

Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?

Perspectives on Psychological Science
7(6) 528–530

© The Author(s) 2012

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/1745691612465253

http://pps.sagepub.com



Harold Pashler¹ and Eric-Jan Wagenmakers²

¹University of California, San Diego and ²University of Amsterdam, The Netherlands

Is there currently a crisis of confidence in psychological science reflecting an unprecedented level of doubt among practitioners about the reliability of research findings in the field? It would certainly appear that there is. These doubts emerged and grew as a series of unhappy events unfolded in 2011: the Diederik Stapel fraud case (see Stroebe, Postmes, & Spears, 2012, this issue), the publication in a major social psychology journal of an article purporting to show evidence of extrasensory perception (Bem, 2011) followed by widespread public mockery (see Galak, LeBoeuf, Nelson, & Simmons, in press; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011), reports by Wicherts and colleagues that psychologists are often unwilling or unable to share their published data for reanalysis (Wicherts, Bakker, & Molenaar, 2011; see also Wicherts, Borsboom, Kats, & Molenaar, 2006), and the publication of an important article in *Psychological Science* showing how easily researchers can, in the absence of any real effects, nonetheless obtain statistically significant differences through various questionable research practices (QRPs) such as exploring multiple dependent variables or covariates and only reporting these when they yield significant results (Simmons, Nelson, & Simonsohn, 2011).

For those psychologists who expected that the embarrassments of 2011 would soon recede into memory, 2012 offered instead a quick plunge from bad to worse, with new indications of outright fraud in the field of social cognition (Simonsohn, 2012), an article in *Psychological Science* showing that many psychologists admit to engaging in at least some of the QRPs examined by Simmons and colleagues (John, Loewenstein, & Prelec, 2012), troubling new meta-analytic evidence suggesting that the QRPs described by Simmons and colleagues may even be leaving telltale signs visible in the distribution of *p* values in the psychological literature (Masicampo & Lalande, in press; Simonsohn, 2012), and an acrimonious dust-up in science magazines and blogs centered around the problems some investigators were having in replicating well-known results from the field of social cognition (Bower, 2012; Yong, 2012).

Although the very public problems experienced by psychology over this 2-year period are embarrassing to those of us working in the field, some have found comfort in the fact that,

over the same period, similar concerns have been arising across the scientific landscape (triggered by revelations that will be described shortly). Some of the suspected causes of unreplicability, such as publication bias (the tendency to publish only positive findings) have been discussed for years; in fact, the phrase *file-drawer problem* was first coined by a distinguished psychologist several decades ago (Rosenthal, 1979). However, many have speculated that these problems have been exacerbated in recent years as academia reaps the harvest of a hypercompetitive academic climate and an incentive scheme that provides rich rewards for overselling one's work and few rewards at all for caution and circumspection (see Giner-Sorolla, 2012, this issue). Equally disturbing, investigators seem to be replicating each others' work even less often than they did in the past, again presumably reflecting an incentive scheme gone askew (a point discussed in several articles in this issue, e.g., Makel, Plucker, & Hegarty, 2012).

The frequency with which errors appear in the psychological literature is not presently known, but a number of facts suggest it might be disturbingly high. Ioannidis (2005) has shown through simple mathematical modeling that any scientific field that ignores replication can easily come to the miserable state wherein (as the title of his most famous article puts it) "most published research findings are false" (see also Ioannidis, 2012, this issue, and Pashler & Harris, 2012, this issue). Meanwhile, reports emerging from cancer research have made such grim scenarios seem more plausible: In 2012, several large pharmaceutical companies revealed that their efforts to replicate exciting preclinical findings from published academic studies in cancer biology were only rarely verifying the original results (Begley & Ellis, 2012; see also Oshervovich, 2011; Prinz, Schlange, & Asadullah, 2011).

Corresponding Author:

Hal Pashler, University of California, San Diego, Department of Psychology, 9500 Gilman Drive #0109, La Jolla, CA 92093-0109
E-mail: hpashler@ucsd.edu

Closer to home, the replicability of published findings in psychology may become clearer with the Reproducibility Project (Open Science Collaboration, 2012, this issue; see also Carpenter, 2012). Individuals and small groups of service-minded psychologists are each contributing their time to conducting a replication of a published result following a structured protocol. The aggregated results will provide the first empirical evidence of reproducibility and its predictors. The open project is still accepting volunteers. With small contributions from many of us, the Reproducibility Project will provide an empirical basis for assessing our reproducibility as a field (to find out more, or sign up yourself, visit: <http://openscienceframework.org/project/EZcUj/>).

This special section brings together a set of articles that analyze the causes and extent of the replicability problems in psychology and ask what can be done about it. The first nine articles focus principally on diagnosis; the following six articles focus principally on treatment. Those readers who need further motivation to change their research practices are referred to the illustration provided by Neuroskeptic (2012). The section ends with a stimulating overview by John Ioannidis, the biostatistician whose work has led the way in exposing problems of replicability and bias across the fields of medicine and the life sciences.

Many of the articles in this special issue make it clear why the replicability problems will not be so easily overcome, as they reflect deep-seated human biases and well-entrenched incentives that shape the behavior of individuals and institutions. Nevertheless, the problems are surely not insurmountable, and the contributors to this special section offer a great variety of ideas for how practices can be improved.

In the opinion of the editors of this special section, it would be a mistake to try to rely upon any single solution to such a complex problem. Rather, it seems to us that psychological science should be instituting parallel reforms across the whole range of academic practices—from journals and journal reviewing to academic reward structures to research practices within individual labs—and finding out which of these prove effective and which do not. We hope that the articles in this special section will not only be stimulating and pleasurable to read, but that they will also promote much wider discussion and, ultimately, collective actions that we can take to make our science more reliable and more reputable. Having found ourselves in the very unwelcome position of being (to some degree at least) the public face for the replicability problems of science in the early 21st century, psychological science has the opportunity to rise to the occasion and provide leadership in finding better ways to overcome bias and error in science generally.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

References

- Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature*, 483, 531–533.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 407–425.
- Bower, B. (2012). The hot and the cold of priming. *Science News*, 181, 26.
- Carpenter, S. (2012). Psychology's bold initiative. *Science*, 335, 1558–1560.
- Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (in press). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology*.
- Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, 7, 562–571.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2, 696–701.
- Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7, 645–654.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science*, 23, 524–532.
- Makel, M., Plucker, J., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7, 537–542.
- Masicampo, E. J., & Lalande, D. R. (in press). A peculiar prevalence of *p* values just below .05. *Quarterly Journal of Experimental Psychology*.
- Neuroskeptic. (2012). The nine circles of scientific hell. *Perspectives on Psychological Science*, 7, 643–644.
- Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7, 657–660.
- Osherovich, L. (2011). Hedging against academic risk. *Science-Business eXchange*, 4(15). doi:10.1038/scibx.2011.416.
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives in Psychological Science*, 7, 531–536.
- Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10, 712–713.
- Rosenthal, R. (1979). An introduction to the file drawer problem. *Psychological Bulletin*, 86, 638–641.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.
- Simonsohn, U. (2012). *Just post it: The lesson from two cases of fabricated data detected by statistics alone*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2114571
- Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, 7, 670–688.

- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology*, 100, 426–432.
- Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, 6, e26828. doi:10.1371/journal.pone.0026828
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61, 726–728.
- Yong, E. (2012). A failed replication attempt draws a scathing personal attack from a psychology professor. *Discover Magazine*. Retrieved from <http://blogs.discovermagazine.com/notrocket-science/2012/03/10/failed-replication-barghpsychology-study-doyen/>