Open Science

Barbara A. Spellman, University of Virginia

Elizabeth A. Gilbert,

Katherine S. Corker, Grand Valley State University

Draft of: 20 September 2017

Abstract

Open science is a collection of actions designed to make scientific processes more transparent and results more accessible. Its goal is to build a more replicable and robust science; it does so using new technologies, altering incentives, and changing attitudes. The current movement toward open science was spurred, in part, by a recent series of unfortunate events within psychology and other sciences. These events include the large number of studies that have failed to replicate and the prevalence of common research and publication procedures that could explain why. Many journals and funding agencies now encourage, require, or reward some open science practices, including pre-registration, providing full materials, posting data, distinguishing between exploratory and confirmatory analyses, and running replication studies. Individuals can practice and promote open science in their many roles as researchers, authors, reviewers, editors, teachers, and members of hiring, tenure, promotion, and awards committees. A plethora of resources are available to help scientists, and science, achieve these goals.

Keywords: data sharing, file drawer problem, open access, open science, preregistration, questionable research practices, replication crisis, reproducibility, scientific integrity

Open Science

When we (the authors) look back a couple of years, to the earliest outline of this chapter, the open science movement within psychology seemed to be in its infancy. Plenty of people were pointing to problems in psychology research, collecting archival data to support the claims, and suggesting how science could be improved. Now it seems that the open science movement has reached adolescence. Things are happening—and they are happening quickly. New professional organizations are being formed to uncover and facilitate ways to improve science, often through new technology, and some old organizations are adopting new procedures to remedy problems created by past practices, often involving revising journal policies. Researchers are changing the way they teach, practice, and convey science. And scientific information (and opinions) are traveling fast. In blogs, tweets, Facebook groups, op eds, science journalism, circulation of preprints, post-print comments, video talks, and so on, more people are engaged in communicating science, and hoping to improve science, than ever before. Thus, any new technology, new procedure, new website, or new controversy we describe is likely to be superseded (or solved) even by the time this chapter is published. But the core values of open science should remain.

The "Open Science" Movement

Science is about evidence: observing, measuring, collecting, and analyzing evidence. And it is about evidence that can be shared across observers and, typically, although not necessarily exactly (Merton, 1973; Popper, 1959), replicated later. Science is about testing hypotheses, using inductive reasoning to create general rules (or theories) from the results, and then using those general rules to make further predictions or explanations of other evidence. And the evidence, the methods of collecting and analyzing that evidence, and the conclusions reached should be

open to the scrutiny and evaluation of other scientists (Lupia, 2014; Ioannidis, 2012). In this way, scientific knowledge can be self-correcting. Or so the story goes.

Beginning in about 2010, a confluence of events thrust questions about the integrity of the experimental practices of psychological science into the limelight. Concerns about non-replicability, post-hoc theorizing, inappropriate use of statistics, lack of access to materials and data, file drawers, and even fraud, had bothered psychological scientists in the past. However, the size, visibility, range, and timing of the recent events were particularly unsettling. The current open science movement grew out of these concerns about the integrity of psychological science (and other sciences, most notably medicine). Open science is a term for some of the proposed reforms to make scientific practices more transparent and to increase the availability of information that allows others to evaluate and use the research (Nosek et al., 2015). We acknowledge that many, or even most, of the proposed reforms of the open science movement are not new (see Bastian, 2016; Spellman, 2015, Table A1). Yet for various reasons they seem to be catching on now, whereas they had not done so in the past (Spellman, 2015; 2016). There are individuals who argue that these reforms will impede psychology research and handcuff the most productive people doing it (Baumeister, 2016; Stroebe & Strack, 2014). However, we believe that open science represents a return to the core beliefs of how science should be practiced (Cronbach & Meehl, 1955; Feynman, 1974), but updated for the present—by technology, by diversity, and by psychological research that explains how and why scientists can go wrong even when they have the best intentions.

The rest of this section describes what we believe to be the major and minor motivations for reform; the following section describes why we believe that now is the right time for reforms to succeed.

Major Motivations for Concern

The nearly concurrent juxtaposition of several different types of events provoked the field to action. Of major importance was the publication of two very different, yet in some ways similar, articles: Bem (2011), which purported to show evidence of precognition (i.e., accurately anticipating future chance events), and Simmons, Nelson, and Simonsohn (2011), which showed evidence of something even more magical (and more disturbing). Another issue was the growing list of high-visibility studies that could not be replicated by respectable and numerous labs. This problem provided a name that would stick for the widespread concern about the robustness of scientific research: the replication crisis. And finally, most shocking, but ultimately least relevant to the reform movement, was the revelation of fraud committed by several well-known psychologists (see Table 1 for a timeline of events).

-- Table 1 - Insert Timeline of Current Events About Here --

Questionable research practices. A maxim for good writing says: Show, don't tell. That's exactly what is done in the very clever and already-classic paper "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant" (Simmons et al., 2011). This paper beautifully illustrates that because of the leeway built into the design, analysis, and reporting of studies, researchers can "discover" just about anything.

The empirical part of the paper contains two experiments. In the first, undergraduate participants were randomly assigned to listen to either a children's song or a control song; afterwards they reported how old they felt and their father's age. An analysis of covariance (ANCOVA) revealed that participants felt significantly older after listening to the children's song. In the second experiment, different participants listened to either The Beatles' song "When

I'm Sixty-Four" or the same control song from the previous study. Afterward, they reported their

birthdate and their father's age. An ANCOVA revealed that participants who had listened to

"When I'm Sixty-Four" were significantly (nearly 1.5 years) younger than the other participants.

Yes. The result is not that those participants *felt* younger; it was that those participants *were*

younger. And given the randomized controlled design of the study, the business-as-usual

conclusion has to be: Listening to The Beatles' song *caused* people to *become* younger. Brilliant.

But, of course, the moral of the paper was that this result could only be reached through

"undisclosed flexibility"—the ability of researchers to collect unreported measures, choose a

sample size (i.e., stop collecting data whenever they want), choose which covariates to use in

analyses and which test statistics to report, change hypotheses post hoc to fit the data they

obtained, and use other not-quite-fraudulent but certainly questionable techniques to come to a

result interesting enough and pretty enough to merit publication in a top journal. Of course, such

a result is not likely to be robust to repeated testing.

To be clear, honestly reported exploratory analyses like those used by Simmons and

colleagues can have tremendous value, especially when later subjected to confirmatory tests on

new data. But psychology's approach for decades has been to advise "HARKing" (hypothesizing

after the results are known; Kerr, 1998)—that is, describing a result derived from data

exploration as though it had been predicted all along—and many of the practices described in

Simmons et al.'s (2011) paper were not only accepted, but also encouraged, by the field of

psychology. For instance, Daryl Bem's famous (2003) chapter on composing publishable articles

encourages authors to write "the article that makes the most sense now that you have seen the

results" (p. 2). Authors are also implored to explore their data: "analyze them [the data] from

every angle … if you see interesting patterns, try to reorganize the data to bring them into bolder

relief" (p. 2). And at the same time, they are advised to rewrite the history of that exploration for the sake of narrative: "the data may be strong enough to justify recentering your article around the new findings and subordinating or even ignoring your original hypotheses" (p. 3). Anecdotally, many reviewers and editors have asked authors to drop nonsignificant manipulations and measures from their manuscripts.[1] To examine how such practices might affect publications, O'Boyle and colleagues (O'Boyle, Banks, & Gonzalez-Mulé, 2014, "The Chrysalis Effect") compared management dissertations to their corresponding published journal articles and found that the ratio of supported to unsupported hypotheses more than doubled due to dropping nonsignificant findings, adding or changing hypotheses, and altering data.

In a survey published shortly after Simmons et al. (2011), many researchers revealed that they knew of people (themselves or others) who had engaged in some of these practices. Though falsifying data was rated as neither prevalent nor defensible, a few practices were prevalent (e.g., failing to report all dependent measures, reporting only studies that confirmed the hypothesis, data peeking), and many were viewed as defensible (John, Loewenstein, & Prelec, 2012). Although Simmons et al. did not use the term, John et al. reinstantiated an earlier label (from Swazey, Anderson, Lewis, & Louis, 1993) for such behaviors: "questionable research practices" (or "QRPs").

The use of QRPs could explain large-scale failures to replicate.

Failures to replicate. At the same time that psychologists were acknowledging the prevalence of QRPs, researchers across the field had become more vocal about being unable to replicate others' findings, including some that were well-cited and ground-breaking (Donnellan, Lucas, & Cesario, 2015; Doyen, Klein, Pichon, & Cleeremans, 2012; Harris, Coburn, Rohrer, &

---

[1]Kerr (1998) reports that in an unpublished study, over half of his respondents said that editors had asked them to alter a hypothesis post hoc.

Pashler, 2013; Johnson, Cheung, & Donnellan, 2014; LeBel & Campbell, 2013; Lynott et al., 2014; McDonald, Donnellan, Lang, & Nikolajuk, 2014; Pashler, Coburn, & Harris, 2012; Shanks et al., 2013; Zwaan & Pecher, 2012). Although there have always been failures to replicate, now many of them were occurring in multiple laboratories and, because of new media, there was increasing recognition that the failures were not isolated to single labs or lines of research. Many (although not all) of these failures to replicate were of experiments regarding social priming effects. Responding to that news in October 2012, Nobel Laureate Daniel Kahneman—whose best selling book *Thinking, Fast and Slow* (2011) extols many of those effects—warned in a widely shared e-mail that with regards to that research area: "I see a train wreck looming" (Kahneman, 2012).

In the "olden days," failures to replicate languished, unpublished, in researchers' file drawers and, if ever discussed, it was only in dimly lit hotel bars at conferences. As alternative publishing venues became available in the mid-'00s (e.g., *PLoS One*), it became possible (albeit still challenging) to publish replication studies. But when researchers couldn't replicate the results of an earlier study, the blame often fell on the researchers trying to do the replication. The assumption was that they were doing something wrong or, more unkindly, that they didn't have the necessary knowledge or experience to run the study (see, e.g., Gilbert, King, Pettigrew, & Wilson, 2016; Luttrell, Petty, & Xu, 2017). At the other extreme, the blame sometimes fell on the original researchers with suspicions that they had failed to communicate all that was necessary to run the study correctly or, more unkindly, that they had used questionable research practices or perpetrated some type of fraud. And there was the third non-judgmental possibility that the subject populations or the world had changed so much that the study would not have the same meaning to the present participants.

To systematically examine the extent and causes of failures to replicate, the early 2010s ushered in multiple coordinated, open replication efforts. For example, the journal *Social Psychology* dedicated an entire special issue to replications of highly cited social psychology studies (Nosek & Lakens, 2014). All replications were peer-reviewed prior to data collection, but at least 10 of the 27 highly cited tested effects failed to replicate.[2] Other journals began to publish replication attempts where replicators collaborated with original authors to ensure study fidelity (Matzke et al., 2015). One critique of these replication attempts was that most were done by single labs and thus failures to replicate could be attributable to lab-specific idiosyncrasies (or biases) or simply statistical flukes.

As a result, psychologists began to combine resources to conduct highly powered replications across multiple lab settings and populations. The Many Labs projects brought together dozens of diverse lab sites to replicate a small set of short, simple studies administered by computer or paper survey. To ensure study quality and mitigate any researcher bias, study methods and materials were peer-reviewed prior to data collection and when possible vetted by the original authors. In Many Labs 1 (Klein et al., 2014), researchers from 36 international laboratories replicated 13 studies, most of which assessed classic, well-established effects like anchoring (Tversky & Kahneman, 1974). Aggregating data across the labs ($N = 6,344$), 10 of 13 effects, including all the classic effects, were replicated as measured by 99% confidence intervals. These results suggested that many classic findings were reliable and that some effects were indeed reproducible, even by relative newcomers. Of course, these studies were not randomly selected, so this high rate of replicability might not generalize to the field of psychology more broadly.

---

[2] See also Marsman et al. (2017), who present an even grimmer view with a Bayesian reanalysis.

Indeed, replication results from a follow-up Many Labs project assessing more contemporary effects were more discouraging. Many Labs 3[3] (Ebersole et al., 2016) aimed to assess whether time-of-semester moderated replication success by testing 10 effects—a mix of classic and high-profile contemporary effects primarily from social and personality psychology—as well as 10 individual differences measures and 3 data quality checks across 20 university labs. Aggregated results ($N = 2,696$ in labs, $N = 737$ online) showed that time of semester and lab site had minimal effects on results. But perhaps more surprising was that only 3 of the 10 selected effects replicated (i.e., 7 of the 10 had null results).

Several other pre-approved, large-scale replications called Registered Replication Reports (RRRs) produced similarly disappointing results. Developed by Dan Simons and Alex Holcombe for *Perspectives on Psychological Science* (Simons, Holcombe, & Spellman, 2014), RRRs invited labs from across the world to replicate pre-approved studies. Unlike Many Labs projects, which assessed multiple short survey- or computer-based studies across the labs, in RRRs each lab replicated a single more labor- or time-intensive study. Researchers proposing a study to be replicated had to explain to the editors why that particular study had "replication value." If approved, the proposers worked with the original authors to develop a protocol for the replication that could be implemented consistently across multiple labs (typically in multiple languages). As of early 2017, only two of six RRRs successfully replicated the basic original result (replicated: Alogna et al., 2014, verbal overshadowing; Bouwmeester et al., 2017, time-pressure cooperation[4]; failed to replicate: Cheung, et al., 2016, commitment and forgiveness[5];

---

[3] Results from Many Labs 2 (Klein et al., *under review*) had not been published at the time this chapter was finalized.

[4] The replicators in Bouwmeester et al., however, found that the original results successfully replicated only when ignoring participant selection effects.

[5] The manipulation check showed that the experimental manipulation had failed.

Eerland et al., 2016, grammatical aspect and intent; Hagger et al., 2016, ego depletion;

Wagenmakers et al., 2016, facial feedback and emotion).

Perhaps the most well-publicized large-scale replication project, however, was the

Reproducibility Project: Psychology (Open Science Collaboration, 2015). Whereas the Many

Labs and RRR projects involved multiple labs replicating a small number of carefully selected

studies, the Reproducibility Project aimed to estimate reproducibility more generally by

coordinating single labs that each replicated one of 100 semi-randomly selected findings. More

than 270 researchers joined the effort, led by Brian Nosek and other researchers at the Center for

Open Science. To decrease selection bias, the replicated studies were chosen from articles

published in 2008 in three top journals (*Psychological Science*; *Journal of Experimental

Psychology: Learning, Memory, & Cognition*; and *Journal of Personality and Social

Psychology*). To make the process transparent, methods and analysis plans were pre-registered

prior to data collection. To increase fidelity, the replication teams worked with original authors

whenever possible to recreate the original studies as closely as they were able. Yet, whereas 97

of the original findings were statistically significant (i.e., $p < .05$), only 35 of the replicated

findings were significant. And fewer than half of the replications had effect sizes that fell within

the 95% confidence interval of original studies.

Regardless of the outcomes, the process of trying to do these replications was quite

informative. First, researchers had to acknowledge the difficulty in doing direct replications.

Much is left out of the method descriptions of a journal and much is left for guessing. Second,

even with open materials and support from original authors, reproducing psychologically

equivalent procedures and measures across populations and time sometimes proved challenging

(Nosek & E. Gilbert, 2016; Spellman, 2013). Third, the field had to acknowledge that there was

no obvious way to interpret what it meant to have a successful replication. Did the effect of

interest have to be significant in the proper direction? Or maybe the measure should be whether

the effect size of the replication fell into the effect size confidence interval of the initial study. Or

perhaps Bayesian statistical analysis—which provides a more continuous estimate of how

consistent a result is with the original result versus a null result—would be more appropriate (Etz

& Vandekerckhove, 2016). However counted, most researchers saw the results as dismal (but see

D. Gilbert et al., 2016).

Although widely noted failures to replicate prompted the name the replication crisis, the

subsequent rise in large-scale replications also garnered substantial attention and gave rise to a

more optimistic name for the focus on improving science: the replication revolution.

Fraud. Though research on QRPs and failed replications provided evidence of

widespread problems in the field of psychology, the most shocking event, at least for social

psychologists, was the revelation of fraud by Diederik Stapel, a prominent Dutch researcher.

Knowledge of his fraud broke in 2011, and the final report about his actions (Levelt Committee

et al., 2012) found evidence of fraud in over 50 of his publications. Retraction Watch

(http://retractionwatch.com) reports his current count[6] as 58, which puts him third on their

leaderboard.

At about the same time, investigations were continuing into misconduct by the Harvard

cognitive-evolutionary psychologist Marc Hauser (Wade, 2010). And in 2011 and 2012, close

analyses by Uri Simonsohn (2013) led to university findings of misconduct and then retractions

by the social psychologists Lawrence Sanna (formerly of University of Michigan, eight papers;

Yong, 2012) and Dirk Smeesters (formerly of Erasmus University Rotterdam, six papers;

---

[6] Current as of September 8, 2017.

Enserink, 2012).

The revelations of fraud could be viewed simply as the result of a few bad actors and bad decisions. However, in combination with the broader concerns described earlier, they helped galvanize the field into action.

Other Motivations for Concern

In addition to pressing concerns about QRPs and failed replications, researchers expressed other concerns about the general practices of science and scientists. Most of these involved the inability to access all relevant information about published studies and the inability to access the complete corpus of scientific work (both published and unpublished).

Lack of access to full methods. Researchers wanting to replicate or simply better understand others' studies were frustrated by their inability to obtain the exact materials and detailed methods used by the original researchers. Of course, all empirical articles contain method sections, but these were often short and incomplete, particularly with the strict adherence to word limits in the short-form articles that had grown in prevalence (Ledgerwood & Sherman, 2012). With the passage of time, materials that were not properly archived were often simply lost or otherwise not recoverable. Some stimuli (such as confederate behavior scripts or video clips) were not often preserved due to technical or media limitations.

Lack of access to data. Some scientific fields make a distinction between *replication* (rerunning a study with the same protocol to gather new data to see whether the same result is obtained) and *reproduction* (reanalyzing the original data to see whether the same result is obtained). Psychologists have been largely concerned with the former—hence the need for access to full methods. But access to original data can be useful.

Prior to the advent of modern computing technologies, psychological researchers did not

routinely archive their data. Records were maintained for some minimal amount of time in personal file cabinets and boxes, after which they were discarded. This may come as a shock to younger researchers who have never experienced professional life without computers, which promise seemingly limitless and permanent storage capabilities, but paper records of research were physically burdensome and extremely tedious to archive. With computing power and online cloud storage increasing, and with more attempts at working cumulatively (either through meta-analysis or direct replication), researchers were becoming more frustrated by their inability to obtain the data from published studies for reanalysis or inclusion in meta-analyses (Morey et al., 2016), despite publication guidelines (e.g., American Psychological Association [APA], 2010a) stating that authors should be willing to share data for such purposes. Some researchers who received many such requests were becoming tired or leery of them (e.g., Schnall, 2014). Certainly, requests for data from older publications can easily be difficult and time-consuming to reconstruct (e.g., in some old format on some old backup technology) or lost (e.g., in the possession of a past graduate student who left the field long ago).

Lack of access to analytic procedures and code. The rise of point-and-click statistical software meant that more researchers than ever had access to advanced analytic techniques. Such software does not force researchers to save the commands they used to complete their data analysis, and many psychologists failed to preserve their analysis code, rendering exact reproduction of analyses and their results challenging even when using the same data as input. Other social sciences (notably economics and political science) have also made note of this issue and have begun taking steps to address it. For instance, the *American Journal of Political Science* requires authors of accepted manuscripts to submit "replication packages" for review to the journal (Jacoby, 2015). Contracted specialists in archiving verify submitted materials to

ensure that results can be reproduced, but psychologists have yet to adopt such practices.

The file drawer problem. Along with failures to replicate was a growing frustration with the inability to publish replication failures (or successes). Researchers were acknowledging that the growing file drawer problem was leading to a mischaracterization of which findings were actually robust (Spellman, 2012a). Coined by Rosenthal (1979), "the file drawer problem" refers to the fact that some research never ultimately makes it into the published literature and instead languishes in researchers' file drawers. In addition to replications, null results and failures to support a hypothesis rarely appear in print (Fanelli, 2012; Franco, Malhotra, & Simonovits, 2014; Makel, Plucker, & Hagerty, 2012). Evidence suggests that researchers might be changing their hypotheses after the results are known to convert null results into hypothesis-supporting results ("HARKing"; Kerr, 1998). These omissions and alterations made cumulative science and meta-analysis challenging, if not impossible.

Lack of access to publications. Researchers were also becoming frustrated by the fact that in a free-access cyberworld most scientific publications are only available for a fee. Although university libraries may make publications available to employees without cost, the libraries pay exorbitant fees for access for their readers, and their subscriptions typically cover neither all relevant journals nor all relevant people. Furthermore, in some countries including the United States, a large portion of research is paid for through (taxpayer funded) federal grants, leaving those same taxpayers wondering why they must pay twice for access to information from research. In terms of social justice, one of the biggest victims of for-profit publishing are researchers in developing nations, who lack institutional or personal access to much of the research literature (see Tennant et al., 2016, for an overview).

Discontent with reporting and use of standard statistics. There has long been

discontent with the use of null hypothesis significance testing (NHST) and with the reporting of

statistics in psychology journals (e.g., Cohen, 1994; Cumming & Finch, 2005; Lykken, 1968;

Meehl, 1967). The vocal dissatisfaction intensified after the publication of Bem's (2011)

precognition paper in the most prestigious social psychology journal. Subsequent post-

publication critique revealed numerous troubling flaws, especially related to data analysis

(Wagenmakers, Wetzels, Boorsboom, & van der Maas, 2011) and research practices (LeBel &

Peters, 2011). We suspect that Bem's paper is unlikely to have garnered such close scrutiny, and

have had those problems identified, had the topic and findings not been so controversial. (The

controversy over appropriate statistics to use and how to report them is not covered in this

chapter.)

Summary and Sequel

Evidence of QRPs, failures to replicate, fraud[7], and difficulties accessing important

information prompted psychologists to acknowledge that scientific reforms should be designed

to address the structural problems that underlie the replication crisis. Moving to the forefront

were ideas for assessing the problems, for decreasing the use of QRPs, and for ameliorating the

file drawer effect (see the November 2012 issue of *Perspectives on Psychological Science* for a

variety of early suggestions about what caused the replication crisis and proposals to address it).

Among the important insights during the early discussion were that the concerns mentioned

earlier were not independent of each other and that many, if not all, result from the problem of

misaligned incentives: that what is good for being a successful scientist is not necessarily what is

good for science itself (Nosek, Spies, & Motyl, 2012). Scientific practices might therefore be

difficult to change if institutions don't change reward structures. Psychologists thus realized that

---

[7] Indeed, although fraud may be revealed by review processes (e.g., Simonsohn, 2013), it is more often
revealed by lab whistleblowers (Stroebe, Postmes, & Spears, 2012).

making science more open would take efforts on the part of various types of stakeholders—

including researchers, editors, reviewers, publishers, funding agencies, and hiring and promotion

committees—some of whom might be resistant to change. Despite these challenges, we are

optimistic that the current time is ripe for it.

Why the Time Is Ripe for Change

Most of the problems just described are not new to the field of psychology, but previous

attempts to fix them had failed (see Spellman, 2015a, Table A1[8]). So what is different now that is

allowing current reforms to take off and take hold? We suggest four broad relevant factors: (1)

technology, (2) demographics, (3) generality of the problem, and (4) insights from psychological

science itself. Of course, there are also countervailing forces, but we save a discussion of the

pushback for the later section called Objections to Open Science. A fifth relevant factor that is

often overlooked is that there were some people and institutions in place and ready to take fast

action to facilitate reform. (For more on the fifth factor, see Table 1 and Chambers, 2017.)

Technology

One thing that is different now is technology, which is both a cause of and a cure for

some of the problems. Technology makes psychological science move faster—from the

presentation of stimuli, to the recruitment of larger participant samples, to the rapid analysis of

data and dissemination of results. And it creates the ability to handle huge data sets—from brain

scans to wearable devices that capture continuous visual fields or physiological measures. But

technology also creates bottlenecks and conflicts; when research is done quickly, expectations

---

[8] Also at:
https://docs.google.com/document/d/1lmnYIcavpXjXo2GA2m7kytKdnZxJPoXVWLoYauxFu5s/edit

are for quick publication, but we still live in a mostly slow and limited print-based publication

world (although that seems to be changing[9]).

The vastness and speed of electronic communication can help cure many of the problems

that led to the replication crisis. First, technology allows for making more information available.

Print journals may keep page limits, but they can create online access to the full materials, data,

and analytic code for journal articles. Or such information can be placed on a third-party server.

Second, technology allows for more people, and more research, to be part of the conversation.

Technology allows researchers to communicate results that would never have gotten published

previously because they were "mere" replications or failures to replicate. It allows distant

researchers to visibly comment on and aggregate thoughts about recently published findings.

However, one challenge related to using technology for reform involves encouraging behavior

change in the research community. Outside of a handful of technophilic individuals, most

researchers will not add a new practice (e.g., pre-registration) or technology (e.g., reproducible

analysis in R) to their workflow without sufficient motivation to change. Researchers must be

convinced of the benefits of new tools, and they must feel sufficiently rewarded for using them.

As we currently see it, new technology has mostly been layered on top of the existing publishing

system. For instance, publishers still produce manuscripts typeset into digital documents that

mimic their physically printed predecessors, whereas open analytic code and data (when it is

shared) is currently linked to a journal's website as seemingly afterthought supplementary

materials. There is little consistency or standards across publishers in terms of how this is done.

Libraries continue to index journals by their volume and issue numbers, as they have for decades

---

[9] The number of online-only journals is growing quickly, as is the number of preprint posting services.

and even centuries, yet data accompaniments to manuscripts do not usually receive their own digital identifiers.

In the face of rapid technological developments and changes, we suspect that libraries and digital curation of research materials, data, and code will become even more valuable. Such curation will enable researchers to take full advantage of the power of these (currently supplementary) materials. For instance, when researchers combine results meta-analytically, they will have better access to relevant studies and less biased estimates as a consequence.

Demographics

A second factor is the changing demographics of psychology researchers. The current young researchers grew up not only with computers but also with faster web-based technology. They seem to be more used to sharing information than older generations. Younger scientists are more diverse in gender, race, and ethnicity than those who inhabited psychology labs during the rapid growth in departments in the decades following World War II (National Science Foundation, National Center for Science and Engineering Statistics, 2015). And, as the field has expanded in both size and content, young researchers have been trained in a larger and more diverse set of labs that, we suspect, vary more in procedures and assumptions than those from the past.

A Science-Wide Problem

Some psychologists may be resistant to reform because they do not want to highlight any problems that might undermine respect for psychological science. However, a factor that might be contributing to the success in changing practices now is the knowledge that our field is not alone in struggling with the issues we have outlined. Non-replicability of findings is rampant in all of the life and social sciences, probably most scarily in medicine. John Ioannidis (2005)

famously pointed that out in his paper "Why most published research findings are false."

Subsequent systematic attempts to replicate clinical drug trials revealed that he was not far off

the mark, and reproducibility rates may be even lower for cutting-edge research. Researchers at

Bayer HealthCare reported replicating only about 25% of 67 preclinical oncology, women's

health, and cardiovascular medicine studies published between 2008 and 2010 (Prinz, Schlange,

& Asadullah, 2011). Scientists at the biotechnology firm Amgen attempted to replicate 53

landmark haematology and oncology studies and only confirmed 6 (11%) (Begley & Ellis,

2012). Recently, the Open Science Collaboration reported initial results from Reproducibility

Project: Cancer Biology studies, and four of seven projects successfully reproduced results

(Kaiser, 2017a, 2017b).

Of course, psychology and medicine are only two examples. In 2010, a high-profile

article about the "Decline Effect"—the way that scientific findings seem to decrease in

magnitude or even disappear with repeated replication—appeared in *The New Yorker* (Lehrer,

2010). In the fall of 2012, Jonathan Schooler organized a conference at UC Santa Barbara during

which researchers from a variety of life science fields reported those same problems.

Other social sciences, like economics, have run their own replication studies and found

their own crises. A recent coordinated replication of 18 between-subject laboratory experiments

from two top economics journals found that only 11 (61%) studies replicated as measured by $p <$

.05, despite being powered at over 90% and sending study plans to original authors for

verification (Camerer et al., 2016). Results for reproducing the results of nonexperimental

studies using the same data and analyses[10] have been even worse, with rates below 50% (Chang

---

[10] These failures to reproduce results given the *same data*—sometimes even with the help of the original
authors--highlights the need to appropriately record and share detailed analysis records. (See the
section: Practicing Open Science: For Researchers.)

& Li, 2015, 49% with original authors' assistance, 33% without; Dewald, Thursby, & Anderson, 1986, 13%; McCullough, McGeary, & Harrison, 2006, 23%).

There is currently a dearth of data on reproducibility rates in other fields such as chemistry, physics, and environmental studies. However, survey data suggests that they, too, are not immune from concerns. In a study conducted by Nature Publishing Group (Baker, 2016), over 1,500 scientists from biology (about half), chemistry, physics/engineering, earth and environmental science, medicine, and "other" sciences reported on their beliefs about replicability of their fields' published research and on results from their own replication attempts. Although some fields were more optimistic than others, across every field over 40% of respondents reported failing to reproduce their own work at least once and over 60% reported failing to reproduce someone else's work.

As a science, psychology is not alone in having a replication crisis; but as scientists, psychologists might be better prepared to address these issues than others.

Insights From Psychological Science Itself

A final, optimistic factor brought home in the talks and writing of Brian Nosek is that, as psychological scientists, we should have a better understanding (than scientists from other disciplines) of the psychology of motivation and decision making. Therefore, we might have insight into both the causes of the problems and how to fix the problems that stem from the previous ways of doing research.

An important insight is that problems have arisen in science because the current incentives for scientists do not align with what is good for science. For example, scientists feel pressured (in the search for jobs or grants or tenure or awards) to publish many papers. Journals want to publish papers in which the authors' hypotheses are novel and cleanly supported, while

also being likely to garner immediate recognition and future citations in order to improve their impact factors.

These values have incentivized scientists to run many small low-powered studies, testing flashy hypotheses. Such perverse incentives may have contributed in the long run to the perpetuation of bad practices. Smaldino and McElreath (2016) persuasively modeled a system in which a process that mimics natural selection could reinforce such practices. If quantity is consistently rewarded over quality, then over time, selection will act against quality—weeding out those individuals who choose to invest effort in high-powered designs or confirmatory replications.

Thus, under the old regime, if the initial hypothesis was not supported, there was a reward for changing the hypothesis (HARKing; Kerr, 1998). If the hypothesis was supported, but the data were "messy," there was a reward for running additional analyses, changing criteria for dropping subjects, or reporting only some of the study's conditions or dependent measures. Of course, not all researchers engaged in such QRPs, but many admitted to some of them (John et al., 2012), and the incentive structure rewarded them. Accordingly, one task for psychologists who were in favor of change was to think about the kinds of incentives that would promote positive change without hurting individuals or the vibrancy of scientific progress.

As psychologists, we should reflect on some psychological processes that can get in the way of scientific progress. First, confirmation bias describes the search for evidence that will confirm a preexisting belief or theory, while ignoring or downplaying evidence that would counter that belief. Second, hindsight bias ("I knew it all along") helps researchers to transform evidence that might have countered a theory into evidence that ultimately supports the researcher's (perhaps new) position. Combine these two biases with flexibility in design and

analysis (John et al., 2012) and mixed messages on which statistical techniques are *really* best

(Silberzahn et al., 2015), and the conditions are right for getting the wrong answer. These biases

pervade daily life, and scientists are unfortunately not immune to them (Chambers, 2017).

On the one hand, knowledge of these biases is not new (Kerr, 1998), and methodologists have

been imploring researchers to pay more attention for many years (Meehl, 1967). But we now

know more than ever about these biases and how to combat them. For instance, preregistration—

in which researchers commit ahead of time to their hypotheses, methods, and analytic strategy—

can serve as an honest signal (both to outsiders and to researchers themselves) about when a

researcher may be falling prey to hindsight bias. Another technique is registered reports

(Chambers, Feredoes, Muthukumaraswamy, & Etchells, 2014), in which reviewers scrutinize a

research plan prior to the onset of data collection, helping to make sure the test is well designed

and fair (i.e., can produce an informative result). Thus, pre-data collection review can be

especially helpful to researchers trying to avoid confirmation bias. Finally, blinding (MacCoun &

Perlmutter, 2017) is a process in which data are perturbed or condition labels are scrambled in

order to decouple data exploration from knowledge of a study's results. Using a holdout sample

(in which exploration occurs on half of the data, while the remaining half is "held out" for later

confirmatory testing) is another technique with a similar purpose. In both cases, after researchers

have thoroughly explored the data, the blind is lifted (or the hold-out sample is unveiled) to

conduct a high quality confirmatory test.

Psychology can thus explain why things have gone wrong, even without scientists

intentionally doing anything wrong. And psychology is important for understanding ways to

incentivize better scientific practices and change behavior (Miller & Prentice, 2013). Many of

these approaches can be observed in the next few sections on how to practice open science.

Practicing Open Science: In General

There is increasing interest in open science, and many journals and science organizations now require some level of open practices as a stipulation of publication (see Table 3). Of course, most psychological scientists, ourselves included, currently fall short of fully implementing open science in practice. How can we all improve?

Before answering this question, we want to acknowledge that there is not one best way to practice open science. Some types of research may require procedures tailored to unique study concerns (e.g., Campbell, Loving, & LeBel, 2014, noting the differences between experimental designs and long-term correlational methods; Patrone, Resnik, & Chin, 2012, discussing biosecurity). And among those who practice open science, there is a spectrum of how transparent people make their research, and there is even room to debate the ideal level of openness (see Axt, 2016).

However, there is growing consensus that, barring special circumstances, transparency and openness imply that all research-relevant information be available, discoverable, and intelligible to others (e.g., Boulton et al., 2012; Christensen, 2016; Destro Bisol et al., 2014; Miguel et al., 2014; Simmons, Nelson, & Simonsohn, 2012; Stodden & Miguez, 2014). This entails recording and publicly sharing data, materials, scripts, relevant hypotheses, and design and analysis choices, as well as any notes or keys needed to understand those materials. Table 2 illustrates how these and other techniques of open science can prevent or remedy potential problems arising from the research process.

Some of these procedures, such as adequately organizing and annotating one's work, are valuable to researchers regardless of the "sharing" component. Benefits—to individual researchers (see McKiernan et al., 2016), to science as a whole, and to the public—conferred by

open science practices include: encouraging collaboration, promoting appropriate allocation of

credit to researchers, enabling scientists to more easily build on prior work, supporting meta-

science, facilitating theory development, increasing the return on investment from grant funding,

and making scientific results more accessible to interested readers. Some journals reward articles

that have open science features above what is required, and some organizations now provide

funding and prizes to encourage open science research.

We recognize that these requirements may sound overwhelming to those new to open

science. But an increasing number of free resources—from online tutorials to regularly updated

manuals—have been designed to support scientists at every step of the process. Plus, to keep

themselves informed and to encourage open practices generally, scientists can attend talks and

join professional groups that discuss open science issues (e.g., the Society for the Improvement

of Psychological Science [SIPS], several Facebook groups). Table 3 provides a list of resources

that support and facilitate open science practices. Many of these resources are maintained and

updated by the Center for Open Science (COS) on the Open Science Framework (OSF; 2015).

Co-founded by Brian Nosek and Jeffrey Spies, COS is a non-profit organization that develops

technology to help researchers increase transparency and reproducibility of their work. OSF is a

free online platform that allows individuals and groups of researchers to store and share research

materials.

In the following section, we outline simple, practical ways to start practicing open

science in some of the many roles of an academic psychologist: researcher, teacher, author, and

reviewer.

-- Table 2 Chart of Connections Across Issues --

-- Table 3 Resource Box --

Practicing Open Science: For Researchers

To make psychology more open, researchers can integrate open practices into their regular workflows. Specifically they can create systems to ensure that their data and materials are available, intelligible, and publicly shared. Except for the sharing part, these are all already good practices within a traditional lab workflow. In particular, they are useful for retaining important information across time and over personnel changes, keeping team members up-to-date on lab projects, preserving documentation for IRBs, and organizing pilot studies and materials relevant for future lab projects or publications. The FAIR guiding principles (Wilkinson et al., 2016) describes data management and archiving best practices that are summed up by: Findable (e.g., have a unique doi), Accessible (e.g., be open sources), Interoperable (e.g., use a common language), Reusable (e.g., be well described).

Make It Available: Save

The first, most basic step to practicing open science is not sexy or new, but it is crucial: Researchers must appropriately save research materials and data. Estimates suggest that the majority of researchers do not adequately provide published data and materials, even when requested to do so by other researchers (Vanpaemel, Vermorgen, Deriemaecker, & Storms, 2015, 62%; Wicherts, Borsboom, Kats, & Molenaar, 2006, 73%). Though this lack of sharing may sometimes be due to purposeful withholding (see Wicherts, Bakker, & Molenaar, 2011), it is more likely that data and materials are simply stored improperly and lost (see Vines et al., 2014). Ideally, researchers would save everything, even specific materials used and raw data—for example, the actual surveys completed and EEGs recorded—for at least a minimal amount of time. This archiving allows for error checking in data interpretation or entry or even for auditing for fraud. For example, original video recordings during experiments provided clues about bias

in coding monkeys' behavior during the fraud investigation of former professor Marc Hauser

(Committee on Professional Conduct of Harvard's Faculty of Arts and Sciences, 2010).

But perhaps more importantly, saving allows for reanalysis or reinterpretation of old data or

materials given new theories, questions, or methods. Chemist Stanley Miller, for example, stored

the laboratory materials and detailed notes for his famous 1954 Miller-Urey study that produced

organic compounds from inorganic ones, which allowed his graduate students in 2008 to re-

analyze his samples using new, more sensitive techniques. Their results indicated that the

original Miller experiment produced over 20 more new compounds than originally reported.

Similarly, in the 1990s cognitive psychologists Ian Deary and Lawrence Whalley rediscovered

original notebooks containing intelligence test data for Scottish children collected by the Scottish

government in the 1930s and 1940s (Deary et al., 2004). The old data have since been linked

with current data, leading to dozens of new findings about how intelligence and other individual

differences are associated with aging and long-term outcomes.

Even if physical materials and raw data cannot feasibly be saved forever, online file

storage systems make it easy to save most compiled datasets, analysis scripts, and notes about

research background, hypotheses (if any), methods, and procedures in perpetuity.

Not only should data be saved for as long as possible, but researchers should also ensure

that materials and data are stored in formats that can be easily reused by others for as long as

possible. Data stored as PDF files, for example, may not be easily searchable or analyzable, and

SPSS files may be usable only by others with access to expensive software.[11] Thus, users should

consider saving data and files in flexible, generic formats such as comma separated spreadsheet

---

[11] There is a free R package that reads .sav files. PSPP and JASP (other free statistical software) also read .sav files.

(.csv), rich text format (.rtf) documents, or text (.txt) files, that might survive the constant

updating of proprietary formats.

Physical storage devices like portable flash drives and general online file storage systems

like Dropbox.com provide easy personal storage, but files may be lost or insufficiently secured.

Several stable, secure online repositories have developed specifically for empirical research (e.g.,

FigShare, OSF). These repositories often include features designed to help researchers' daily

workflows, such as version control systems and the ability to selectively share with colleagues or

the public more generally. OSF also allows researchers to assign DOIs (i.e., digital object

identifiers) to archived materials, including data. These DOIs can be assigned separately for

different components of a project. Specialized repositories also exist for subtypes of research,

such as neuroimaging (e.g., openfmri.org), developmental psychology videos (databrary.org).

and genomics (e.g., dbGaP), which often incorporate tools for addressing specialized concerns

like participant anonymity and large file sizes. Researchers continue to work on tools for saving

massive datasets—those too big, for instance, to download onto one computer—but in the future

even this type of data should be saved and shared (Crosas, King, Honaker, & Sweeney, 2015).

Another option is to publish the dataset as a "data descriptor." The idea here is that high quality,

reusable datasets are published in their own articles, garnering credit for the research teams that

produced them, as well as subjecting the data and accompanying documentation to peer review

to ensure quality and future use value. *Scientific Data* (http://www.nature.com/sdata/) began

accepting social science datasets in late 2014 after debuting in biology in late 2013 (Nature

Publishing Group, 2014), and the *Journal of Open Psychology Data* publishes and archives

psychology datasets with high reuse potential. The Transparency and Openness Promotion

(TOP) Guidelines, discussed later, encourage journals to ask authors for citations to such information (see Table 4).

Make It Intelligible: Organize, Label, Annotate

Saving materials and data alone is not sufficient, though—to be open, data, materials, and supporting documentation must also be organized, labeled, and annotated in a way that third parties can understand and use them. All potentially relevant manipulations and measures should be labeled—independent and dependent variables, covariates, demographics, and so forth. Often it is helpful to create a separate key or codebook describing each variable. Data analysis scripts also should be annotated and explained, such that the purpose of each analysis can be isolated and understood. And information about hypotheses as well as relevant prior pilot work should also be recorded. Developing a version control system or using a repository that automatically assigns version control (e.g., OSF) may also help clarify and record changes in the research process and datasets over time.[12]

We recommend that the process of labeling and note taking occur *throughout* the research project, rather than when a paper is ready for review (see Donnellan & Conger, 2007). As many have pointed out, memory is fallible. Most scientists juggle multiple projects, many projects take months or years to complete, and hypotheses may change across studies or over time. Even the most thoughtful researcher, therefore, is likely to forget or misremember relevant details of the research process. Moreover, as discussed earlier, after researchers have completed data analysis, they may be unintentionally biased toward remembering only hypotheses, data plans, and such that are consistent with whatever story *post hoc* data analysis supports. In turn

---

[12] Version control means that there is a clear, transparent, and documented system for tracking a file as it is revised over time. Rather than manually saving files on the local computer as (for instance) "Version1," and "Version 1.1," a well-functioning version control system preserves both newer and older versions of files and makes it apparent to outsiders how changes are occurring over time.

they may fail to adequately note measures, materials, hypotheses, and analyses that failed to

return interesting or consistent results (see Feynman, 1986; Nosek et al., 2012; Nuzzo, 2015).

Record keeping at each step of the research process helps capture information that would thus be

lost or forgotten.

To help make record keeping easier, some scientists have developed checklists and

templates for what should be recorded at different steps of the research process (see Table 3).

These templates vary in level of detail but generally outline key information that should be

recorded prior to data collection (e.g., power analyses, procedures, variables and their

explanations, and additional information for confirmatory studies like exclusion criteria,

hypotheses, and analysis plans), after data collection (e.g., notes about final sample size or

anything unusual that occurred), and after data analyses (e.g., annotated analysis script files and

data keys, exploratory analyses). Some of these checklists have themselves been made open

access, allowing anyone to use them. Additionally, software programs like Sweave and LaTeX

can also be used in combination with programs like R to create or apply standard, annotated

analyses to datasets (see Harrell, 2014).

Make It Discoverable: Share

Of course, being open requires that, ultimately, researchers publicly share their work in a

place that is easy to find. Published work should state where relevant data and supporting

materials are available, using stable URLs or other stable digital identifiers (e.g., DOI).

Repositories such as OSF and FigShare allow single documents or entire folders to be shared

with individuals or made completely public with the click of a button. Public releases are then

searchable and can be linked to papers, blog posts, and personal webpages.

When and how much to share. Some researchers have concerns about when sharing should occur and whether some information should be withheld for ethical or pragmatic reasons.

Practices vary among researchers. Many wait until the relevant research has been published and then share information about only the specific variables and data analyses reported. To support this choice, some mention concerns that sharing large amounts of information would be onerous to researchers and decrease incentives to collect data, thus leading to less discovery and publication of true findings. Publishing large quantities of complex materials and data might even burden readers, who could have a hard time distinguishing important from unimportant information. Such concerns might be particularly relevant to research involving complex, longitudinal designs involving numerous variables (Finkel, Eastwick, & Reis, 2015). Indeed, journals (e.g., *PLOS ONE*) and badge systems (e.g., *Psychological Science*'s Open Data badge) currently require that authors share only the data and materials necessary to reproduce reported results.

Other researchers, however, release data and materials—including variables that are not reported—as soon as the research is under review or even sooner. Jeff Rouder, for example, automatically uploads his lab's raw data nightly, a process he calls "born open" data. Rouder notes a belief that data are collected for the common good and that researchers should view themselves as data "stewards" rather than "owners" (Rouder, 2016). To the benefit of science, broad data sharing may accelerate the accumulation of new knowledge by encouraging others to analyze existing data in new ways. (For an example of researchers using existing data to publish new analyses see Corker, Donnellan, Kim, Schwartz, & Zamboanga, 2015).[13] To the benefit of the researcher who collected the data, it also often leads to increased citations (McKiernan et al.,

---

[13] Ironically, Gilbert, King, Pettigrew, & Wilson, 2016, which criticizes reproducibility and changing norms, is based on openly shared data.

2016) and may lead to future collaborations and authorship (*see* Ebersole, 2017, discussing

IJzerman, Pollet, Ebersole, & Kun, 2017).

Concerns about sharing early and broadly. One oft-cited concern about sharing early

is that researchers will be "scooped" (see, e.g., #scoopingsci for one online Twitter discussion).

Indeed, we know of at least one example of a team's data being used for another's publication

submission before the team who collected the data had published on it (Capaldi, 2015, using data

from Ebersole et al., 2016).[14] And, given that our field often rewards "first to publish" rather than

"best to publish" a finding, quickly publishing using someone else's data could hinder the

original collector's ability to carefully and thoroughly work with or explain the data prior to

publication (see Overbye, 2005, discussing the controversy over the discovery of the dwarf

planet Haumea). These concerns may be particularly applicable to researchers who conduct

resource-intensive large-scale, longitudinal studies (Finkel, et al., 2015; but see Campbell,

Loving, & LeBel, 2014).

But fears of "scooping" likely are outweighed by benefits not only to science generally

but also to individual data-collecting researchers. First, publically sharing study information

communicates to others that you have been working on that question, likely encouraging others

(or at least those at earlier stages of the research process) to choose different research questions.

Alternatively, researchers working on similar questions might choose to collaborate. To decrease

concerns about scooping or free-riding, as a community we may want to develop guidelines for

when collaboration would be appropriate or for when independent, dual publication on the same

finding would benefit science. Additionally, researchers may benefit from sharing study data and

materials by receiving increased citations (Belter, 2014; Piwowar & Vision, 2013). Scientists

---

[14] The publications answered different research questions. We believe that this would often be the case for
people re-using data.

should create stronger norms for citing shared data and materials and—crucially—highly value

such citations (Belter, 2014; Chen, Downs, & Shumacher, 2013; Nosek et al., 2015).

Another concern about sharing is that doing so may violate the privacy of participants, reveal

other sensitive information (e.g., security concerns like military decisions or biohazard data), or

be protected by intellectual property rights (e.g., trade secrets). If this is the case, then some

information may reasonably be withheld or only disclosed in aggregate. Efforts should be made

to ensure that researchers share only what is not sensitive. To protect participants, some of us, for

example, often simply delete any identifying information (e.g., I.P. addresses for online workers,

unique demographics) from publicly shared data. Internal Review Boards (IRBs) can also

provide guidance to help inform participants during the consent process that anonymized data

will be shared, and they can help researchers develop protocols to ensure the ethical release of

information. Authors may also work with funding and oversight agencies (see Fouchier et al.,

2013).

Preregister Confirmatory Research

Why should researchers preregister research? As noted earlier, confirmatory research

(which tests *a priori* hypotheses specified before data collection) and exploratory research

(which creates new *a posteriori* hypotheses based on examining collected data) are both

important parts of the scientific process. But distinguishing between the two types of research is

crucial when interpreting inferential statistics. This is because confirmatory research is limited to

focused tests of key variables, thereby reducing the chance of finding a statistically significant

result when a real effect does not actually exist (i.e., type I error, or "false positive"), whereas

exploratory research decreases the chance of missing a potentially generative future direction

(i.e., "theoretical false negative"; see Fiedler, Kutzner, & Krueger, 2012).

Historically, however, many researchers had reported exploratory research as though it

had been confirmatory (Bem, 2003; John, et al., 2012; Kerr, 1998 [coining the term "HARKing,"

for hypothesizing after results are known]). This practice likely led many scientists to cherry-

pick "good results" from many analyses and to dismiss less interesting or inconsistent ones, in

turn increasing false-positive rates (Gelman & Loken, 2014; Ioannidis, 2005). Simmons et al.

(2011) termed related processes "p-hacking," meaning analyzing data multiple ways to obtain a

statistically significant result. Preregistration of confirmatory research—or publically committing

to a specific research question, method, and data-analysis plan prior to data collection[15]—helps

decrease this possibility[16] (Bishop, 2013; Chambers, 2013; Wagenmakers, Wetzels, Borsboom,

van der Maas, & Kievit, 2012).

Currently there is no field-wide standard for how or what to register, and pre-registration

takes a variety of forms. To be truly confirmatory, however, hypotheses, methods, and data

analysis plans should be stated precisely, rather than in general terms that could be interpreted as

supporting multiple findings. For example, stating the hypothesis that "X will be bigger than Y

in condition Z" is more precise than stating, "X and Y will be different in some conditions."

Researchers also should specify which specific statistical tests and data plans will be used to

answer which questions (e.g., *t*-tests, multilevel modeling). Of course, sometimes appropriate

statistical tests or analysis plans may not be determined until data has been collected—data may

be surprisingly binary rather than normally distributed, measures may be at ceiling, manipulation

---

[15] In the case of archival research or preexisting data sets, preregistration would occur prior to data compilation, cleaning, and analysis.

[16] Currently two, somewhat different, varieties of pre-registration exist, which roughly correspond to the pre-registration and pre-registration+ badges developed by the Center for Open Science. To qualify for the pre-registration badge, one must fully specify the research design (including study materials) ahead of time. For the pre-registration+ badge, one must additionally specify the planned statistical analyses. Note that directional hypotheses are not required; research can be exploratory in the sense of not having a singular prediction from theory, but confirmatory insofar as statistical tests are clearly specified a priori.

checks may fail, and so forth. Thus, researchers should also consider how, if at all, manipulation checks, confounds, covariates, or unexpected data details will be incorporated into the analysis plan. To disallow post-hoc researcher flexibility while still preparing for such issues, labs may want to develop a standard operating plan to be used to guide decisions for common issues (Lin & Green, 2016). Publicly available pre-registration templates can provide step-by-step guidance (Corker, 2016; van 't Veer & Giner-Sorolla, 2016).

Registration may be done in various ways and places. Because the purpose of preregistration is simply to record one's hypotheses and planned analyses for confirmatory research prior to data collection, on the most informal end of the spectrum, researchers could simply write an e-mail to colleagues or otherwise store preregistration information as part of their regular research note-taking. Ideally, however, this information should be stored in a way that is time-stamped, can be publically shared, and cannot be accidentally lost or changed. Multiple online resources assist preregistration. For example, OSF has a function to freeze and timestamp preregistrations online and provides users with an online link to their pre-registration to include in submitted papers. AsPredicted provides a similar stand-alone service in which researchers answer eight basic research questions (e.g., main question and hypotheses, planned sample, whether any data has already been collected), which are time-stamped, saved, and available for public dissemination if the author chooses (see AsPredicted.org). Examples of public registrations are also available at each of these websites. A major difference between the two systems concerns whether eventual public disclosure of a preregistration is mandatory. At OSF, authors may choose to embargo their registrations for up to 4 years, at which point they become public, whereas AsPredicted registrations may remain private forever. Making registrations public is a standard feature of existing biomedical registries, and it helps to prevent

fraud (multiple alternate registrations). It can also help meta-researchers locate file-drawered studies for inclusion in reviews and meta-analyses. Thus, although some researchers may currently prefer to keep registrations private indefinitely, it is in the long-term interest of the scientific community to have open and accessible registries.

Join a Large-Scale Replication Project

In addition to incorporating open science practices into one's individual research workflow, opportunities exist to join large-scale organized replication projects involving open science. For example, as of the writing of the final edits to this chapter, recruitment was still underway for Many Labs 4, on "variability in replication implementation" (see Klein et al., under review), and there may be more Many Labs projects in the future. There also certainly will be more calls for joining Registered Research Reports (RRRs). As described earlier, in these projects, a team of researchers proposes a study to replicate and develops a protocol to do so. Other teams that then sign on to run the study are provided with clear guidelines on how to preregister and share relevant materials and data openly as part of the project. These RRRs were formerly published in *Perspectives on Psychological Science* but will soon be published in the new journal *Advances in Methodologies and Practices in Psychological Science.* A list of RRR projects seeking participating labs is available at the APS website (see Table 3).

<p align="center">Practicing Open Science: For Teachers</p>

As teachers and mentors, we can teach open science as part of our courses and research-training efforts. For example, some supervisors expect lab members to incorporate open science practices as part of their regular workflow, even using shared lab documents to track project annotation, organization, and sharing.

Basic information about reproducibility and open practices can be integrated into introductory psychology courses. Numerous researchers have developed syllabi for semester-long courses on reproducibility and open science for both the undergraduate and graduate level (Course Syllabi for Open and Reproducible Methods at https://osf.io/vkhbt/ and Modular Course Materials for Research Methods Courses: A SIPS Hackathon Product at https://osf.io/zbwr4/wiki/home/).

Project TIER (Teaching Integrity in Empirical Research) provides resources to help train undergraduate and graduate students on why and how to practice open, replicable research. Developed by economist Richard Ball and librarian Norm Medeiros with the goal of making replication documentation as "ubiquitous and routine as providing a list of references" (Ball & Medeiros, n.d.), Project Tier supports workshops for faculty and students, syllabi and other course materials, and paid fellowships for outreach and curriculum development.

Teachers may also work with students in psychology methods classes to conduct replication studies. Every year over 115,000 undergraduates receive bachelor's degrees in psychology (National Center for Educational Statistics, 2016), and the majority of graduates complete research methods courses or capstone projects as part of their training. Such training provides a vast opportunity to collect replication data which, with reasonable oversight to ensure data quality, could be compiled into large replication databases (Grahe et al., 2012). Teachers who have already begun incorporating replications into their methods classes report that students are excited to contribute to cutting-edge research beyond the classroom, produce quality data, and benefit from focusing on the details of study design and procedures (Frank & Saxe, 2012; Grahe et al., 2012).

Practicing Open Science: For Authors and Reviewers

Editors, publishers, and journals have provided many of the incentives and opportunities to make reforms happen. About a year after publishing "False-Positive Psychology," Simmons et al. (2012), proposed the "21 Word Solution"—a 21-word statement that authors could put in their method sections (and journals could require) to indicate full disclosure of methods. It reads: "We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study." A few journals adopted this procedure. In late 2014, a group of journal editors, funding source representatives, and research experts (largely within the social and behavioral sciences), met and developed the Transparency and Openness Promotion (TOP) Guidelines—a template for concrete policies and procedures to encourage more open science. (See Table 4.) The guidelines cover eight different research components: data citation, data transparency, analytic methods (code) transparency, research materials transparency, design and analysis transparency, preregistration of studies, preregistration of analysis plans, and replication. Within each component, the guidelines differentiate four levels, ranging from zero (journal encourages or says nothing about the research component) to three (the journal generally requires the research component). Journals choose which, if any, component at which level to adopt and can tailor each component to their preferences or needs (Nosek et al., 2015). As of September 2017, over 3,200 journals, including the *Science* and *Nature* family of journals, and about 65 organizations have become signatories of the TOP Guidelines.

We treat authors and reviewers (and editors) in one section here because so many of their concerns interact. But the most important interaction for protecting science is this: Authors should provide all information that reviewers want and need, but authors should be rewarded and not penalized for doing so. "Imperfections" in data or failures to replicate an effect in one of

several studies should not prevent publication; demanding the perfect pure story incentivizes

QRPs and other practices that are bad for science.

-- Table 4 Insert TOP Guidelines About Here --

Choosing Where to Publish (and Share) Research

Authors may consider supporting open science by submitting their work to journals that

endorse, reward, or even require open practices. They might consider whether the journal has

some equivalent to the 21-word statement or is sufficiently far along in the TOP Guideline

standards. Instead of or in addition those practices, other publication outlets have adopted

journal-specific programs or award systems. Several journals, for example, have adopted

"badges" to mark papers that follow open practices. (For a regularly updated list, see OSF's

Badges & Endorsements, Table 3.) For example, *Psychological Science* provides badges—

literally icons posted on the title page of papers—for papers that include open data, open

materials, or preregistration (Eich, 2013). Such incentives can work to encourage the publication

of open science; in the 18 months following the implementation of these badges at *Psychological*

Science, the rate of making data publically available increased over tenfold, from less than 3% to

over 39% (Kidwell et al., 2016).

If authors want to pre-register everything about a study, they may choose to preregister as

part of a registered report (Chambers et al., 2014). Registered reports are journal articles that are

peer-reviewed prior to data collection. As such they are provisionally accepted for publication

based on the soundness of their ideas, hypotheses, methods, and analysis plans rather than their

results. (See Table 3 for a link to a list.)

To reach the widest possible audience, authors may also consider publishing in open

access journals, which remove price and permission barriers (e.g., subscription fees and

copyright restrictions), so that publications can be consumed and shared for free by anyone in the world.[17] The Directory of Open Access Journals (available at doaj.org) lists over 9,000 open-access journals. Some open-access journals are financially supported by institutions, whereas others require authors to pay an article-processing charge. Importantly, the quality of open access journals varies widely, and peer review may be limited at some journals where authors pay for publication. For example, hoax papers containing intentional research impossibilities have been accepted at some open access journals (Bohannon, 2013; N. Gilbert, 2009). Many high-quality journals such as PLoS, BioMedCentral, and University of California Press' *Collabra: Psychology*, however, provide peer-reviewed, high-impact options. Open-access journal articles may even be downloaded and cited more than other articles (Björk & Solomon, 2012; Teplitskiy, Lu, & Duede, 2016; but see Davis, 2011).

Another way to reach a wide audience is to share preprint versions. Even work published in traditional pay-for-access journals can generally be publicly shared in a preprint version—that is, an author-created version that does not incorporate the journal's formatting.[18] Sherpa Romeo (see Table 3) compiles an easily searchable database of specific journals' sharing policies. Free online sharing platforms like ResearchGate.net, Academia.edu, the Social Science Research Network (SSRN), and PsyArXiv make sharing pre-print papers quick and easy.[19] They also often assign papers permanent, citable DOIs and provide space for pre-publication online feedback. PsyArXiv is specifically designed for psychology research and links preprints with supporting materials on OSF.

---

[17] There are often variations in how much sharing is allowed depending on the journal and the source and amount of financial support.

[18] This is, of course, not legal advice.

[19] Critics fear that some of these repositories may become less open as they are acquired by large, profit-driven publishers (e.g., Nagel, 2016). ResearchGate and Academia.edu are for-profit, and SSRN was recently acquired by Elsevier (a large for-profit publisher).

Choosing Where to Review

Reviewers can practice and encourage open science in several different ways. To make a forceful statement about the importance of open science, reviewers could refuse to review for a journal that didn't have sufficiently high open-science standards (e.g., did not ask for the 21-word statement; had not progressed far enough along the TOP guideline standards). Journal reviewers may require that research materials and data be shared as a stipulation of review (Morey et al., 2016). Or upon receiving a manuscript to review, the reviewer might note that there was no 21-word statement or that the authors had not followed open-science practices, and then simply refuse to review the manuscript until the relevant actions were taken. Of course, editors might not have a kindly view of such preemptive tactics.

Researchers can also choose whether to review for journals that publically post reviews or that allow for post-publication review. Public posting of reviews started in the 1990s, and an increasing number of journals give reviewers the option to post reviews alongside published articles. Critiques of *requiring* open posting include that its effect on review quality may be minimal, and that reviewers may be more likely to decline reviewing in the first place (van Rooyen, Delamothe, & Evans, 2010[20]). Arguments for open peer review, however, include that it increases dissemination and discussion of critiques (positive and negative) and incentivizes good reviews (Mueller, 2016; Swoger, 2014). Many editors and authors report that they view open reviews as being helpful (e.g., Callaway, 2016).

There is also growing interest in post-publication review—that is, public review of articles that have already been published. A handful of journals use post-publication review instead of traditional pre-publication review (e.g., *Atmospheric Chemistry and Physics*, *The*

---

[20] However, these authors still argue that the benefits of open review outweigh any burdens.

Winnower), and some argue that this model should replace traditional journal-arbitrated publication decisions because it is more efficient, more accessible, and less arbitrary (Eisen & Vosshall, 2016). Other journals combine pre- and post-publication review. For example *PLoS ONE* has established an online comment system for peer-reviewed-and-published articles. But perhaps most well known are the stand-alone post-publication review repositories such as PubMedCommons and PubPeer. These forums allow for discussion and critique of work published in any journal. Though people may still abuse these forums by flooding them with disrupting, malevolent, or simply low-quality comments (see Stirling, 2015), they have been responsible for uncovering numerous well-known cases of study errors and potential fraud (see, e.g., Cyranoski & Hayden, 2013; Otake, 2014).

Choosing How to Review

Most journals allow authors to choose whether to sign their reviews. However, in pushing for a more open review processes, some advocates have proposed that reviewers be *required* to do so. Such open reviewing might have some benefits for the reviews; for example, reviewers are likely to be more measured in their tone, more justified in their criticisms, less self-promoting, and less ad hominem, if they must attach their names to their reviews. Identification of reviewers could be useful in later discussions between authors and editors. There are a few conflicting studies on whether signing affects review quality (Walsh, Rooney, Appleby, & Wilkinson, 2000, argue that signed reviews are better). However, requiring signed reviews may also have the unintended consequence of affecting people's willingness to review (e.g., van Rooyen, Godlee, Evans, Black, & Smith, 1999).

Of course, signing reviews can have consequences for the reviewer. Revealed reviewers may be thanked by the manuscript's authors, and together they may engage in extended offline

discussion. In fact, one of the authors of this chapter once suggested some potential future studies in a review and then ended up collaborating with the manuscript's author to perform (and publish) them. However, there are reasonable worries that even constructively critical reviews might be received badly, damaging friendships and leaving a bad taste in the mouths of potential future reciprocal reviewers or tenure letter writers. Thus, not surprisingly, only one of us—the tenured full professor—always signs her reviews; the others do not (but expect to in the future).[21] Note that a good way to sign reviews is to write—if true—"I always sign my reviews. <Name>." This signature signals that you don't sign only positive reviews; it also means that unpleasant reviews should not be falsely attributed to you.

More important, of course, is that reviewers must decide what to put into the review and what action to recommend based on it. Editors then must decide what action to take based on the set of reviews from the experts and their own (independent) reading of the manuscript. (We are assuming this model of the reviewer/editor relation for our comments.)

The paper on False-Positive Psychology (Simmons et al., 2011, Table 2) contained four suggestions for reviewers. First, reviewers should "ensure" that authors comply with transparency requirements (e.g., disclose all variables measured and all experimental conditions). Second, they should also recognize that real results are likely to be messy—for instance, it would be (very) unlikely to perfectly replicate a 2 by 2 interaction five times in a row. This suggestion is unique in tone: it notes that if authors are required to reveal all their variables, conditions, measures, and so forth, then reviewers (appropriately) need to be more tolerant of imperfections when making recommendations to the editor. Third, reviewers should ask authors to show that their results are robust to unusual analytic choices (e.g., excluding participants above an arbitrary

---

[21] The tenured professor had a previous rule: Regardless of the evaluation, do not sign reviews of papers by senior people in her field, do sign all other reviews.

age or with reaction times slower than a cutoff chosen post hoc). Finally, reviewers are advised

to demand exact replications if answers to any of the previous inquiries are judged insufficient.

These suggestions are oddly phrased. Yes, it is fine for reviewers to *check* that authors have

followed their requirements for authors. But reviewers cannot *require* anything of authors for

publication—that is the job of the editor. Reviewers could read a manuscript and recommend

revision or rejection to the editor based on the failure to meet the Simmons et al. (2011)

standards. However, to reiterate, if the field is to expect researchers to adopt the new reporting

standards, then reviewers must be more tolerant of imperfections.

Objections to Open Science

Many scientists object to some or all of the reforms desired by the open science

movement; others fully support open science but fear that practicing it may harm them as

individual researchers. These people see various undesirable consequences of the reforms. Some

of those consequences are imagined, but others are quite real and should be addressed by the

scientific community.

Acknowledging and Confronting the Problem

Psychologists (and scientists generally and even the public) have many differing beliefs

about whether there is a replication crisis and what, if anything, should be done about it.

Although a large proportion of the 1,500 respondents in the *Nature* sample (Baker, 2016,

described above) agreed that there was "a significant crisis" (52%) or "a slight crisis" (38%) of

reproducibility in science, and a very small proportion said there was no crisis (3%) or that they

"didn't know" (7%), representative data on researchers' views and knowledge of the current

controversy is scarce. Certainly, some have never heard of the replication crisis, and some likely

have heard about it but ignored or dismissed the information for reasons of limited time or interest.

Other scientists have acknowledged that there are colleagues claiming that there is a replication crisis, but deny its existence, arguing that failures to replicate are what one would expect from normal science (D. Gilbert et al., 2016). Such scientists have declared that it is all science as usual—there have always been failures to replicate, and there is nothing new or remarkably different now (Fiske, Schacter, & Taylor, 2016). Some argue that estimates of QRPs from self-reports are inaccurate (Fiedler & Schwartz, 2016) and that failures to replicate are an acceptable part of a scientific culture that encourages cutting-edge discovery and theory generation (Fiedler et al., 2012).

Other scientists acknowledge that something is amiss, but have recommended that, more or less, we should not air our dirty laundry in public because it would tarnish the reputation of psychology. In remarks at a featured panel at the Society for Personality and Social Psychology, Harry Reis encapsulated these fears:

I was talking to a grants officer not too long ago, who was musing about [having] $10 million to put into a new initiative to develop some field. Where would we put it? And he said, why would I put it in a field where people in the field are saying everything that's being done is garbage and wrong? (Reis, 2017)

This argument was particularly common early on, when, between the publication of Bem's (2011) precognition paper and the revelation of the Stapel fraud, it felt like psychology was particularly in the limelight and looking frail. Worries included implications for future research funding, hiring researchers, credibility of good prior research, and general respect for the field as a science. These worries may have abated somewhat as flaws in other sciences have

been revealed and psychology has taken a leadership role in reform (Chambers, 2017; Spellman, 2015).

Others have pushed back against this perspective, arguing persuasively that the only way for psychology to be seen as a trustworthy field is to increase levels of openness and transparency. Vazire (2017) likens the scientific publishing process to a used-car sales transaction. In both cases, there is the potential for information asymmetry—the used car salesperson (and the researcher) have access to more information than the potential buyer (research consumer). Economists have shown that in cases of information asymmetry, the most likely outcome is devaluation of the end product. If one can't be sure about the quality of the product, that product becomes worth less and less over time. The solution for psychology is clear – we *must* improve our practices, and we *must* increase transparency, so that research consumers don't have to trust us, but can instead see the quality of our work for themselves. Transparency reduces information asymmetry, stopping a negative cycle of decreasing trust and valuation of our work.

A quick look at publications, conferences, blogs, Facebook groups, and other types of discussion arenas, suggests that there might be a generational divide in preference for action— with older more established researchers (in general) urging caution and younger researchers (in general) pushing for change.[22] Such a division should not be surprising to psychology researchers (nor to students of history) given that older people were habituated to the status quo and had succeeded under it (see Watkins, 2017). But such a division creates massive problems for changing incentive structures to favor open science. To change journal, granting agency, and university policies, people in positions of power (i.e., the field's more senior members) at those

---

[22] Susan Fiske and colleagues have data on who writes and comments on methods blogs in her talk at the National Academy of Sciences: `www.youtube.com/watch?v=8ykftugZ44Y`

places must want to effect change.

The Injustice of Rules

Some scientists think that even if there are problems in science, the solutions suggested by the open science movement would hurt science by imposing unnecessary or unrealistic burdens. The typical argument is that the proposed changes are most relevant to, and easily implemented in, relatively simple laboratory studies in social and cognitive psychology. Thus, one fear was that journals would create requirements that other types of studies could not meet. Many of these feared requirements aren't necessarily about open science per se but were about other practices that would improve replicability. For example, some researchers worry that journals (or grant panels) will require specified minimum sample sizes or specified minimum power to publish (or fund) a study. Of course, sometimes populations are so difficult to collect data from (e.g., infants), or limited or hard to find (e.g., prisoners, a distant native population), or the research is so time sensitive (e.g., immediately after a terrorist attack) that perhaps not enough participants can be recruited to run a highly powered study. To the extent that journals would adopt a minimum-subject or minimum-power policy, they would likely simply require that researchers report their sample sizes; report any relevant choices regarding sample sizes such as power analyses, stopping points, and exclusions; and interpret findings appropriately given the relatively small amount of data. A suggestion of minimum sample size was made by Simmons et al. (2011) but was quickly recognized as misguided, even by the authors themselves (Simmons, Nelson, & Simonsohn, in press). Acknowledging the necessary diversity of standards across fields, there is nothing about such factors in the TOP guidelines, and, as far as we know, there are no journals that have required specific minimums without exception.

Another fear is that to be published, a study would be required to have fully open data. Researchers note that sometimes data sets contain a lot of information that is not used in the current paper but will be analyzed for future publications by the collecting authors, who don't want to share all their data—yet. And sometimes data is confidential, so that the authors cannot legally disclose some or all of their data. Most proponents of this proposed reform call it a default, recognizing that not all studies must meet the open data standard, and would straightforwardly allow for exceptions (with justifications) to the default rule. An early study of almost 1,300 psychologists (Fuchs, Jenny, & Fiedler, 2012) asked researchers how they felt about adopting various open science practices; the title of their paper describes their results well: "Psychologists are Open to Change, yet Wary of Rules."

Slows Down Scientific Discovery (and Wastes Time)

Another objection to open science is that by encouraging certain practices it will slow down scientific discovery. For example, by increasing power (typically larger sample sizes), experiments will take more time and money to run. By expecting clearly annotated and publicly posted materials and data, researchers will have additional tasks to complete prior to publication. By encouraging more replications, both of one's own and other people's research, researchers may be unable to complete as many new studies. All these practices involve opportunity costs—the extra resources they require could slow down scientists' pursuit of new research that could lead to new discoveries.

A similar but even bigger fear is that someday *all* psychology studies will be required to be preregistered and that preregistration will eliminate researchers' ability to fully explore their data and publish their interesting-but-not-preregistered findings. As discussed earlier, however, this fear is likely ungrounded. Although preregistration is currently mandatory for government

grant-funded clinical trials (clinicaltrials.gov), no one is arguing for requiring it in psychology[23].

The current proposals simply encourage preregistration as a way of demonstrating that

predictions were made before the data were collected, thus enhancing the epistemic value of the

data and increasing interpretability of inferential statistics.

But would such reforms really slow down scientific discovery or development? First,

there may be ways to limit the extra resources it takes to incorporate these practices. For

example, within-subjects designs and sequential analyses (Lakens, 2014) may decrease the

number of participants needed for high-powered studies. And as discussed earlier, teachers of

methods classes may incorporate publishable replication projects into their regular curriculum.

Second, as noted earlier, researchers can do exploratory analyses of data (like before)—it's just

that such analyses must be noted as exploratory (unlike before).

However, even if these processes do slow down publication, they could still help speed

up scientific development. For example, if open science practices result in fewer false positives,

meta-analyses would be built on stronger evidence, making theory development stand on firmer

ground. It should also help to remove unsupported theories from the literature, thus reducing

time wasted testing unproductive paths and theorizing about statistical noise (Schimmack, 2012).

In a pair of intriguing simulations, Will Gervais investigated the issue of trade-offs. First, he

showed that conducting many small studies ($N = 40$/study) versus a few large studies ($N =$

300/study) with a fixed maximum annual sample size, and under typical conditions in

psychology research, will lead to greater numbers of significant (and, therefore, publishable)

---

[23] Our preregistered prediction might already be on shaky grounds. The National Institutes of Health (NIH) recently announced that all NIH-funded clinical trials are expected to register with clinicaltrials.gov. The definition of clinical trials encompasses experiments that measure health-related, biomedical, or behavioral outcomes. These regulations would sweep in many basic research studies from psychology and neuroscience. https://grants.nih.gov/policy/clinical-trials/reporting/steps.htm

results for the researcher who does many small studies (Gervais, 2014). However, more publishable results do not necessarily mean more true results. A second simulation (Gervais, 2016) showed that the large study (high powered) approach resulted in more "findings" (true positives) than the small study approach (casting a wide net). Thus, as incentives shift and researchers are more concerned about the reproducibility of their findings, we expect that more researchers will choose to pursue a high-powered approach (which will ultimately be more efficient in the long run).

Produces Inequities in Evaluation

In the short run, the adoption of new norms, in any field or endeavor, may create problems and inequalities in the evaluation of people and products. In the case of open science, its procedures do slow down the production and publication of research (e.g., through the push for larger samples), and they are more likely to reveal imperfections in research results (e.g., through author guidelines that require disclosure of variables, conditions, and measures that did not "work"). Ironically, and nonobviously, imperfect results may provide better evidence for a true effect, and perfect results may be evidence of p-hacking (Giner-Sorolla, 2012; Lakens & Etz, 2017; Schimmack, 2012).

Thus, researchers who adopt open science early in their careers are likely to (potentially accurately) believe that they are disadvantaged compared both to peers who have not adopted open science and to predecessors who were not faced with that choice. The early adopters might wonder how their smaller number of publications[24], greater susceptibility to criticism of their publications, and perhaps simply their participation in some organized replication projects, will

---

[24] Note that this concern is sadly not new: "Finally, social psychologists must find ways of adjusting to or moderating academic pressures to publish, at the same time that they reduce their own rate of publishing findings that are so incomplete as to be misleading" (Elms, 1975, p. 975).

be viewed when they go on the job market, come up for tenure, or compete against others for grants or awards.[25]

There have been calls for universities and grant agencies to be sensitive to the new norms (e.g., Chambers, 2017). One suggestion has been that instead of considering the number of publications a researcher has, a committee (e.g., for hiring or tenure) should consider only the researcher's three (for hiring) or seven (for tenure) best papers (https://storify.com/hardsci/n-best-evaluation-of-scientists). Some institutions have already begun to reward or require open-science practices. Other people have disclosed that they, or their department, checks the power and p-values within researchers' publications during hiring or promotion.[26] But whether these or any other procedures to even the playing field will be adopted is still unknown, and it is likely that things will remain in flux for some time.

… And Wouldn't Help With Some Other Related Concerns

A final objection to implementing open science practices is that with all the additional work it entails, it will still prove ineffective at catching clever fraudsters. That may be true. But it certainly could help catch less clever ones. Finding unlikely patterns in researchers' data facilitated the discovery of fraud in several cases (e.g., Simonsohn, 2013). Looking at full data sets can be quite revealing. However, one can imagine that a very clever fraudster could create realistic data. A deceptive fraudster could "preregister" experimental hypotheses *after* running the study—thus being open to HARKing, exactly what reforms were supposed to help us avoid. However, even if it hardly ever helped with fraud, open science isn't a plan to uncover the (probably) rare cases of fraud. Rather, open science can help protect against the wider intended

---

[25] We mention possible benefits to citations and recognition of open science in the section on Practicing Open Science: For Authors.

[26] Disclosed by some faculty in business schools.

and unintended uses of questionable research practices (QRPs). That reduction, plus the other

reforms described earlier, should lead to more robust and replicable science.

The Future of Open Science

What Is Here to Stay

Right now (in late 2017), it is difficult to imagine that some of the basic reforms of the

open science movement will fall by the wayside—especially the ones that capture the

fundamental values of science. For example, open methods—the idea that as a default, scientists

should make the descriptions of their methods as complete as possible, including providing

verbatim instructions, videos, vignettes, scales, and the like—seems likely to become nearly

ubiquitous. (The TOP guidelines include levels for citing such contributions.) Scientists should

reveal the conditions, measures, and procedures that "didn't work" and the reasons why subjects

were excluded from analyses (Simmons, Nelson, & Simonsohn, 2011). None of these processes

should require (much) additional labor, but they do provide important information for future

researchers who desire to replicate, understand, or evaluate the research.

Increasingly, it appears that open data in some form will soon become a given. Funders (e.g., the

Gates Foundation) and societies (e.g., the German Psychological Society, DGPs; see Schönbrodt,

Gollwitzer, & Abele-Brehm, 2017) are already starting to require or strongly encourage sharing

of data. Given the increasing ease of storing and sharing data files, we expect this trend will

continue.

What Is Still to Come

Even as we have been working on this chapter, things that we thought were in the future

have already come to pass. Among the things we are hoping and expecting to see is for the open

science movement to standardize, simplify, and streamline by (among other things) refining

procedures, articulating values, and creating technology.

A major overarching issue is that open science still resembles the Wild West (or for non-

U.S. readers, a place where there are different rules and enforcers in every town, and no one

knows who is in charge overall). The standards for the different aspects of open science are often

not clear. For example, what data really needs to be shared (Only specific published variables?

Also potential moderators/confounds?); when must they be shared (Only after the collector has

published? After a certain set time?); and exactly how easily reusable must it be (Are authors

ethically required to clean and annotate?)? Similarly, exactly what needs to be included to call

something a pre-registration? Just as APA style (APA, 2014b) standardizes what sections appear

in empirical articles, we can imagine that there will be certain standardized ways of

preregistering. It is likely that repositories like OSF will drive this effort. Some people might feel

uncomfortably uncertain about how to proceed and find it easier to navigate if there were more

clear expectations or standards.

In order to decide what those standards should be, the field needs to clarify the purpose of

open science within psychology. We have focused on it as a tool to ensure that published

research is reliable and that it can be used by other researchers (e.g., in meta-analyses). But as a

field we may want to discuss whether it has other core values for us—for example, to

disseminate science more broadly, speed up progress, encourage citizen science among non-

professionals, and so forth.

And when the field is clear on which benefits of open science it values most, it can

determine the best incentive structures to reward open science practices. For example, how

should such practices be valued not only at research institutions but also at teaching-focused

institutions? How can open science ensure that people who collect really resource-intensive data get rewarded for it? Should we create some new "impact index" that weights shared data and materials, and use by others of those shared things, along with number of publications and article citations?

Finally, new, useful, technology for practicing open science is invented every month. But more is still needed (especially things that make open science easier to practice). As of now, for example, some researchers still don't have good tools to deal with sharing really massive data files; even posting hours of video online is probably challenging (but see Databrary, 2012). And we need more ways to curate and aggregate the vast amount of information we are generating into forms that would be useful in furthering open science. Some such next steps are described below.

Moving Even Further Forward

Even if fully implemented to Stage 3 (the strictest level), the TOP guidelines do not address some of the other related problems that we see as slowing the overall progress of psychological science. In particular, as Spellman (2015) noted, the suggested procedures almost all address improving the reliability and reproducibility of individual studies—what she referred to as "making better bricks." What the various open science initiatives are not doing (yet), is impelling the *connection* of such better bricks: they are neither facilitating the ordered accumulation of information nor helping to create links between studies in and across fields. These are important activities for theory development. They are also important for improving scientific self-correction—devaluing and discarding less-supported theories (our field is not good at this process; Ferguson & Heene, 2012; Greenwald, 2012).

Extending open science should make us better at not only keeping but also aggregating information. Making information that is relevant to assessing studies more open has already begun. Some journals have lessened restrictions on the length of method sections either by dropping word counts or adding options for online supplemental materials. (And, of course, some journals now require full method details and something like the 21-word statement.) Some journals now also ask that researchers reveal and describe all of the studies in a series of about-to-be-published research—that is, to reveal the studies that "don't work" as well as the ones that do. And researchers need places to make available studies like failed, or successful, simple replications (e.g., like psychfiledrawer.org)—studies that are not "publishable" in the current sense but that would reduce the file drawer problem and be useful in meta-analyses.

As a greater proportion of relevant results become findable, meta-analyses should become more complete and more prevalent. They should not be practiced by only a few people but should be taught more broadly and used more consistently. The field needs to agree on "best practices" within meta-analyses and needs better ways to call for papers to include in meta-analyses. The current process is to search the published research using some specific terms (see problems with that later) and to put out a notice on various list serves or websites. For various purposes, there should be a more general repository site regarding meta-analyses in progress. There should also be a more systematic way to save the work that goes into meta-analyses (see Braver, Thoemmes, & Rosenthal, 2014; Lakens, Hilgard, & Staaks, 2016), so that others may access the database to check it and build on it. The open science of meta-analysis should be thought of in analogous ways to the open science for individual studies. (A goal might be to devise TOP guidelines for meta-analyses similar to those for standard empirical research.) And, perhaps papers that include multiple studies of the same basic effect should include within-paper

meta-analyses—this can help readers understand how mixed or "imperfect" results in aggregate can actually provide good evidence when there is no *p*-hacking (Lakens & Etz, 2017).

The field also needs to do a better job connecting research. As the body of empirical results grows, there should be better ways to access and use literature related to our own empirical, review, and theory projects. Researchers need to find relevant research, but the keyword system has been broken by idiosyncrasies, and researchers rely too heavily on literal word searches that miss relevant research if the same terms are not used by different authors in their manuscripts (see, e.g., Ranganath, Spellman, & Joy-Gaba, 2010).

Citations within papers should provide information about why another paper is being cited. There are many reasons to cite, for example, general background, methods, and—important for creating a cumulative science—acknowledging whether the findings are consistent or inconsistent with the previous findings or theories described. Conference speakers are fond of noting that retracted papers are often cited as heavily as they were before they were retracted—and sometimes even more so. But the (long) list of citing papers does not reveal why. As authors we could annotate our in-text references with reasons for citing them; acquiring those annotations in a citation search will help us select which papers we need to read and will help the field to uncover whether those papers are being disconfirmed, confirmed, limited, or expanded, by subsequent papers (as is done for appellate judicial opinions in the United States; Anicich, 2014; Spellman, 2012b). And maybe with those annotations, we could devise a reproducibility index (Oransky, 2013) that grades journals on how well the articles they publish stand up to replication attempts.

Conclusion

We view the open science movement as a return to implementing the values that science

had back when it was simpler and slower and the community of scientists was smaller. The

argument that open science will *slow down* science might be correct, but it will certainly not

prevent discovery. In many cases, it might even accelerate scientific progress. What is important

is for it to slow down the rate and decrease the proportion of "false discoveries" (i.e., false

positives) and help the field rid itself of long-standing misconceptions based on unreplicable

findings. More reliable individual findings should help create a science that gives us more solid

grounding for exporting our findings to education, business, law, and other important applied

domains, and for a sounder theoretical understanding of mind, brain, and behavior.

Definitions and Terms

21-word solution (or 21-word statement): a statement put in a method section to indicate full disclosure. It reads: "We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study."

confirmatory versus exploratory analysis: confirmatory research tests *a priori* hypotheses that are specified before data collection; exploratory research creates new *a posteriori* hypotheses based on examining already-collected data.

"file drawer" effect: phenomenon in which research is not published—often because results are null. This makes cumulative or self-correcting science difficult and likely leads to an over-representation of false positive findings in published literature.

"HARKing" (hypothesizing after results are known): describing a result derived from data exploration as though it had been predicted from the outset.

open science: a collection of actions designed to make scientific processes more transparent and their results more accessible.

Open Science Framework (OSF): a free online platform for researchers to store their research materials in one location. OSF has a function to freeze and time-stamp pre-registrations online, allows users to share materials with specific collaborators or the public generally, and provides users with online links to their pre-registration and other materials to include in submitted papers.

p-**hacking:** analyzing data multiple ways (often not consistent with the original hypothesis) in order to obtain a statistically significant result.

questionable research practices (QRPs): behaviors during the research process that inflate the chance of a false-positive finding. Examples include failing to report inconsistent or null

findings; excluding data points such as outliers, or running additional statistical analyses based on post hoc criteria in order to find statistically significant results ("p-hacking"); running additional participants to reach statistical significance ("data peeking"); and reporting unexpected findings as though they were hypothesized from the start ("HARKing").

Registered Replication Reports (RRR): preapproved, large-scale replications in which multiple labs replicate a single study, usually with input from the original author. Study methods, materials, and analysis plans are approved and registered by the publishing journal prior to data collection.

Transparency and Openness Promotion (TOP) guidelines: a template for concrete policies and procedures that journals can use when implementing open science standards. See Table 4.

List of Abbreviations

APA = American Psychological Association

APS = Association for Psychological Science

BITSS = Berkeley Initiative for Transparency in the Social Sciences

COS = Center for Open Science

DOI = digital object identifier

HARKing = hypothesizing after results are known

OA = open access

OS = open science

OSF = Open Science Framework

PRO = Peer Reviewers' Openness (Initiative)

QRP = questionable research practice

RRR = Registered Replication Report

TOP = Transparency and Openness Promotion

References

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Bornstein, B., . . . Zwaan, R. A. (2014). Contribution to Alogna et al (2014). Registered replication report: Schooler & Engstler-Schooler (1990). *Perspectives on Psychological Science, 9*(5), 556–578. 10.1177/1745691614545653

American Psychological Association. (2010a). *Ethical principles of psychologists and code of conduct.* Retrieved from http://apa.org/ethics/code/index.aspx

American Psychological Association. (2010b). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Anicich, E. M. (2014). What lies within: Superscripting references to reveal research trends. *Perspectives on Psychological Science*, *9*, 682–691.

Axt, J. R. (2016, January). So you've decided to do more open science [PowerPoint slides]. In K. Corker (chair), *Translating open science into daily practice.* Society for Personality and Social Psychology, San Diego, CA. Retrieved from https://osf.io/buvwf/

Ball, R. & Medeiros, N. (n.d.). Project Tier Mission Statement. Retrieved from www.projecttier.org/about/mission-statement/

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533,* 452–454. doi:10.1038/533452a

Bastian, H. (2016, December 5). *Reproducibility crisis timeline: Milestones in tackling reproducibility [*Blog post]. Retrieved from http://blogs.plos.org/absolutely-maybe/2016/12/05/reproducibility-crisis-timeline-milestones-in-tackling-research-reliability/

Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology, 66,* 153–158.

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature, 483,* 531–533.

Belter, C. W. (2014). Measuring the value of research data: A citation analysis of oceanographic data sets. *PLOS ONE*, *9*(3), e92590. doi:10.1371/journal.pone.0092590

Bem, D. J. (2003). Writing the empirical journal article. In J. M. Darley, M. P. Zanna, & H. L. Roediger, III (Eds.), *The compleat academic: A practical guide for the beginning social scientist* (2nd ed., pp. 171–201). Washington, DC: American Psychological Association.

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.

Bishop, D. (2013, July 26). *Why we need pre-registration* [Blog post]. Retrieved from http://deevybee.blogspot.com/2013/07/why-we-need-pre-registration.html

Björk, B. C., & Solomon, D. (2012). Open access versus subscription journals: A comparison of scientific impact. *BMC Medicine, 10*(1), 73. doi:10.1186/1741-7015-10-73

Bohannon, J. (2013). Who's afraid of peer review? *Science, 342,* 60–65.
    doi:10.1126/science.342.6154.60

Boulton, G., Campbell, P., Collins, B., Elias, P., Hall, W., Laurie, G., . . . Walport, M. (2012).
    Science as an open enterprise. The Royal Society Science Policy Centre Report 02/12.
    Retrieved from
    https://royalsociety.org/~/media/policy/projects/sape/2012-
    06-20-saoe.pdf

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis
    and replicability. *Perspectives on Psychological Science, 9,* 333–342.

Callaway, E. (2016). Open peer review finds more takers: Journal experiments and surveys
    suggest scientists are coming round to the idea of publishing review reports. *Nature, 539,*
    343. Retrieved from www.nature.com/news/open-peer-review-finds-
    more-takers-1.20969

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., . . . Hang, W.
    (2016). Evaluating replicability of laboratory experiments in economics. *Science,
    351*(6280), 1433–1436. doi:10.1126/science.aaf0918

Campbell, L. (2015; last update 2016). Campbell Lab: OSF Research Milestones / Disclosure
    Templates. https://osf.io/m7f8d/

Campbell, L., Loving, T. J., & LeBel, E.P. (2014). Enhancing transparency of the research
    process to increase accuracy of findings: A guide for relationship researchers. *Personal
    Relationships, 21*(4), 531–545. doi:10.1111/pere.12053

Capaldi, C. A. (2015, May). *Graduating from undergrads: Are MTurk workers less attentive
    than undergraduate participants?* Poster presented at the 4th Annual Psychology Outside
    the Box Conference, Ottawa, Ontario.

Chambers, C. D. (2017). *The seven deadly sins of psychology: A manifesto for reforming the
    culture of scientific practice*. Princeton, NJ: Princeton University Press.

Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex [Editorial].
    Cortex, *49*(3), 609–610. doi:10.1016/j.cortex.2012.12.016

Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., & Etchells, P. J. (2014). Instead of
    "playing the game" it is time to change the rules: Registered reports at *AIMS
    Neuroscience* and beyond. *AIMS Neuroscience, 1,* 4–17.

Chen, R. S., Downs, R. R., & Schumacher, J. (2013, December). *Assessing the interdisciplinary
    use of socioeconomic and remote sensing data in the earth sciences.* Presentation at the
    Annual Meeting of the American Geophysical Union, San Francisco, CA.

Chang, A. C., & Li, P. (2015). Is economics research replicable? Sixty published papers from
    thirteen journals say "usually not." Finance and Economics Discussion Series 2015-083.
    Washington, DC: Board of Governors of the Federal Reserve System.
    doi:10.17016/FEDS.2015.083

Cheung, I., Campbell, L., & LeBel, E. P., Ackerman, R. A., Aykutoğlu, B., Bahník, Š., . . . Yong, J. C. (2016). Registered Replication Report: Study 1 From Finkel, Rusbult, Kumashiro, & Hannon (2002). *Perspectives on Psychological Science, 11*(5), 750–764. doi:10.1177/1745691616664694

Christensen, G. (2016). *Manual of best practices in transparent social science research.* Retrieved April 2, 2017, from www.bitss.org/education/manual-of-best-practices/

Cohen, J. (1994). The earth is round (*p* < .05). *American Psychologist, 49,* 997–1003.

Committee on Professional Conduct of Harvard's Faculty of Arts and Sciences. (2010). *Report of Investigating Committee following an Inquiry to assess allegations of scientific misconduct by Professor Marc D. Hauser (Psychology).* Retrieved from http://cache.boston.com/news/pdfs/harvardreport.pdf?p1=Article_Related_Box_Article

Corker, K. S. (2016, January 25). *Preregistration: Best project ever* [Template]. Retrieved from https://osf.io/h48uw/

Corker, K. S., Donnellan, M. B., Kim, S. Y., Schwartz, S. J., & Zamboanga, B. L. (2015). College student samples are not always equivalent: The magnitude of personality differences across colleges and universities. *Journal of Personality.* doi:10.1111/jopy.12224

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Crosas, M., King, G., Honaker, J., & Sweeney, L. (2015). Automating open science for big data. *The ANNALS of the American Academy of Political and Social Science, 659*(1), 260–273. doi:10.1177/0002716215570847

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60,* 170–180.

Cyranoski, D. & Hayden, E. C. (2013, May 23). Stem-cell cloner acknowledges errors in groundbreaking paper: Critics raise questions about rush to publication. *Nature News.* doi:10.1038/nature.2013.13060

Davis, P. M. (2011). Open access, readership, citations: A randomized controlled trial of scientific journal publishing. *The FASEB Journal, 25*(7), 2129–2134. doi:10.1096/fj.11-183988

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*(1), 130-147. doi: 10.1037/0022-3514.86.1.130

Dewald, W. G., Thursby, J. G., & Anderson, R. G. (1986). Replication in empirical economics: The journal of money, credit and banking project. *The American Economic Review, 76*(4), 587–603. Retrieved from www.jstor.org/stable/1806061

Destro Bisol, G., Anagnostou, P., Capocasa, M., Bencivelli, S., Cerroni, A., Contreras, J. L., ... Boulton, G. (2014). Perspectives on Open Science and scientific data sharing: an

interdisciplinary workshop. *Journal of Anthropological Sciences, 92,* 179–200. doi:10.4436/JASS.92006

Donnellan, M. B., & Conger, R. D. (2007). Designing and implementing longitudinal studies. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.) *Handbook of research methods in personality psychology* (2nd ed., pp. 21–36). New York, NY: Guilford Press.

Donnellan, M. B., Lucas, R. E., & Cesario, J. (2015). On the association between loneliness and bathing habits: Nine replications of Bargh and Shalev (2012) Study 1. *Emotion, 15,* 109–119.

Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind?. *PLOS ONE* 7(1): e29081. doi:10.1371/journal.pone.0029081

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., . . . Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology, 67,* 68–82. doi:10.1016/j.jesp.2015.10.012

Ebersole, C. R. (2017, Winter). Open science as both prosocial and proself. *Psi Chi, 21*(2), 26–29. Retrieved from `www.psichi.org/page/212EyeWin17eEbersole#.WIlk-vkrI2z`

Eerland, A., Sherrill, A. M., Magliano, J. P., Zwaan, R. A., Arnal, J. D., Aucoin, P., . . . Prenoveau, J. M. (2016). Registered replication report: Hart & Albarracín (2011). *Perspectives on Psychological Science, 11*(1), 158–171. doi:10.1177/1745691615605826

Eich, E. (2014). Business not as usual. *Psychological Science, 25,* 3–6.

Eisen, M., & Vosshall, L. B. (2016, January 21). *Coupling pre-prints and post-publication peer review for fast, cheap, fair, and effective science publishing* [Blog post]. Retrieved from `www.michaeleisen.org/blog/?p=1820`

Elms, A. C. (1975). The crisis of confidence in social psychology. *American Psychologist, 30,* 967.

Enserink, M. (2012, June 25). Rotterdam marking psychologist resigns after university investigates his data. *Science.* Retrieved from `www.sciencemag.org/news/2012/06/rotterdam-marketing-psychologist-resigns-after-university-investigates-his-data`

Etz, A., & Vandekerckhove, J. (2016). A Bayesian perspective on the reproducibility project: Psychology. *PLOS ONE, 11*(2), e0149794. doi:10.1371/journal.pone.0149794

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90,* 891–904. doi:10.1007/s11192-011-0494-7

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science, 7,* 555–561.

Feynman, R. P. (1986). Personal observations on the reliability of the shuttle. *Report of the Presidential Commission on the Space Shuttle Challenger Accident, 2,* 1–5. Retrieved from `www2.lns.mit.edu/fisherp/Appendix-F.txt`

Feynman, R. P. (1974). Cargo cult science. *Engineering and Science, 37,* 10–13.

Fiedler, K. (2011). Voodoo correlations are everywhere—not only in neuroscience. *Perspectives on Psychological Science, 6,* 163–171.

Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The long way from α-error control to validity proper: Problems with a short-sighted false-positive debate. *Perspectives on Psychological Science, 7,* 661–669.

Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science, 7*(1), 45–52. doi:10.1177/1948550615612150

Finkel, E. J., Eastwick, P. W., & Reis, H. T. (2015). Best research practices in psychology: Illustrating epistemological and pragmatic considerations with the case of relationship science. *Journal of Personality and Social Psychology, 108*(2), 275–297. doi:10.1037/pspi0000007.

Fiske, S. T., Schacter, D. L., & Taylor, S. E. (2016). Introduction. *Annual Review of Psychology, 67,* v. doi:10.1146/annurev-ps-67-121415-100001

Fouchier, R. A. M., García-Sastre, A., Kawaoka, Y., Barclay, W. S., Bouvier, N. M., Brown, I. H., . . . Cox, N. J. (2013). Transmission studies resume for avian flu. *Nature, 493*(7434), 609–609. doi:10.1038/nature11858

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science, 345*(6203), 1502–1505. doi:10.1126/science.1255484

Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science*, *7*(6), 600–604. doi:10.1177/1745691612460686

Fuchs, H. M., Jenny, M., & Fiedler, S. (2012). Psychologists are open to change, yet wary of rules. *Perspectives on Psychological Science, 7,* 639–642.

Gelman, A., & Loken, E. (2014). The statistical crisis in science: Data-dependent analysis—a "garden of forking paths"—explains why many statistically significant comparisons don't hold up. *American Scientist, 102*(6), 460. doi:10.1511/2014.111.460

Gervais, W. (2014). `http://willgervais.com/blog/2014/3/5/more-power`

Gervais, W. (2016). `http://willgervais.com/blog/2016/2/10casting-a-wide-net`

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science." *Science*, *351*(6277), 1037a. doi:10.1126/science.aad7243

Gilbert, N. (2009, June 15). Editor will quit over hoax paper: Computer-generated manuscript accepted for publication in open-access journal. *Nature News.* doi:10.1038/news.2009.571

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.

Grahe, J. E., Reifman, A., Hermann, A. D., Walker, M., Oleson, K. C., Nario-Redmond, M., & Wiebe, R. P. (2012). Harnessing the undiscovered resource of student research projects. *Perspectives on Psychological Science, 7*(6), 605–607. doi:10.1177/1745691612459057

Greenwald, A. G. (2012). There is nothing so theoretical as a good method. *Perspectives on Psychological Science, 7,* 99–108.

Hagger, M. S., Chatzisarantis, N. L., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., . . . Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science, 11*(4), 546–573. doi:10.1177/1745691616652873

Harrell, F. E. (2014, October). *Reproducible statistical reporting with R, LaTex, and knitr.* Presentation at the Novartis Biostatistics Conference, East Hanover, NJ. Retrieved from `http://biostat.mc.vanderbilt.edu/wiki/pub/Main/SweaveLatex/fhsweave.pdf`

Harris, C. R., Coburn, N., Rohrer, D., & Pasher, H. (2013). Two failures to replicate high performance goal priming effects. *PLOS ONE* 8(8): e72467. doi:10.1371/journal.pone.0072467

IJzerman, H., Pollet, T. V., Ebersole, C. R., & Kun, D. (2017, June 29). What predicts Stroop performance? A conditional random forest approach. Retrieved from psyarxiv.com/scpe2. doi: 10.17605/OSF.IO/SCPE2

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124. doi:10.1371/journal.pmed.0020124

Ioannidis, J. P. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science, 7,* 645–654.

Jacoby, W. G. (2015, March 25). *American Journal of Political Science* guidelines for replication files. Retrieved from `https://ajpsblogging.files.wordpress.com/2015/03/ajps-guide-for-replic-materials-1-0.pdf`

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. doi:10.1177/0956797611430953

Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014). Does cleanliness influence morality judgments? A direct replication of Schnall, Benton, and Harvey (2008). *Social Psychology, 45*(3), 209–215.

Kahneman, D. (2012). *A proposal to deal with questions about priming effects* [Open letter]. Retrieved from `www.nature.com/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman%20Letter.pdf`

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus & Giroux.

Kaiser, J. (2017, January 18). Rigorous replication effort succeeds for just two of five cancer papers. *Science.* doi:10.1126/science.aal0628

Kaiser, J. (2017b, June 27). Cancer studies pass reproducibility test. *Science.* doi: 10.1126/science.aan7016

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217.

Kidwell, M. C., Lazarevic, 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), e1002456. doi:10.1371/journal.pbio.1002456

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., . . . Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology, 45,* 142–152. doi:10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F. Adams, B. G., Adams, R. B., Alper, S., . . . Friedman, M. (under review). *Many Labs 2: Investigating variation in replicability across sample and setting.* Retrieved from `https://osf.io/8cd4r/`.

Klein, R. A., Ebersole, C., Atherton, O., Nosek, B., Vaughn, L. A., Schmidt, K. . . . Ratliff, K. (in prep). Many Labs 4: Variability in replication implementation. Retrieved from `https://osf.io/ph29u/`

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology, 44,* 701–710. doi:10.1002/ejsp.2023

Lakens, D., & Etz, A. J. (2017). Too true to be bad: When sets of studies with significant and non-significant findings are probably true. *Social Psychological and Personality Science*, 1–7. doi:10.1177/1948550617693058.

Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychology, 4*(1), 24. doi:10.1186/s40359-016-0126-3

LeBel, E. P., & Campbell, L. (2013). Heightened sensitivity to temperature cues in individuals with anxious attachment: Real or elusive phenomenon? *Psychological Science, 24,* 2128–2130.

LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology, 15,* 371–379.

Ledgerwood, A. (2016). Ledgerwood Lab Experiment Archive Form. https://ucdavis.app.box.com/s/f8hn7rqtwwf6aa6hjtkthdbiuehup312

Ledgerwood, A., & Sherman, J. W. (2012). Short, sweet, and problematic? The rise of the short report in psychological science. *Perspectives on Psychological Science, 7,* 60–66.

Lehrer, J. (2010, December 13). The truth wears off: Is there something wrong with the scientific method? *The New Yorker*. Retrieved from `http://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off`

Levelt Committee, Noort Committee, & Drenth Committee. (2012). *Flawed science:* The fraudulent research practices of social psychologist Diederik Stapel. Retrieved from

https://www.tilburguniversity.edu/upload/3ff904d7-547b-
40ae-85fe-
bea38e05a34a_Final%20report%20Flawed%20Science.pdf

Lin, W., & Green, D. P. (2016). Standard Operating Procedures: A safety net for pre-analysis plans. *PS: Political Science & Politics, 49*(03), 495–499. doi:10.1017/S1049096516000810

Luttrell, A., Petty, R. E., & Xu, M. (2017). Replicating and fixing failed replications: The case of need for cognition and argument quality. *Journal of Experimental Social Psychology, 69,* 178–183.

Lupia, A. (2014). What is the value of social science? Challenges for researchers and government funders. *PS: Political Science & Politics, 47,* 1–7.

Lykken, D. E. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70,* 151–159.

Lynott, D., Corker, K. S., Wortman, J., Connell, L., Donnellan, M. B., Lucas, R. E., & O'Brien, K. (2014). Replication of "Experiencing physical warmth promotes interpersonal warmth" by Williams and Bargh (2008). *Social Psychology, 45,* 216–242.

MacCoun, R. J., & Perlmutter, S. (2017). Blind analysis as a correction for confirmatory bias in physics and in psychology. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 297–322). Hoboken, NJ: Wiley.

Marsman, M., Schönbrodt, F. D., Morey, R. D., Yao, Y., Gelman, A., & Wagenmakers, E. J. (2017). A Bayesian bird's eye view of "Replications of important results in social psychology." *Royal Society Open Science, 4*(1), 160426. doi:10.1098/rsos.160426

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research how often do they really occur? *Perspectives on Psychological Science, 7*(6), 537–542. doi:10.1177/1745691612460688

Matzke, D., Nieuwenhuis, S., van Rijn, H., Slagter, H. A., van der Molen, M. W., & Wagenmakers, E. J. (2015). The effect of horizontal eye movements on free recall: A preregistered adversarial collaboration. *Journal of Experimental Psychology: General, 144*(1), e1–e15. doi:10.1037/xge0000038

McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2006). Lessons from the JMCB Archive. *Journal of Money, Credit, and Banking, 38*(4), 1093–1107. doi:10.1353/mcb.2006.0061

McDonald, M. M., Donnellan, M. B., Lang, R., & Nikolajuk, K. (2014). Treating prejudice with imagery: Easier said than done? *Psychological Science, 25,* 837–839.

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., & Yarkoni, T. (2016). How open science helps researchers succeed. *Elife, 5,* e16800. doi:10.7554/eLife.16800

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations.* Chicago, IL: University of Chicago Press.

Meehl, P. E. (1967). Theory testing in psychology and physics: A methodological paradox. *Journal of Consulting and Clinical Psychology, 46,* 806–834.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., ... Van der Laan, M. (2014). Promoting transparency in social science research. *Science, 343*(6166), 30–31. doi:10.1126/science.1245317

Miller, D. T., & Prentice, D. A. (2013). Psychological levers of behavior change. In E. Shafir (Ed.) *The behavioral foundations of public policy* (pp. 301–309). Princeton, NJ: Princeton University Press.

Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., ... Vanpaemel, W. (2016). The Peer Reviewers' Openness Initiative: Incentivizing open research practices through peer review. *Royal Society Open Science, 3*(1), 1–7. doi:10.1098/rsos.150547

Mueller, A. (2016, May 16). *The case for open review* [Opinion post]. Retrieved from www.insidehighered.com/views/2016/05/16/open-peer-review-journal-articles-offers-significant-benefits-essay

Nagel, D. (2016, May 18). Elsevier stirs up controversy with SSRN Acquisition. *Campus Technology.* Retrieved from https://campustechnology.com/Articles/2016/05/18/Elsevier-Buys-Up-SSRN-Stirs-Up-Controversy.aspx?Page=1

National Center for Educational Statistics. (2016). Table 322.10. Bachelor's degrees conferred by postsecondary institutions, by field of study: Selected years, 1970–71 through 2014–15. Retrieved from https://nces.ed.gov/programs/digest/d16/tables/dt16_322.10.asp?current=yes

National Science Foundation, National Center for Science and Engineering Statistics. (2015). *Women, minorities, and persons with disabilities in science and engineering: 2015.* (Special Report NSF 15-311). Retrieved from http://www.nsf.gov/statistics/wmpd/

Nature Publishing Group. (2013). Journal policies: Availability of data, materials, and methods. Available at http://www.nature.com/authors/policies/availability.html#requirements

Nature Publishing Group. (2014). Scientific Data *launches with first Data Descriptors* [Press release]. Retrieved from: http://www.nature.com/press_releases/scientific-data.html

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science, 348,* 1422–1425. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4550299/

Nosek, B. A., & Gilbert, E. A. (2016, March 7). Let's not mischaracterize the replication studies. *Retraction Watch.* Retrieved from

http://retractionwatch.com/2016/03/07/lets-not-mischaracterize-replication-studies-authors/

Nosek, B.A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45,* 137–141. doi:10.1027/1864-9335/a000192

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7,* 615–631.

Nuzzo, R. (2015). Fooling ourselves. *Nature, 526*(7572), 182–185. doi:10.1038/526182a

NSF (2015). Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science: *Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences.* Retrieved from: http://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf

O'Boyle, E. H., Banks, G. C., & Gonzalez-Mulé, E. (2014). The Chrysalis Effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management, 43*(2), 376–399. doi:10.1177/0149206314527133

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349,* 6251. doi:10.1126/science.aac4716

Open Science Framework (2015). Course syllabi for open and reproducible methods. Retrieved from https://osf.io/vkhbt/

Oransky, I. (2013, July 8). *Time for a scientific journal Reproducibility Index.* Retrieved from: http://retractionwatch.com/2013/07/08/time-for-a-scientific-journal-reproducibility-index/

Otake, T. (2014, April 20). "STAPgate" shows Japan must get back to basics in science: Misconduct, and covering it up, common in nation's labs: Experts. *The Japan Times Online.* Retrieved from www.japantimes.co.jp/news/2014/04/20/national/stapgate-shows-japan-must-get-back-to-basics-in-science/#.WJJcD1UrJtS

Overbye, D. (2005, September 13). One find, two astronomers: An ethical brawl. *New York Times.* Retrieved from http://www.nytimes.com/2005/09/13/science/space/one-find-two-astronomers-an-ethical-brawl.html

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PLOS ONE, 7*(8), e42510. doi:10.1371/journal.pone.0042510

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the Special Section on Replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528–530. doi:10.1177/1745691612463401

Patrone, D., Resnik, D., & Chin, L. (2012). Biosecurity and the review and publication of dual-use research of concern. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science, 10*(3), 290–298. doi:10.1089/bsp.2012.0011.

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ, 1*, e175. doi:10.7717/peerj.175

Popper, K. R. (1959). *The logic of scientific discovery*. London, England: Hutchinson.

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.

Ranganath, K. A., Spellman, B. A., & Joy-Gaba, J. A. (2010). Cognitive "category-based induction" research and social "persuasion" research are each about what makes arguments believable: A tale of two literatures. *Perspectives on Psychological Science, 5,* 115–122.

Reis, H. (2017). Speaker in symposium: *What now? Moving past the replication debate. A conversation about the future of social and personality psychology.* Society for Personality and Social Psychology. www.youtube.com/watch?v=MUrnJ8j7c7o

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

Rouder, J. N. (2016). The what, why, and how of born-open data. *Behavior Research Methods, 48*(3), 1062–1069. doi:10.3758/s13428-015-0630-z

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods, 17*(4), 551–566. doi:10.1037/a0029487

Schnall, S. (2014, November 18). *Social media and the crowd-sourcing of social psychology* [Blog post]. Retrieved from www.psychol.cam.ac.uk/cece/blog

Schönbrodt, F., Gollwitzer, M., & Abele-Brehm, A. (2017). Der Umgang mit Forschungsdaten im Fach Psychologie: Konkretisierung der DFG-Leitlinien. *Psychologische Rundschau, 68,* 20–35. doi:10.1026/0033-3042/a000341. English version available at https://osf.io/preprints/psyarxiv/vhx89

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., Kavvadia, F., & Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *PlOS ONE, 8*(4), e56515. doi:10.1371/journal.pone.0056515

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., … Nosek, B. A. (2015). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. Retrieved from https://osf.io/j5v8f

Simmons, J. P., Nelson, L. N., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue, The Official Newsletter of the Society for Personality and Social Psychology, 26*(2), 4–7. Retrieved from http://ssrn.com/abstract=2160588.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (in press). False-positive citations. *Perspectives on Psychological Science*. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2916240

Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An introduction to Registered Replication Reports at *Perspectives on Psychological Science*. *Perspectives on Psychological Science, 9,* 552–555. doi:10.1177/1745691614543974

Simonsohn, U. (2013). Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological Science, 24,* 1875–1888.

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science, 3*, 160384. doi:10.1098/rsos.160384

Spellman, B. A. (2012a). Introduction to the Special Section: Data, data, everywhere… Especially in my file drawer. *Perspectives on Psychological Science, 7,* 58–59. doi:10.1177/1745691611432124

Spellman, B. A. (2012b). Introduction to the Special Section on Research Practices. *Perspectives on Psychological Science, 7,* 655–656.

Spellman, B. A. (2012c). Scientific utopia… or too much information? Comment on Nosek and Bar-Anan. *Psychological Inquiry, 23,* 303–304.

Spellman, B. A. (2013). There is no such thing as replication but we should do it anyway: Comment on "Recommendations for increasing replicability in psychology." *European Journal of Personality, 27,* 120–144. doi:10.1002/per.1920

Spellman, B. A. (2015). A short (personal) future history of Revolution 2.0. *Perspectives on Psychological Science, 10,* 886–899. doi:10.1177/1745691615609918

Spellman, B. A. (2016). A different kind of scientific revolution. *The New Atlantis: A Journal of Science and Technology, 49,* 46–50.

Spies, J. R. (2013). The Open Science Framework: Improving science by making it open and accessible. Ph.D. Dissertation at the Department of Psychology, University of Virginia. https://search.proquest.com/docview/1419438360

Stirling, J. (2015, February 5). The dark side of post-publication peer review [Blog post]. Retrieved from http://physicsfocus.org/dark-side-post-publication-peer-review/

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science, 7*(6), 670–688. doi:10.1177/1745691612460687

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science, 9,* 59–71. doi:10.1177/1745691613514450

Swazey, J. P., Anderson, M. S., Lewis, K. S., & Louis, K. S. (1993). Ethical problems in academic research. *American Scientist*, *81*(6), 542–553. http://www.jstor.org/stable/29775057

Stodden, V. & Miguez, S., (2014). Best practices for computational science: Software infrastructure and environments for reproducible and extensible research. *Journal of Open Research Software, 2*(1), p. E21. doi:10.5334/jors.ay

Swoger, B. (2014, March 26). Post publication peer-review: Everything changes, and everything stays the same [Blog post]. Retrieved from https://blogs.scientificamerican.com/information-culture/post-publication-peer-review-everything-changes-and-everything-stays-the-same/

Tennant, J. P., Walder, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, C. H. J. (2016). The academic, economic, and societal impacts of open access: An evidence-based review [version 3]. *F1000 Research, 5,* 632. doi:10.12688/f1000research.8460.3

Teplitskiy, M., Lu, G., & Duede, E. (2016). Amplifying the impact of Open Access: Wikipedia and the diffusion of science [Online article]. *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.23687

Tilburg University (2011, September 7). Prof. Diederik Stapel suspended. https://uvtapp.uvt.nl/tsb11/npc.npc.ShowPressReleaseCM?v_id=4082238588785510

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185* (4157), 1124–1131. doi:10.1126/science.185.4157.1124.

van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology, 67,* 2–12.

Vanpaemel, W., Vermorgen, M., Deriemaecker, L., & Storms, G. (2015). Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra. 1*(1, Pt. 3). doi:10.1525/collabra.13

van Rooyen, S., Delamothe, T., & Evans, S. J. (2010). Effect on peer review of telling reviewers that their signed reviews might be posted on the web: Randomised controlled trial. *British Medical Journal, 341,* c5729. doi:10.1136/bmj.c5729

van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. (1999). Effect of open peer review on quality of reviews and on reviewers' recommendations: A randomised trial. *British Medical Journal, 318*(7175), 23–27. doi:10.1136/bmj.318.7175.23

Vazire, S. (2017, March 3). *Looking under the hood* [Blog post]. Retrieved from http://sometimesimwrong.typepad.com/wrong/2017/03/looking-under-the-hood.html.

Vines, T. H., Albert, A. Y., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., ... Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology, 24*(1), 94–97. doi:10.1016/j.cub.2013.11.014

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4,* 274–290.

Wade, N. (2010, August 20). Harvard finds Marc Hauser guilty of scientific misconduct. *New York Times*. Retrieved from http://www.nytimes.com/2010/08/21/education/21harvard.html

Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., ... Zwaan, R. A. (2016). Registered Replication Report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science*, *11*(6), 917–928. doi:10.1177/1745691616674458

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*(3), 426–432. doi:10.1037/a0022790

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. doi:10.1177/1745691612463078

Walsh, E., Rooney, M., Appleby, L., & Wilkinson, G. (2000). Open peer review: A randomised controlled trial. *British Journal of Psychiatry*, *176*(1), 47–51. doi:10.1192/bjp.176.1.47

Watkins, H. (2017, May 31). System justification in social psychology? A survey of responses to the replicability debate in psychology. Retrieved from https://psyarxiv.com/nn67y/. doi:10.17605/OSF.IO/NN67Y

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*(11), 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*(7), 726–729. doi:10.1037/0003-066X.61.7.726

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. doi:10.1038/sdata.2016.18

Yong, E. (2012, July 12). Uncertainty shrouds psychologist's resignation. *Nature News*. doi:10.1038/nature.2012.10968

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLOS ONE, 7*(12), e51382. doi:10.1371/journal.pone.0051382

Table 1: Timeline of Events Relevant to the Move to Open Science in Psychology

	Meetings and Organizations	Publications and Communications	Technology and Other
2005		Ioannidis, "Why Most Research Findings are False" in *PLoS Med*	
2007			Dataverse launches
2008		Neuroskeptic begins blogging	Dropbox launches Github launches
2009		Vul et al., "Puzzling high Correlations in fMRI Studies…" in *Perspectives on Psychological Science*	
2010		Launch of Retraction Watch http://retractionwatch.com Hauser investigation finds "scientific misconduct"(Wade, 2010)	Open Science Framework created (J. Spies, Dissertation, UVA)
2011	First Reproducibility Project by OSC begins.	Bem, "Feeling the Future" in *Journal of Personality and Social Psychology* Stapel fraud inquiry becomes public.	Psychfiledrawer.org launches

2012	Decline Effect Conference at UC Santa Barbara (Oct) (J. Schooler) BITSS Founded – First meeting of the Berkeley Initiative for Transparency in the Social Sciences	Simmons, Nelson, Simonsohn, "False-positive Psychology…" in *Psychological Science* (Jan) 2012 Levelt Report published (documenting Stapel fraud) Daniel Kahneman sees "a train wreck looming" (Oct) for social priming research *Perspectives on Psychological Science* full issue on Replicability and Research (eds., Pashler & Wagenmakers; Spellman) (Nov)	Figshare launches PubPeer starts
2013	COS founded (March) APS – Double session on Methods (May) President Obama signs Executive Order making open government data the new default (May) White House honors Champions of Change in Open Science (June) https://www.whitehouse.g	*Cortex* (ed. Chambers) first to launch Registered Reports format (March)	

	ov/blog/2013/06/21/celebrating-open-science-champions-change-white-house		
2014	==Pnomics ?== ==Jonathan Baron? (open data at JDM?)== TOP development meeting (Nov)	"*p-Curve:* The key to the file drawer" (Simmons, Nelson, & Simonsohn) published *Psychological Science* now requires authors to disclose excluded data, variables, and conditions Eich (ed.) announces badges for open materials, data, and pre-registration in *Psychological Science* (Jan.) *Social Psychology* special issue of Registered Reports (eds., Nosek & Lakens; 15 studies) First RRR Perspectives *JPSP* starts taking replications	
2015		Nosek et al., Promoting an Open Research Culture (aka "TOP Guidelines", *Science*) (June) Establishment of Psych Methods (June) and PsychMAP (December) Facebook groups	Curatescience.org launches

		CRSP (*Comprehensive Results in Social Psychology*), a pre-registration only journal founded	
		Open Science Collaboration, "Estimating the Reproducibility of Psychological Science" (100 replications)	
2016	SIPS – First meeting of the Society for the Improvement of Psychological Science (June)	*Psychological Science* pilots StatCheck for submitted papers	PsyArXiv, a dedicated preprint server for psychology, launches
2017		*Advances in Methodologies and Practices in Psychological Science* founded.	
		Chambers, *The Seven Deadly Sins of Psychology: A Manifesto for Reforming the Culture of Scientific Practice*	

Table 2: How Open Science Can Remedy Problems Arising from the Research Process

Stage of Research Process	Sources of Error & Bias	Open Science Solutions
Research planning	Quantity >> Quality Emphasis on smallest publishable unit Underpowered designs	Changing norms and reward structures Technological solutions
Data collection and methods reporting	Optional stopping Failure to disclose all treatments, covariates, and dependent measures	21-word solution Open materials Open workflow Pre-registration
Data analysis and results reporting	HARKing p-hacking Flexible data cleaning and analytic choices	Open data and code Pre-registration Publishing replications Improved statistical training and re-training Distinguishing exploratory from confirmatory results
Publication process	Sexy >> Solid Emphasis on $p < .05$ Overclaiming Tidy stories Lack of publication of replications	Open peer review Post-publication peer review Pre-registered reports Alternatives to traditional manuscripts (e.g., data descriptors) Preprints and new media Tolerance of "messy/imperfect" findings
Storing and archiving	File drawer problem Information loss creates bias in meta-analyses	Open materials, data, and workflow Rewards for sharing Innovations in curation and indexing

Note. Influenced by various presentations over last few years by Chris Chambers and Brian Nosek.   >> means "viewed as much more important than"

Table 3: Open Science Resource Box: Some Materials for Understanding and Practicing Open Science

Organizations		
	Society for the Improvement of Psychological Science (SIPS)	http://improvingpsych.org
	Berkeley Initiative for Transparency in the Social Sciences (BITSS)	http://www.bitss.org/
	Center for Open Science (COS)	https://cos.io
Teaching/Learning Resources		
	Course Syllabi	
	COS list of syllabi	https://osf.io/vkhbt/
	Project TIER materials	http://www.projecttier.org/tier-classroom/course-materials/
	Online Tutorials/Guidelines	
	Video tutorials on the Open Science Framework and many other open science topics	https://www.youtube.com/channel/UCGPlVf8FsQ23BehDLFrQa-g
	Tutorial on Open, Reproducible Workflows	https://www.youtube.com/watch?v=kzUtpDBo8wk
	BITTS Manual of Best Practices	Christenson, 2016, Manual at https://osf.io/kmjc2/
	Pre-registration	https://www.youtube.com/watch?v=EnKkGO3OM9c
Planning / Completing a Replication		
	Journals Accepting Registered Reports	https://cos.io/rr/?_ga=1.62517349.857803987.1480710442#journals

	Calls for Large-scale Registered Registrations (APS)	http://www.psychologicalscience.org/publications/replication/ongoing-projects
	Collaborative Replications and Education Project (CREP)	https://osf.io/wfc6u/
Journal Standards and Aspirations		
	Transparency and Openness Promotion (TOP)	
	TOP Guidelines	https://cos.io/top/
	Journal Signatories to TOP Guidelines	https://cos.io/our-services/top-guidelines/
	Journals Level of Implementation of TOP Guidelines	TBD
	Catalogue of Journal-Specific Pre-Print & Postprint Rules	
	Sherpa/Romeo List of Publisher Copyright Policies and Self-Archiving	http://www.sherpa.ac.uk/romeo/index.php
Open Sharing and Reviewing Resources: Online Repositories (Note: There are several well-used proprietary services we have not listed.)		
	For Data	
	Nature's Recommendation List	http://www.nature.com/sdata/policies/repositories
	Open Access Directory's List	http://oad.simmons.edu/oadwiki/Data_repositories
	Open Science Framework	https://osf.io
	For Preprints/Postprints	
	Academia (general)	https://academia.edu

	PsyArXiv	https://osf.io/preprints/psyarxiv
	OSF Preprints (general)	https://osf.io/preprints/
	For Pre-registrations	
	As Predicted	https://aspredicted.org
	Open Science Framework	https://osf.io https://osf.io/registries/
	For Post-publication Review	
	PubPeer (anonymous option)	https://pubpeer.com/
	Directory of Open Access Journals	https://doaj.org
Checklists and Templates		
	For Pre-registration	
	Pre-registration checklist from OSF	https://osf.io/ncqg7/
	Pre-registration sample template	Van't Veer & Giner-Sorolla, 2016 (https://osf.io/t6m9v/)
	Lab communication and re-registration powerpoint	Corker, 2016 (https://osf.io/c36dq/)
	For Lab Workflow	
	Campbell Lab Disclosure Template	Campbell, 2015 (https://osf.io/m7f8d/)
	Ledgerwood Lab Experiment Archive Form	Ledgerwood, 2016 (https://ucdavis.app.box.com/s/f8hn7rqtwwf6aa6hjtkthdbiuehup312)
	Checklists for Research	Nosek et al., 2014 (https://osf.io/mv8pj/)

	Workflow	
	Corker Lab Template	https://osf.io/sjtyr/
	Project Tier template	http://www.projecttier.org/tier-protocol/specifications/ https://osf.io/ybzxe/
Blogs/Online Groups/Feeds		
	Blogs/Feeds	
	PsychBrief Psychological Methods Blog List and Sign-Up Option	http://psychbrief.com/psychological-methods-blog-feed/
	DataColada (Simmons, Nelson, Simonsohn)	http://datacolada.org
	The Hardest Science (Sanjay Srivastava)	https://hardsci.wordpress.com/
	Sometimes I'm Wrong (Simine Vazire)	sometimesimwrong.typepad.com/
	Invariances (Jeff Rouder)	http://jeffrouder.blogspot.com/
	Statistical Modeling, Causal Inference, and Social Science (Andrew Gelman)	http://andrewgelman.com/
	Facebook Groups	
	PsychMAP (Psychological Methods and Practices)	https://www.facebook.com/groups/psychmap
	PsychMethods (Psychological Methods Discussion Group)	https://www.facebook.com/groups/853552931365745/
	List of Active Psych Groups on Facebook (many but not all re: Open Science)	https://docs.google.com/document/d/1wXU0bc23ulRN0cBsFcNTJm5t-QqMOjGe5lsRWzfwGPo/edit

Awards/Incentives		
	Leamer-Rosenthal Prizes for Open Science	http://www.bitss.org/lr-prizes/
	The Preregistration Challenge from COS	https://cos.io/prereg/
	Badges for Open Science	https://osf.io/tvyxz/wiki/home/

Table 4. Summary of the eight standards and three levels of the TOP guidelines (as published in *Science*).

	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Citation standards	Journal encourages citation of data, code, and materials—or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
Data transparency	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Analytic methods (code) transparency	Journal encourages code sharing—or says nothing.	Article states whether code is available and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Research materials transparency	Journal encourages materials sharing—or says nothing	Article states whether materials are available and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Design and analysis transparency	Journal encourages design and analysis transparency or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.

Preregistration of studies	Journal says nothing.	Journal encourages preregistration of studies and provides link in article to preregistration if it exists.	Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Preregistration of analysis plans	Journal says nothing.	Journal encourages preanalysis plans and provides link in article to registered analysis plan if it exists.	Journal encourages preanalysis plans and provides link in article and certification of meeting registered analysis plan badge requirements.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies—or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts blind review of results.	Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes.

Note: Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.