

Swan Song Editorial

Psychological Science
2019, Vol. 30(12) 1669–1673
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0956797619893653
www.psychologicalscience.org/PS



Early in 2012, Geoff Cumming blew my mind with a talk that led me to realize that I had been conducting underpowered experiments for decades. In some lines of research in my lab, a predicted effect would come booming through in one experiment but melt away in the next. My students and I kept trying to find conditions that yielded consistent statistical significance—tweaking items, instructions, exclusion rules—but we sometimes eventually threw in the towel because results were mad-deningly inconsistent. For example, a chapter by Lindsay and Kantner (2011) reported 16 experiments with an on-again/off-again effect of feedback on recognition memory. Cumming's talk explained that p values are very noisy. Moreover, when between-subjects designs are used to study small- to medium-sized effects, statistical tests often yield nonsignificant outcomes (sometimes with huge p values) unless samples are very large. For example, if Cohen's d equals 0.50 for a between-subjects comparison and there are 20 subjects in each group, then about two thirds of the time, p will be greater than .05, a Type II error. Even if the sample size is 50 per condition, nearly one in three between-subjects experiments with an effect size (d) of 0.50 would yield nonsignificant results. Experiments in my lab were trying to detect a small mixed-model interaction with two or three dozen subjects per group. No wonder our results were inconsistent.

Hard on the heels of Cumming's talk, I read Simmons, Nelson, and Simonsohn's (2011) "False-Positive Psychology" article, published in *Psychological Science*. Then I gobbled up several articles and blog posts on misuses of null-hypothesis significance testing (NHST). The authors of these works make a convincing case that hypothesizing after the results are known (HARKing; Kerr, 1998) and other forms of " p hacking" (post hoc exclusions, transformations, addition of moderators, optional stopping, publication bias, etc.) are deeply problematic. Such practices are common in some areas of scientific psychology, as well as in some other life sciences. These practices sometimes give rise to mistaken beliefs in effects that really do not exist. Combined with publication bias, they often lead to exaggerated estimates of the sizes of real but small effects. For a review, see Nelson, Simmons, and Simonsohn (2018).

In 2013, I became an Associate Editor of *Psychological Science* under Editor in Chief Eric Eich. I enthusiastically supported Eich's groundbreaking work to improve statistical standards. Eich—along with Alan Kraut, then Executive Director of the Association for Psychological Science (APS), and Bobbie Spellman, then Editor in Chief of *Perspectives in Psychological Science*—contributed to the development of the Transparency and Openness Promotion guidelines (<https://cos.io/top/>), to which APS was an original signatory. Eich instituted the use of badges designating articles with open data, open materials, and preregistered research plans. He required authors to aver that they had disclosed all data exclusions, manipulations, and measures and how they had determined their studies' sample sizes. And he removed word-count limits from Method and Results sections so that authors had space to report the key details of their studies, analyses, and findings. Eich also encouraged authors to shift from NHST to an emphasis on estimation of effect sizes. Eich's (2014) editorial, "Business Not as Usual," still makes good reading.

Since becoming Acting Editor in July 2015 (and Editor in Chief in January 2016), I have extended Eric Eich's initiatives. As discussed in my 2015 editorial, I encouraged authors to provide sound rationales for their studies' sample sizes and to use graphs that reveal the shapes of data distributions. I endeavored to discourage misinterpretations of nonsignificant effects in low-powered hypothesis tests. I instituted the use of statcheck (<http://statcheck.io/>) to detect errors in reporting of common inferential tests. Articles in *Psychological Science* now name the Action Editor; this credits editors and holds them accountable for their work. As of 2018, each empirical article includes an Open Practices statement that addresses the availability of data and materials and whether or not any of the studies reported in the article were preregistered. I have also asked authors to address constraints on the generality of the findings they report (Simons, Shoda, & Lindsay, 2017).

In 2017, I added Preregistered Direct Replications (PDRs) to the *Psychological Science* submission categories. As explained in my 2017 editorial, PDRs report attempts to recreate the essential conditions and measures of a study originally published in *Psychological*

Science. Authors are encouraged to submit PDRs following the Registered Report model, meaning that the initial submission is a proposal that is sent for review before the study is conducted (see <https://cos.io/rr/>). The addition of PDRs was inspired by Sanjay Srivastava's (2012) argument that journals should take responsibility for articles they have published by being open to publishing direct replication attempts (successful or not). Only a few PDRs have been published in *Psychological Science* to date. More are in the pipeline.

Psychological Science in recent years published several articles emphasizing evidence for the absence of a nontrivial effect—that is, evidence supporting the null. Statisticians are still debating the fine points, but there is wide agreement that Bayes factor or equivalence tests can be used to assess the strength of evidence for the null. This, I believe, is an important step forward for life science.

APS generously supported the appointment of six Statistical Advisors during my term. These individuals served in a variety of ways to improve the statistical rigor and clarity of the works published in *Psychological Science*. The Statistical Advisers also provided crucial support when questions emerged regarding statistical matters in articles previously published in *Psychological Science*.

During my term, *Psychological Science* issued many Corrigenda (corrections of errors made by authors) and several Retractions (withdrawals of articles because of errors so fundamental that they vitiated the value of the work). Several of these have been issued at the behest of authors who stepped forward and requested the Corrigendum or Retraction themselves. We all make mistakes, and I believe that researchers should be saluted for owning up to them when appropriate. In other cases, errors came to light because authors had posted their data, and other scientists, examining them, found evidence of problems. That too is progress.

Did the changes that Eric Eich and I instituted (and that our Associate and Senior Editors implemented) improve *Psychological Science*? It is hard to say. Submissions rates have declined over our editorships (from a peak of 2,700 in 2011 to about 1,700 per year in the last 3 years). Perhaps the new standards (and the added work involved in completing the online submission form) selectively discouraged scientifically weaker submissions, but they probably also cost us some gems. The number of articles published per year has decreased, offset by an increase in average article length (e.g., mean length was 6.6 pages in 2010 and 11.2 pages in 2018). *Psychological Science* became hugely successful in part because it published short, snappy, easy-to-read articles with surprising findings. Now the journal's articles are longer and include more technical and statistical detail, and their take-home messages are more

nuanced. Journal impact factor (JIF) increased monotonically from 2011 to 2017 but declined in 2018; JIF is subject to a host of critiques, but no editor likes to see it dip.

Happily, there are other indicators that the reforms Eric Eich and I implemented are having the intended effects. Figure 1 is from Kidwell et al. (2016). It shows, for *Psychological Science* and four comparison journals, the percentage of articles published between 2012 and 2015 in which the authors claim that readers can directly access the data underlying the reported effect. That percentage shot up in *Psychological Science* when badges were introduced.

Figure 2, also from Kidwell et al. (2016), shows what happened when the researchers attempted to access and reproduce the analyses from articles claiming that data were directly accessible. As the figure shows, their success rate was vastly higher for *Psychological Science* than for the comparison journals. I would love to see a follow-up to this study.

Figure 3 shows the percentage of Research Articles published in *Psychological Science* that earned Open Data, Open Materials, and Preregistration badges in each year since badges were introduced. These curves are gratifying. I invite comparisons with other journals.

Badges are not an end in themselves. The aim is to make it easy for scientists to access data and materials that support assessments of reproducibility, robustness, and replicability and to encourage detailed preregistrations that enable researchers and readers to differentiate between a priori hypothesis tests and exploratory analyses (along with other good things; Lindsay, Simons, & Lilienfeld, 2016). To date, we have relied primarily on authors' attestations of badge worthiness. Providing thorough assessments of badge applications (i.e., reproducing data analyses, assessing the completeness of the materials, comparing preregistrations with the reported methods and analyses) would require a nontrivial investment of time and resources.

Many professional societies such as APS team with commercial publishers to produce and distribute journals. This enables societies to achieve very high levels of distribution and to generate income that they use to support robust peer review, provide editorial honoraria, and promote the societies' goals (e.g., underwriting the cost of conferences, advocating for our science, conducting public outreach and education efforts). But this model is under intense fire. Critics argue that commercial publishers' profits are unconscionably high, that the goals of publishers are at odds with the goals of scientists, and that the commercial publication system blocks many people from accessing reports of publicly funded research. Plan S, propounded by 11 European funding agencies, would require immediate open access to reports of research they fund (Else, 2018).

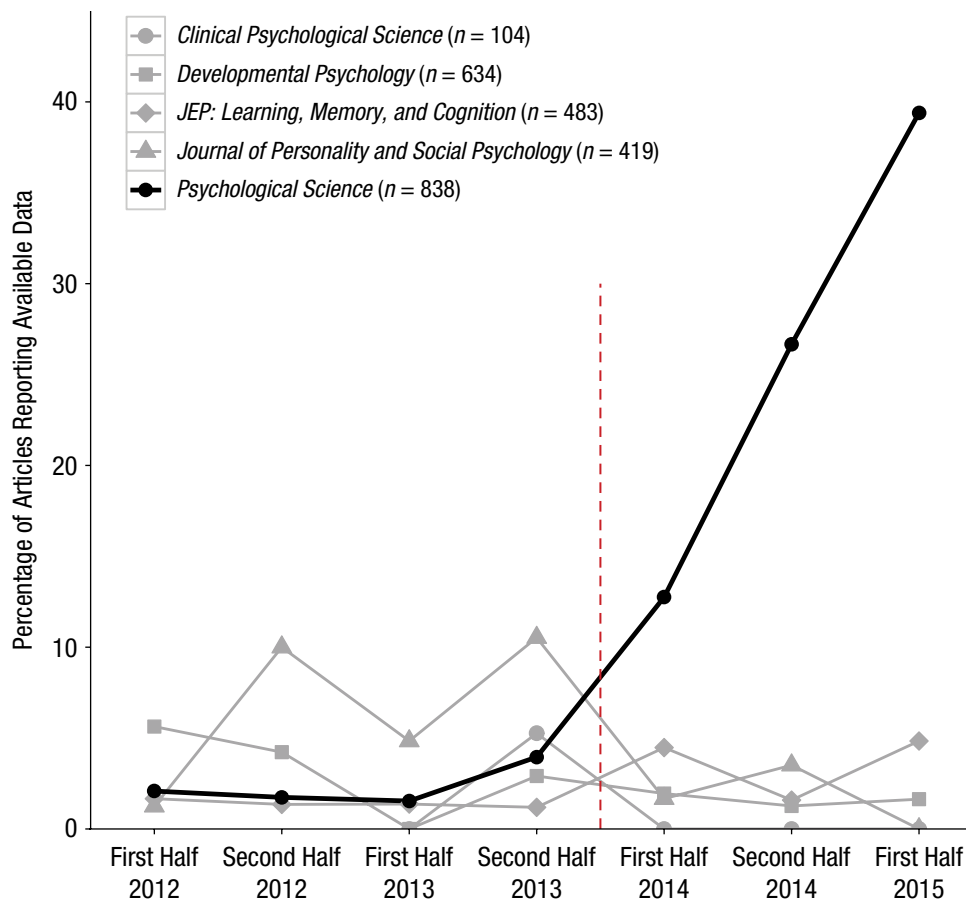


Fig. 1. Percentage of articles published in *Psychological Science* and four comparison journals stating that the data underlying the reported effect were publicly available. Results are shown for articles published from the first half of 2012 to the first half of 2015. The dashed red line indicates where *Psychological Science* introduced the Open Practices badge program. Figure adapted from Kidwell et al. (2016, Fig. 2). *JEP* = *Journal of Experimental Psychology*.

The University of California recently dropped subscriptions to Elsevier journals (McKenzie, 2019). Early career researchers increasingly rely on and extol the virtues of “preprints” on centralized repositories (e.g., PsyArXiv.com). Academic Twitter seems awash in scorn for journals that publish articles behind a paywall.

I believe in the value of curated, peer-reviewed, copy-edited journal articles. Journals can set policies that quickly and dramatically increase transparency, as shown above. It is great that professional societies use revenue from journals to support their proscience agendas. But I believe that professional societies must think carefully about how and why they publish journals and how they can continue to do so in the future. More specifically, I believe that societies such as APS should invest more resources in adding value to their publications.

Serving as Editor in Chief of *Psychological Science* has been a huge honor, challenge, and thrill. I thank the many individuals who have generously helped me

along the way. Here I can name only a few of the most crucial. Roddy Roediger (then-Chair of the APS Publications Committee) and Alan Kraut (Executive Director Emeritus of APS) were tremendously inspiring and supportive as I began my term as Acting Editor, and I thank them warmly. I am also grateful for the superb work of the many fine Senior and Associate Editors who served *Psychological Science* during my term—their diligent and thoughtful contributions were crucial. Kudos are due to the Statistical Advisors, to the members of the Editorial Board, and to the ad hoc reviewers. Fellow editors Dan Simons, Simine Vazire, and Scott Lilienfeld were models, sounding boards, and supporters. David Mellor of the Center for Open Science has been a terrific resource. I also thank former Peer Review Coordinator Kristen Medeiros and current Peer Review Coordinator Ami Getu—super responsive, good humored, and helpful. Amy Drew, Senior Manager, Publications, was always generous, patient, knowledgeable,

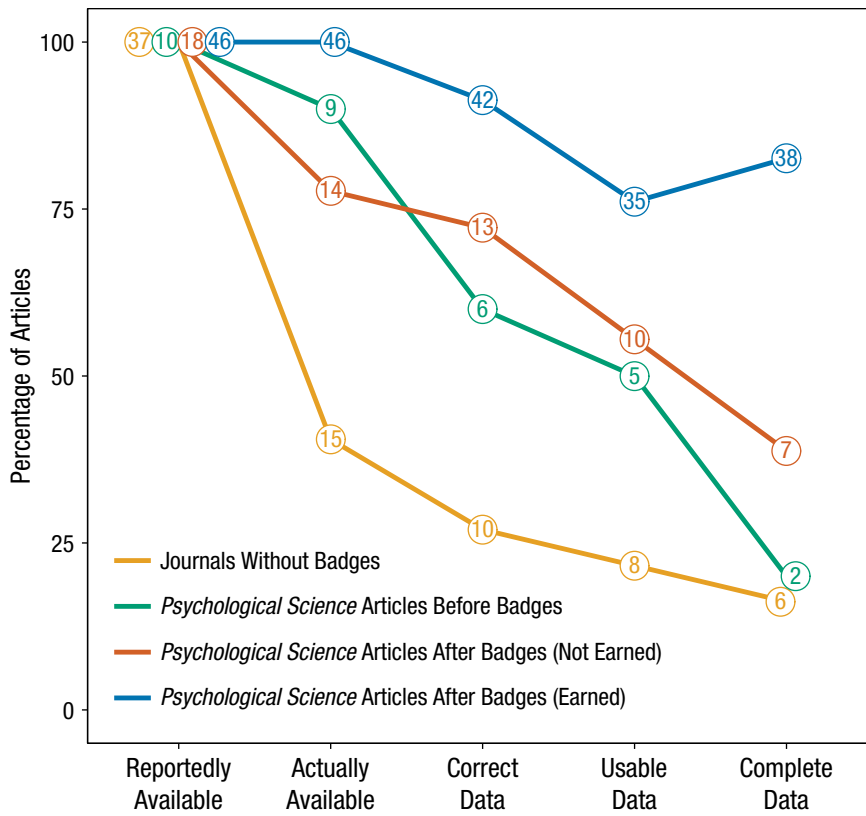


Fig. 2. Percentage of articles in which data were reportedly available at an independent archive or personal website, actually available, correct, usable, and complete. Results are shown separately for *Psychological Science* and other journals that do not participate in the badge program. For *Psychological Science*, results are broken down for articles printed before and after the journal began awarding badges, with articles printed after badge availability further broken down by whether or not the badge was earned or applied for. The numbers in circles are the total number of articles in each category. Figure adapted from Kidwell et al. (2016, Fig. 4).

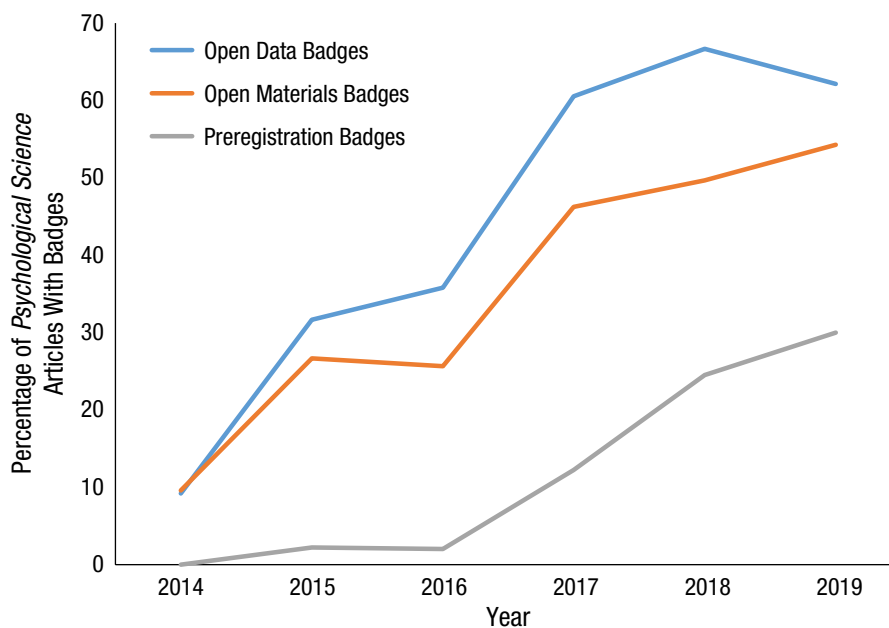


Fig. 3. Percentage of Research Articles published in *Psychological Science* that earned Open Data, Open Materials, and Preregistration badges each year since badges were introduced.

and skilled. I further thank Brian Winters and Michele Nathan for superb work as Managing Editors extraordinaire. I very much appreciate that APS Executive Director Sarah Brookhart and Publications Committee Chair Deanna Barch gave me great latitude and generous support. I close with my very best wishes to the new Editor in Chief of *Psychological Science*, Patricia Bauer. I hope that she finds the role as engaging and rewarding as I have.

—D. Stephen Lindsay

Acknowledgments

I thank Dan Simons, Geoff Cumming, Eric Eich, Roddy Roediger, Aime Ballard-Wood, Melissa Welch-Ross, and Brian Winters for feedback on a draft of this editorial.

References

- Eich, E. (2014). Business not as usual. *Psychological Science*, 25, 3–6. doi:10.1177/0956797613512465
- Else, H. (2018). Radical open-access plan could spell end to journal subscriptions. *Nature*, 561, 17–18. doi:10.1038/d41586-018-06178-7
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2, 196–217. doi:10.1207/s15327957pspr0203_4
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., . . . Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLOS Biology*, 14(5), Article e1002456. doi:10.1371/journal.pbio.1002456
- Lindsay, D. S. (2015). Replication in psychological science. *Psychological Science*, 26, 1827–1832. doi:10.1177/0956797615616374
- Lindsay, D. S. (2017). Preregistered direct replications in *Psychological Science*. *Psychological Science*, 28, 1191–1192. doi:10.1177/0956797617718802
- Lindsay, D. S., & Kantner, J. (2011). A search for influences of feedback on recognition of music, poetry, and art. In P. A. Higham & J. P. Leboe (Eds.), *Constructions of remembering and metacognition: Essays in honour of Bruce Whittlesea* (pp. 137–154). Houndmills, England: Palgrave Macmillan.
- Lindsay, D. S., Simons, D. J., & Lilienfeld, S. O. (2016). Research preregistration 101. *Observer*, 29(10), 14–17.
- McKenzie, L. (2019, March 1). UC drops Elsevier. *Inside Higher Ed*. Retrieved from <https://www.insidehighered.com/news/2019/03/01/university-california-cancels-deal-elsevier-after-months-negotiations>
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, 69, 511–534. doi:10.1146/annurev-psych-122216-011836
- 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. doi:10.1177/0956797611417632
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12, 1123–1128. doi:10.1177/1745691617708630
- Srivastava, S. (2012, September 27). A Pottery Barn rule for scientific journals [Web log post]. Retrieved from <https://thehardestscience.com/2012/09/27/a-pottery-barn-rule-for-scientific-journals/>