubtful about the existence of PSI. *J. Conscious. Stud.* 10, 29–50.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association*, 6th Edn, Washington, DC: American Psychological Association.
- Bezeau, S., and Graves, R. (2001). Statistical power and effect sizes of clinical neuropsychology research. *J. Clin. Exp. Neuropsychol.* 23, 399–406.
- Bouwmeester, D., Pan, J. W., Mattle, K., Eibl, M., Weinfurter, H., and Zeilinger, A. (1997). Experimental quantum teleportation. *Nature* 390, 575–579.
- Busemeyer, J. R., Pothos, E. M., Franco, R., and Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychol. Rev.* 118, 2, 193–218.
- Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 1, 155–159.
- Cohen, J. (1994). The earth is round ( $p < .085$ ). *Am. Psychol.* 49, 997–1003.
- Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York: Routledge.
- Dijksterhuis, A., Bos, M. W., Nordgren, L. F., and Van Baaren, R. B. (2006). On making the right choice: the deliberation-without attention effect. *Science* 311, 1005–1007.
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191.
- Genovese, M. (2005). Research on hidden variable theories, a review of recent progresses. *Phys. Rep.* 413, 319–396.

the hypothesis that the effects or differences are substantial can be rejected. Hence, equivalence tests are appropriate when researchers want to show little difference or effect (Levine et al., 2008).

## EVALUATING INFORMATIVE HYPOTHESES

Evaluating specific expectations directly produces more useful results than sequentially testing traditional null hypotheses against catch-all rivals. Researchers are often interested in the evaluation of informative hypotheses and already know that the traditional null hypothesis is an unrealistic hypothesis. This presupposes that prior knowledge is often available; if this is not the case, testing the traditional null hypothesis is appropriate. In most applied studies, however, prior knowledge is indeed available in the form of specific expectations about the ordering of statistical parameters (Kuiper and Hoijsink, 2010; Van de Schoot et al., 2011).

## BAYESIAN APPROACH

Another alternative is to abandon the frequentist approach and use a Bayesian one (Wagenmakers et al., 2011). With a Bayesian approach the problem of statistical power is substituted with parameter estimation and/or model comparison (Kruschke, 2011). In the first approach, assessing null values, the analyst simply sets up a range of candidate values, including the null value, and uses Bayesian inference to compute the relative credibility of all the candidate values. In the model comparison approach, the analyst sets up two competing models of what values are possible. One model posits that only the null value is possible whereas the alternative model posits that a broad range of other values is also possible. Bayesian inference is used to compute which model is more credible, given the data.

## FINAL COMMENT

Is there a chance to abandon “The Null Ritual” in the near future and to think of science as cumulative knowledge? The answer is “yes” if we approach scientific discovery thinking meta-analytically (Cumming, 2012), that is, simply reporting observed (standardized) ES and the corresponding confidence intervals, both when NHST is refuted and when it is not refuted (Nickerson, 2000; American Psychological Association, APA, 2010) without drawing dichotomous decisions. The statistical approaches listed above are good tools to achieve this goal.

How many editors and reviewers are committed to pursuing it?

## ACKNOWLEDGMENTS

Suggestions and comments by the reviewers were greatly appreciated for improving the clarity and quality of the paper. Proof Reading Service revised the English.

- Genovese, M. (2010). Interpretations of quantum mechanics and measurement problem. *Adv. Sci. Lett.* 3, 249–258.
- Gigerenzer, G., Krauss, S., and Vitouch, O. (2004). “The null ritual what you always wanted to know about significance testing but were afraid to ask,” in *The Sage Handbook of Quantitative Methodology for the Social Sciences*, ed. D. Kaplan (Thousand Oaks, CA: Sage), 391–408.
- Gutiérrez, R., Caetano, R., Woiczikowski, P. B., Kubar, T., Elstner, M., and Cuniberti, G. (2010). Structural fluctuations and quantum transport through DNA molecular wires, a combined molecular dynamics and model Hamiltonian approach. *New J. Phys.* 12, 023022.
- Kennedy, J. E. (2001). Why is PSY so elusive? A review and proposed model. *J. Parapsychol.* 65, 219–246.
- Khrennikov, A. Y. (2010). *Ubiquitous Quantum Structure from Psychology to Finance*. Berlin: Springer-Verlag.
- Kline, R. B. (2004). *Beyond Significance Testing. Reforming Data Analysis Methods in Behavioral Research*. Washington, DC: APA.
- Kruschke, J. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspect. Psychol. Sci.* 6, 299–312.
- Kuiper, R. M., and Hoijtink, H. (2010). Comparisons of means using exploratory and confirmatory approaches. *Psychol. Methods* 15, 69–86.
- Levine, T. R., Weber, R., Sun Park, H., and Hullett, C. R. (2008). A communication researchers’ guide to null hypothesis significance testing and alternatives. *Hum. Commun. Res.* 34, 188–209.
- Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychol. Methods* 9, 147–163.
- Milton, J. (1997). Meta-analysis of free-response ESP studies without altered states of consciousness. *J. Parapsychol.* 61, 279–319.
- Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychol. Methods* 5, 241–301.
- Pratte, M. S., and Rouder, J. N. (2009). A task-difficulty artifact in subliminal priming. *Atten. Percept. Psychophys.* 71, 1276–1283.
- Richard, F. D., Bond, C. E., and Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Rev. Gen. Psychol.* 7, 331–363.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Rev. Gen. Psychol.* 13, 90–100.
- Sedlmeier, P., and Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychol. Bull.* 105, 309–316.
- Sio, U. N., and Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychol. Bull.* 135, 94–120.
- Storm, L., Tressoldi, P. E., and Di Risio, L. (2010). Meta-analysis of free-response studies, 1992–2008: assessing the noise reduction model in parapsychology. *Psychol. Bull.* 136, 471–485.
- Storm, L., Tressoldi, P. E., and Di Risio, L. (in press). Meta-analysis of ESP studies, 1987–2010, assessing the success of the forced-choice design in parapsychology.
- Strick, M., Dijksterhuis, A., Bos, M. W., Sjoerdsma, A., van Baaren, R. B., and Nordgren, L. F. (2011). A meta-analysis on unconscious thought effects. *Soc. Cogn.* 29, 738–762.
- Valentine, J. C., Pigott, T. D., and Rothstein, H. R. (2009). How many studies to you need? A primer on statistical power for meta-analysis. *J. Educ. Behav. Stat.* 35, 215–247.
- Van de Schoot, R., Hoijtink, H., and Jan-Willem, R. (2011). Moving beyond traditional null hypothesis testing: evaluating expectations directly. *Front. Psychol.* 2:24. doi:10.3389/fpsyg.2011.00024
- Van den Bussche, E., den Noortgate, W., and Reynvoet, B. (2009). Mechanisms of masked priming: a meta-analysis. *Psychol. Bull.* 135, 452–477.
- von Lucadou, W., Römer, H., and WAlach, H. (2007). Synchronistic phenomena as entanglement correlations in generalized quantum theory. *J. Conscious. Stud.* 14, 4, 50–74.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., and Van der Maas, H. (2011). Why psychologists must change the way they analyze their data: the case of psi. *J. Pers. Soc. Psychol.* 100, 426–432.
- Walach, H., and von Stillfried, N. (2011). Generalised quantum theory. Basic idea and general intuition: a background story and overview. *Axiomathes* 21, 185–209.
- Waroquier, L., Marchiori, D., Klein, O., and Cleeremans, A. (2009). Methodological pitfalls of the unconscious thought paradigm. *Judgm. Decis. Mak.* 4, 601–610.

**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.

Received: 13 April 2012; accepted: 12 June 2012; published online: 04 July 2012.

Citation: Tressoldi PE (2012) Replication unreliability in psychology: elusive phenomena or “elusive” statistical power? *Front. Psychology* 3:218. doi: 10.3389/fpsyg.2012.00218

This article was submitted to *Frontiers in Quantitative Psychology and Measurement*, a specialty of *Frontiers in Psychology*.

Copyright © 2012 Tressoldi. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.