Original Article

Sociological Theory
2018, Vol. 36(3) 289–313

DOI: 10.1177/0735275118794987
st.sagepub.com

The Emergence of Statistical Objectivity: Changing Ideas of Epistemic Vice and Virtue in Science

Jeremy Freese[1] and David Peterson[2]

Abstract

The meaning of objectivity in any specific setting reflects historically situated understandings of both science and self. Recently, various scientific fields have confronted growing mistrust about the replicability of findings, and statistical techniques have been deployed to articulate a "crisis of false positives." In response, epistemic activists have invoked a decidedly economic understanding of scientists' selves. This has prompted a scientific social movement of proposed reforms, including regulating disclosure of "backstage" research details and enhancing incentives for replication. We theorize that together, these events represent the emergence of a new formulation of objectivity. Statistical objectivity assesses the integrity of research literatures in the results observed in collections of studies rather than the methodological details of individual studies and thus positions meta-analysis as the ultimate arbiter of scientific objectivity. Statistical objectivity presents a challenge to scientific communities and raises new questions for sociological theory about tensions between quantification and expertise.

Keywords

objectivity, meta-analysis, transparency, replication, false positives

INTRODUCTION

"Objectivity" is a core aspiration of conventional science. Yet the goals of producing objective knowledge often come into conflict with the expertise needed to produce it (Daston 1992; Daston and Galison 1992, 2007; Porter 1995). Scientists are lauded for uncommon skill and judgment, but these may come to be regarded as barriers to the universality and transparency implicit in objectivity. During periods of scandal or controversy, scientists' judgment may come to be seen as a potential source of bias and even corruption.

[1]Stanford University, Palo Alto, CA, USA
[2]University of California, Los Angeles, Los Angeles, CA, USA

Corresponding Author:
David Peterson, University of California Los angeles, 621 Charles E. Young Dr., South Box 957221, 3360 LSB Los Angeles, CA 90095-7221, USA.
Email: davidapeterson@g.ucla.edu

Presently, various scientific fields are said to be threatened by a "crisis of credibility" that centers on concerns about the replicability of published research. National Institutes of Health director Francis Collins described replicability concerns as a "cloud" over biomedical research (Hughes 2014). This includes findings of poor rates of successful replication in drug target (Prinz, Schlange, and Asadullah 2011) and cancer (Begley and Ellis 2012) research and influential papers raising alarms about neuroscience (Button et al. 2013) and medical genetics (Greene et al. 2009). With 30 other journals, *Science* and *Nature* have published an unprecedented joint editorial making specific commitments to replicable science (Center for Open Science [COS] 2015).

Psychology in particular has received much attention regarding its replicability. This includes a widely publicized effort to replicate 100 sampled findings from leading psychology journals, which reported only 39 percent success in terms of statistical significance and only 59 percent success in finding even "moderately similar" results (Baker 2015). Psychology's chronically insecure status as a science has historically led psychologists to aggressively pursue new technologies of objectivity that are later adopted by other sciences (Danziger 1990; Porter 1995). Again, psychologists are at the vanguard, advocating for significant changes in scientific practice. Of the field's recent contributions to cross-disciplinary debates about replication, Reinhart (2016:413) notes, "social psychology is at the center." Significantly, a social psychologist co-founded the Center for Open Science, an organization dedicated to promoting changes in research practice across the sciences. Since its formation in 2013, it has received over $25 million in private and public funding.

According to Daston and Galison (2007), debates about validity in science revolve around specific, historically situated articulations both of *epistemic vices* and the *epistemic virtues* that scientists are pressed to adopt to overcome these vices. We use recent events in social psychology to develop the theoretical argument that fears about replicability across the sciences reflect the emergence of a new and powerful means of articulating epistemic vice. In response, epistemic activists within science have promoted a correspondingly novel formulation of objectivity, which we call *statistical objectivity*.[1] The central feature of statistical objectivity is the projection of debates about objectivity and subjectivity onto the patterns of results produced by *collections* of studies rather than the methodological details of *individual* studies. This not only undermines traditional interpretations of scientific evidence but reveals, in ways that are invisible when studies are evaluated in isolation, how currently acceptable forms of expert discretion can lead to systematic problems in literatures.

By reframing objectivity as a cumulating achievement, activists have simultaneously redefined epistemic vice. Rather than the incursion of individual subjectivity into objective research, they target the collective failure that results from the misalignment of institutional incentives. In what follows, we outline how this understanding has inspired a package of institutional reforms, which present fundamental challenges to both disclosure practices and data interpretation. We argue that recent changes to scientific practice represent the restatement of classical debates regarding objectivity onto a new, collective plane.

Statistical objectivity is a scientific social movement (Frickel and Gross 2005) that demands sociological attention for two related reasons. First, it highlights significant changes in the nature of expertise. It focuses its floodlights on the Goffmanian "backstage" of science, the private domain in which individual scientists typically have been granted the authority to package and present their work. Demands for increased transparency transform scientific experts from the producers of finished science to data farmers, producing grist for a meta-analytic mill. Second, statistical objectivity represents a new frontier of quantification. Although science is already dominated by statistical methods, the urge toward replication creates a second-order quantification that makes the meta-analyst the ultimate arbiter of scientific disputes.

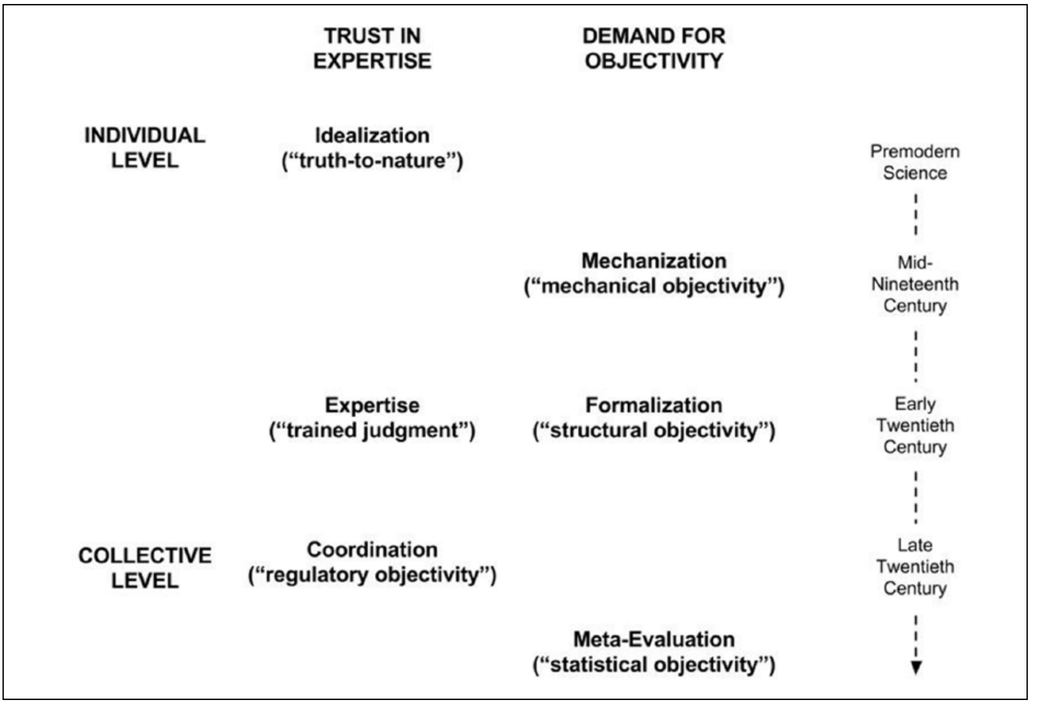

Figure 1. Movements in expertise and objectivity.

EXPERTISE AND OBJECTIVITY

"All epistemology begins in fear," write Daston and Galison (2007:372). During periods of high anxiety and suspicion—periods in which *what we know* is no longer secure—issues of *how we know* come to the fore. Historically, scientific epistemology has been motivated by the fear of subjectivity because the achievement of objective knowledge has been understood to be possible only through "the suppression of some aspect of the self, the countering of subjectivity" (Daston and Galison 2007:36). Through a historical study of scientific atlases, they outline the "epistemic virtues" and "vices" that dominated different historical periods. Although they differ along many particulars, each form of knowing may be understood as another movement in an interplay between the valorization of expertise, which is the personified unification of objectivity and subjectivity and the drive to erect strong barriers between objectivity and subjectivity that occurs when experts lose credibility.

Changes in social and technical conditions continually challenge prevailing practices, creating novel vices that in turn motivate new epistemic virtues. This interplay pulls fields from trusting experts to demanding objectivity and back again (Figure 1).

Daston and Galison (2007) argue that before the historical advent of the modern concept of "objectivity," science was guided by the Platonic belief that nature provided only imperfect examples of pure, objective forms ("truth-to-nature"). For example, naturalists were responsible for synthesizing their observations into "ideal" or "characteristic" portrayals. Yet early atlases led to anxieties about the potential for researchers to subjectively aestheticize or theorize images.

In the mid–nineteenth century, concerns over subjective bias motivated a turn toward "mechanical objectivity." Researchers restrained from idealizing depictions of nature through the use of machines and a strict adherence to protocols. In scientific atlases, advances in photography were purported to produce representations free from human input. Subsequent developments reaffirmed the tension between the need for expertise to interpret complex data ("trained judgment") and the need to overcome individual bias through the development of a universal language of science ("structural objectivity").

However, recent increases in the scale and interconnectedness of science has resulted in challenges to objectivity that earlier scientists could not have envisioned. As the scientific community grew in the postwar era, so did concerns that differences in research objects, protocols, and data analysis procedures produced incongruous literatures that threatened to fragment research and make aggregation impossible. To counter this problem, some fields have developed top-down systems of coordination (Collins 2004). Researchers—especially in medical fields—have increasingly turned to guidelines, rules, standards, and regulations to enforce integration (Berg et al. 2000; Timmermans and Epstein 2010).

Cambrosio and colleagues (2006, 2009) have labeled this move toward centralized coordination *regulatory objectivity*. Like previous regimes, regulatory objectivity depends on trust in expertise. However, in place of the trained judgment used to make sense of individual findings, groups of experts develop tools designed to orchestrate the "collective production of evidence" (Cambrosio et al. 2009:654). With its concern for the coordination of entire fields, regulatory objectivity represents a break from the concern with the individualist epistemic virtues that dominated earlier periods.

Statistical Objectivity

We theorize here that these major developments in the history of objectivity have recently been joined by another. The signature of *statistical objectivity* is the grounding of objectivity in an aggregate assessment of the coherence of results reported by multiple studies. It is a meta objectivity, embracing a statistical logic by which findings, once presented as self-sufficient, are recast as data points in a higher order analysis.

Statistical objectivity is a response to the fear that various interests involved in producing and publishing results can be so profound and pervasive that they enable a self-reinforcing pair of problems. One is a vast proliferation of exaggerated knowledge claims; the other is a weakened capacity for exaggerated claims to be subsequently corrected.

As with mechanical objectivity, the reforms we associate with statistical objectivity are rooted in a concern with how researchers' subjectivities may prompt erroneous idealizations of data. However, unlike mechanical objectivity, the primary object of scrutiny in statistical objectivity is not the individual study but a *population of studies*. This move toward conceptualizing objectivity as a quality of collected studies is anticipated by the emphases on standardization and regulation of research practice that mark regulatory objectivity. But while regulatory objectivity centralizes expertise and integrates research communities by implementing rules regarding the production of data, statistical objectivity emphasizes transformation in how research is reported and interpreted.

Many recent studies have investigated aspects of statistical objectivity, including recent literature on "meta" science (Edwards et al. 2011; Zimmerman 2008), the expansion of forensic science (Kruse 2012; Lynch et al. 2008), evidence-based medicine (Timmermans and Berg 2003), and the explicit codification of rules for conducting and reporting research (Frow 2012; Leahey 2008). Moreover, statistical objectivity is a pointed critique of the emphasis on "framing" scientific findings to maximize their theoretical potency (Healy 2017; Strang and Siler 2015), suggesting that such work obscures vital detail and creates a culture of outlandish claims disconnected from questions of truth or falsity (Wakeham 2017). Yet, as we document here, what makes statistical objectivity a unique and significant development is that it combines these different ideas into a potent package that includes a philosophy of science, set of statistical tools, and list of demands regarding changes in scientific practice.

We develop our theory about the emergence of statistical objectivity by offering sustained attention to the case of recent developments in social psychology. Our argument has two parts. In the first, we detail how scientists have reframed prevailing ideas about the scientific self. We describe how epistemic activists have raised the possibility of a "crisis of false positives" by analyzing collections of studies, making visible a threat to objectivity that is hidden when studies are considered on their own. Then, we explain how this threat to objectivity has been dominantly depicted by epistemic activists in terms of a particular view of scientists' selves: namely, scientists as economic actors led to bad practices by a poorly aligned system of incentives.

In the second part, we discuss how this understanding shapes the two complementary and mutually reinforcing reforms that epistemic activists have pressed. One is constraining the ability of scientists to control the interpretation of their findings by keeping their methods and data sequestered on an inaccessible "backstage" of science. Instead, activists have sought to increase and standardize the disclosure of details of the research process that were previously unreported. The other is the cultivation of collections of studies that allow techniques of collective evaluation to be more powerfully applied. This represents a quantification of quantification in which scientific evidence that once stood on its own is transformed into mere data points in higher order, meta-analytic data sets.

THE CHALLENGE OF COLLECTIVE ASSESSMENT TO THE SCIENTIFIC SELF

The capacity for inferential statistics to *articulate unlikeliness* can give it enormous rhetorical force. In fingerprint analysis, for example, fingertips are represented as a series of standardized, categorizable "points of identification" (Cole 1998). Many people share particular points, but as the number of points considered increases, statistics allows investigators to make claims regarding the unlikeliness that anyone other than a suspect could have produced a particular fingerprint. Likewise, in accounting investigations, deviations from expected distributions in the frequencies of digits often serve as initial evidence of an irregularity in how numbers were generated: revealing, in some cases, fraud (Durtschi, Hillison, and Pacini 2004).

Similar statistical demonstrations have initiated the detection of outright fraud in science, including psychology. Three social psychologists have resigned from positions as a result of charges of fabrication that were instigated by methodologists who showed that patterns in published results were too consistent across studies given expected natural fluctuations of real data. (Borsboom, van der Mass, and Wagenmakers 2014; Simonsohn 2013; van der Heijden, Groenen, and Zeelenberg 2014). For example, in one case, investigators estimated that the probability of real data achieving a claimed level of linearity in a suspect series of studies would be 1 in 508 quintillion (Borsboom et al. 2014).

Importantly, this logic is not limited to detecting fraud. For the developments we describe in this paper, the use of forensic-style statistics to reveal fraud is emphatically of secondary concern. In psychology, biomedicine, and elsewhere, statistical tools are used for evaluating the plausibility of literatures being infested with "false positives"—findings that are greatly overstated, if not simply wrong.

One tool in such work is a funnel plot, shown in Figure 2. Each point represents the results of one study. If a set of experiments all estimate the same effect, effect sizes should be symmetrically distributed around the average effect size (the dashed lines in

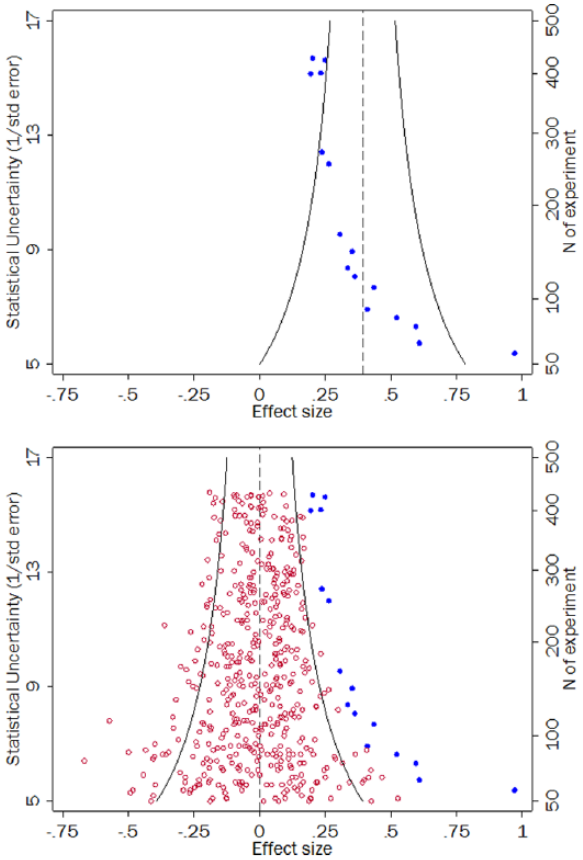

Figure 2. Funnel plots. Each dot represents a simulated study estimating a true effect size of zero (see supplemental material for simulation design). The top panel is a collection of studies that report positive, statistically significant findings. Bias in the collection is evident from the negative association between the observed effect size (*x*-axis) and its statistical uncertainty (*y*-axis). The bottom panel includes the effect sizes from all the simulated studies that were not statistically significant in the predicted direction (hollow circles). Note both that (1) only in the bottom panel do results correspond to the expected funnel shape and (2) the average effect size in the biased collection (dashed line) diverges sharply from the average of zero in the unbiased collection.

Figure 2). However, estimates should narrow as the statistical uncertainty of results decrease (e.g., studies with a larger sample size), producing a "funnel" shape (as in the bottom plot of Figure 2).

On the other hand, if the set of studies available in the published record is biased because only statistically significant findings are being published, larger studies will have systematically smaller effect sizes. This leads to a greater concentration of studies in the bottom right and upper left quadrants of the funnel (the top plot in Figure 2). Consequently, even though the top plot would appear to depict a set of studies with consistent, positive results in favor of a hypothesis, the funnel plot may be taken to demonstrate that the literature in question is biased. In fact, as the bottom plot in Figure 2 shows, the pattern of published findings shown in the top plot could be observed even if no true effect exists.

Table 1. Problems with *p*.

The standard interpretation of a *p* value in an experiment is as the probability of observing an equal or greater difference between treatment and control groups if the manipulation had no actual effect. Without adjustment, this interpretation assumes a given hypothesis test is the only test. The following practices all make *p* fictive in ways that make it easier to obtaining a publishable *p* value when no true effect exists.	
File drawer problem	A researcher conducts many experiments of many hypotheses but selectively reports experiments based on whether p is significant.
Dropping studies	A researcher conducts multiple experiments that test a hypothesis but selects which to report based on whether *p* is significant.
Data peeking	A researcher computes *p* as data are being collected, deciding to stop collecting data if results are significant but continuing otherwise.
p-hacking	A researcher tests the hypothesis by analyzing the data in various ways and determining which analyses to present based on whether the results are significant.
HARKing	A researcher conducts exploratory analyses, devises a post hoc explanation for an analysis for which a significant result is found, and then interprets the result as if it were an a priori prediction.

Although the funnel plot was developed in the 1980s, the growing volume of experimental studies and increasing ease of accessing them through online databases has made it a more potent tool. Alongside, a growing variety of related techniques have been developed. Particularly influential in psychology has been the *p*-curve method (Simonsohn, Nelson, and Simmons 2014), which is based on one way that the results of a set of experiments detecting a true effect would differ from a set of experiments in which statistically significant results were the illusory result of the biased analytic decisions known as "*p*-hacking" (described in Table 1).

Two features typical of these demonstrations are worth highlighting. First, they juxtapose the results of actual literatures and statistical expectations in clear, visual terms. The rhetorical strength of public demonstrations has been posited as one of the pillars of modern science (Rosental 2013; Shapin and Shaffer 1985), and data visualizations remain a central tool of scientific persuasion (Burri and Dumit 2008). What makes these demonstrations especially potent is that they are based on purely formal statistics and purport to present a clear view of what a literature *should* look like.

Second, the demonstrations analyze and produce conclusions about *collections* of results. These tools are most powerful for demonstrating general "fishiness." That is, they reveal questionable collective properties that can raise doubt and prompt investigations. Yet, even when analytics point to extreme fishiness about a set of results, any particular result in the collection could have been arrived at through unimpeachable practice.

These forensic-style demonstrations thereby produce a new plane on which the integrity of claims may be scrutinized. When the credibility of an individual claim is considered on its own, judgments of its objectivity focus on *how* it was produced. Once aggregated into a collection of claims, however, they may be expected to exhibit particular statistical properties if they are to appear collectively credible. When individually credible numbers do not have credible collective properties, doubt can pervade a literature. Individual credibility is threatened even if one cannot identify any specific problem in how any specific number was produced. That is, the demonstrations make possible that a collection of studies, which previously appeared impressively consistent in their findings and impeccable in methods, might be instead shown to be consistent with a "crisis of false positives," in which the true effect is either radically smaller than what had been reported or even potentially nonexistent.

One might view the challenge of collective assessment as simply adding to the "trials of strength" (Latour 1987) that scientists must withstand to be published. But this misses the more fundamental disruption that these demonstrations pose: the introduction of a new and separate assessment of claims based on collectively analyzing *sets* of published studies. This raises scientific criticism from discussions of individual findings to scientific cultures. The results of these collective assessments can provoke new skepticism and doubt about individual studies and their authors. But more significantly, such failures undermine the claims that scientists are ascetic actors unaffected by worldly incentives or that the system of self-policing has regulated shortsighted impulses. As such, it demands a systematic response.

ECONOMIC REASONING AND SCIENTIFIC SELVES

The capacity for statistical demonstrations to cast compelling doubts about collections of studies does not in itself determine how the problem is understood or what potential solutions for it are posed. As Daston and Galison (2007:36–37) argue, new objectivities are typically posited as solutions for "a certain kind of willful self, one perceived as endangering scientific knowledge." These serve as "collective action frames," which help both identify problems and point toward remedies (Benford and Snow 2000). In the present case, subjective influence could be viewed as a moral failure of individuals to resist temptations or a socialization failure by epistemic cultures. Some arguments have been made to each effect. Yet in both psychology and science more broadly, what is striking about discussions of the causes and potential solutions of the "replicability crisis" is how thoroughly dominated they are by an economic view of the self.

By *economic*, we mean a view of self that emphasizes responsiveness to incentives provided by institutions rather than one driven by morals or socialization to scientific norms. (We do not mean *economic* to imply that these incentives are focused on money per se.) Epistemic activists locate the root cause of biased literatures as a "dysfunctional reward structure" for scientific selves (Miguel et al. 2014) that prizes framing over exacting empirics and methods (Strang and Siler 2015). One incisive description of the incentive problem frames it as a conflict between "getting it published" and "getting it right" (Nosek, Spies, and Motyl 2012).

This disconnect is highlighted by the title of perhaps the most influential paper prompting concerns about false positives: "Why Most Published Findings Are False," by Ioannidis (2005).[2] Mixing statistical theory and rational-choice style reasoning, the paper presents its provocative title as a logical inevitability given prevailing incentives and standards. Specifically, Ioannidis argues that the proportion of false positive findings in a literature depends on (1) the likeliness of hypotheses and (2) the strength of the evidence required to publish, while the extent to which false positive findings remain unrefuted in a literature depends on (3) the strength of mechanisms of self-correction. In recent challenges to social psychology, failures in all three aspects have been asserted.

Unlikely Hypotheses

Novelty and discovery are, of course, vital to scientific progress. Social psychology, however, has often been criticized for overvaluing highly counterintuitive findings. Counterintuitive hypotheses are understood as having particular popular appeal, and popular interest is rewarded in the field in many ways (Love 2013). Epistemic activists argue that this "Gladwellization" of the field promotes both the pursuit of counterintuitive findings by authors and a preference for them among editors (Nosek et al. 2012; Posner 2014). They have contended this not only directs attention toward hypothesis that are likely false but, alluding to Kuhn (1962), also reduces incentives to produce the "normal science" that is more incremental but has a greater chance of enduring (OSC 2015).

Although epistemic activists have criticized many unlikely hypotheses in social psychology, one study in particular has been significant in galvanizing opposition: a 2011 publication in social psychology's leading journal, the *Journal of Personality and Social Psychology* (*JPSP*), in which eminent Cornell psychologist Daryl Bem (2011) presented experimental evidence of precognition. His paper supposedly demonstrated that subject behavior was influenced by a randomly assigned *future* event. If the work had been conducted by an unfamiliar investigator, perhaps fabrication may have been suspected, but Bem was a high-profile psychologist with a long history of contributions to psychological science. For those unwilling to entertain paranormal claims, this took the pursuit of unlikely hypotheses to its logical extreme: a hypothesis with zero chance of being true. That a false hypothesis could be published with experimental evidence that, if anything, exceeded prevailing evidential standards in the field provided an obvious prompt for reflection.

Weak Evidentiary Safeguards

Epistemic activists have strongly targeted a standard that has long served as the primary gatekeeper in many behavioral and biomedical fields: null hypothesis significance testing (NHST), and particularly the reliance on a threshold of $p < .05$. While the idea that NHST provided an "objective" method to evaluate results was key to its rise in psychology and social science (Danziger 1990; Porter 1995), it has since come under growing criticism. Some believe problems would be greatly diminished by requiring a more stringent threshold, like $p < .005$ (Benjamin et al. 2017), and others contend that the whole statistical framework on which p values are based invites trouble (Woolston 2015).

In an article that has been credited with crystalizing anxieties regarding false positives in psychology, Simmons, Nelson, and Simonsohn (2011) provided a dramatic demonstration of the potential mischief produced by "undisclosed flexibility" in analyses as they ran an experiment for a nonsense hypothesis and nevertheless were able to manipulate analyses to yield statistically significant results using common practices of the field. Anonymized surveys of psychologists suggest that these activities have been widespread (John, Loewenstein, and Prelec 2012). Although such practices have long existed in the gray areas of professional discretion, activists have encouraged their reinterpretation as "soft fraud" (Chambers 2014). The principal axes of undisclosed flexibility are summarized on Table 1.

Weak Self-correction

In his classic discussion of science as a self-correcting enterprise, Merton (1973:276) describes scientists as "subject to rigorous policing, to a degree perhaps unparalleled in any other field of activity." However, Stapel's (2014:116–17) memoir of his fraud presents a different picture:

> It was very, very easy. . . . Nobody ever checked my work; everyone trusted me. . . . I did it all myself, with a big cookie jar right next to me . . . and nobody watching . . . and next to me was a big jar of cookies . . . with nobody even near. I could take whatever I wanted.

For Merton, the fear of losing prestige was vital to maintaining scientists' discipline and "rigorous policing" connected prestige to quality of work. In contrast, a persistent complaint in social psychology is that the types of replications that might identify false positive studies are infrequently undertaken and even more rarely published (Makel, Plucker, and Hegarty 2012). The journal that published Bem's precognition study refused to review a failed replication of Bem's findings, citing a (since changed) policy against publishing replication studies (Aldhous 2011).

For several reasons, incentives for conducting replication studies in social psychology may be especially low. First, compared to many "bench" sciences, the ability to extend social psychological findings with a new experiment is less contingent on being able to replicate the prior experiment (Peterson 2015). Second, repeating experiments provide very little opportunity for displays of technical virtuosity. Thus, repeating experiments is easily derided as time-wasting and diagnostic of a lack of ideas of one's own. Third, in contrast to many "applied" biomedical sciences, the low external stakes regarding whether a given claim is true or false increases the interpretability of attempts at replication as personal attacks—as "bullying" (Schnall 2014) or, even, "methodological terrorism" (Singal 2016). Consequently, at least until recent developments, even some results regarded as "classics" had no published record of anyone simply trying to repeat the original experiment as closely as possible (Klein et al. 2014).

Runaway Expectations

Ioannidis (2005:700) raises the gloomy prospect that some areas of science could be "null fields," in which all the positive findings comprising the literature are simply reflections of the potential bias in their incentives and standards. If published findings shape what scientists subsequently regard as plausible, then false positive findings can inspire and beget other false positive findings. This may be especially true whenever there is weak gatekeeping and few consequences to publishing studies that cannot be replicated.

In these ways, problems originating in bad incentives can have a runaway character, in which problematic practices raise expectations and the push to meet those expectations begets even more problematic practices. The use of questionable research practices by psychologists to make an individual study's finding more compelling has been likened to the use of performance-enhancing drugs in sports (John et al. 2012). In both cases, the level of competition is artificially raised, putting fair competitors at a disadvantage. And like performance-enhancing drugs, these practices can produce outcomes that appear increasingly dubious to outsiders. Some vocal recent critics of social psychology, like the statistician and political scientist Gelman (Gelman and Carlin 2013; Gelman and Loken 2014), for example, have focused attention on large published effects that outside audiences might find implausible on their face, like a study finding that women's ovulatory cycles have large effects on their approval of Barack Obama (Durante, Rae, and Griskevicius 2013).

Conclusion: False Positives as a Social Dilemma

As analyses of sets of findings reveal hidden problems in literatures, researchers in those fields may be under considerable pressure to address them. But what to do? An economic understanding of the problem yields a straightforward social dilemma. Individual researchers have an incentive to produce studies that are compelling as possible. Yet, the cumulative consequence is a literature that cannot withstand statistical analysis. This in turn raises doubts about the whole field, regardless of whether specific methodological flaws are apparent. As with other social dilemmas, when the pursuit of individual interest is insufficiently constrained, the long-term welfare of the group as a whole suffers (Everett and Earp 2015). Moreover, moral exhortations may be regarded as insufficient to produce change in the absence of a realignment of institutional incentives. Consequently, solutions involving deep structural changes are favored—perhaps even regarded as necessary for meaningful progress—and in the second half of the article, we outline reforms advocated by epistemic activists that seek to change the incentives provided by science institutions.

STATISTICAL OBJECTIVITY I: PROJECTIVE DISCLOSURE

Goffman (1959:112) famously argued that social performances require hidden, "backstages" where "illusions and impressions are openly constructed." This behind-the-scenes work is necessary for performances yet needs to be obscured because it would undermine the desired impression. Laboratory ethnographies have long made the point that the actual practice that takes place on the "backstage" of scientific labs is messier and more interpretive than what is presented to audiences in journal articles (Gilbert and Mulkay 1984; Holton 1978; Knorr Cetina 1983, 1995; Woolgar 1982). The discrepancy between backstage practice and front-stage presentation might be regarded as ultimately benign, even if sociologically interesting, under the premise that the science ultimately "works." When forensic-style demonstrations cast doubt on whether the science actually does work, however, they raise the prospect that the tidying process is not merely sparing readers unnecessary details but instead is obscuring systematic subjective bias.

Unlike previous regimes of objectivity, which have sought to constrain the discretion of researchers, statistical objectivity makes no direct effort to constrain the role of experts' subjective judgments in producing findings. Instead, its epistemic activists pursue a policy of "frontstaging" in which practices and decisions that were previously allowed to remain in the backstage of scientific practice are made public. As with other surveillance technologies, such as police body cameras (Lyon 2001), practices designed to increase transparency produce possible material for some unknown future investigation. As advocates are quick to point out, the mere existence of a reviewable record can be enough to change behavior.

Classical liberal arguments present "transparency" as a central requirement for both rationalization and democracy (Krippner 2007; Meyer and Bromley 2014). Advocates for statistical objectivity have embraced this logic. Thus, a watchword of statistical objectivity is *open*: open data, open materials, open practice, open science. Advocates argue that openness addresses threats to objectivity in three ways. First openness increases the verifiability of findings because it reduces the extent to which readers need to take an author's claims on faith. In periods when there is trust in experts, a lack of disclosure may not be perceived as a problem. Once rising cynicism and doubt become seen as threats to the broader credibility of fields, however, explicit verifiability becomes available as an obvious mechanism for enhancing credibility and earning trust. Second, openness improves the quality of individual papers by discouraging questionable or incautious practices by introducing the risk of revelation and its reputational cost. Third, openness enhances what can be detected and learned from analyses of collections of studies. Making more details of studies available leads to opportunities for more powerful meta-analyses.

Statistical objectivity provokes two movements toward open practice: standardization about what research details are expected to be explicitly reported within a journal article and increasing expectations about the extensiveness of supplementing materials that are made publicly available as part of publication but are not part of the article itself.

Standardized Reporting

Transparent reporting practice promotes explicit expectations about which details of data collection and analysis will be reported. Elaborate guidelines have emerged in recent years in biomedical domains, most notably the CONSORT guidelines for reporting results of randomized clinical trials (Montgomery et al. 2013; Schulz et al. 2010; Simera et al. 2010). As part of the changes that the high-profile journal *Psychological Science* has recently implemented, researchers who submit manuscripts are now required to complete a checklist. It requires authors to affirm that for each experiment, their paper explicitly and accurately reports how the sample size of the experiment was determined, how many observations were excluded from analysis and why, all independent variables or manipulations ("whether successful or failed"), and all outcomes that were analyzed (Eich 2014).

Two features of using checklists like this bear emphasis. First, the checklist transforms backstage decisions into explicit moments of potential misconduct by mandating an occasion for truth-telling or lying where before there was the possibility of strategically ambiguous silence. The checklist thus "draws lines" about permissible conduct (Frow 2012). Second, the checklist does not directly regulate what researchers do. They are permitted full use of their judgment in designing experiments and analyzing data. Disclosure may require articulating details that might make the paper less credible or compelling for readers, and anticipation of such reactions of readers might influence how data are collected and analyzed. But unlike the strategies associated with regulatory objectivity (Cambrosio et al. 2006), any implications for practice are only indirect.

Another suggestion has been to introduce explicit mechanisms that allow researchers to certify transparent research practices. A growing number of psychological journals have agreed to publish badges for articles that meet particular guidelines (American Psychological Association 2017; Eich 2014). These are displayed as colorful icons appearing just below the title of articles as well as in listing of articles on their website and serve as mechanisms for signaling open practices. For instance, in addition to badges for "Open Data" and "Open Materials," a "Preregistration" badge is available for articles that include an experiment for which the experimental design and details of planned analysis were deposited in an independent archive prior to the data being collected (COS 2013). Because it constrains experimental and analytic choices, preregistration is intended to eliminate the possibility that significant results are due to the forms of *p*-hacking reported in Table 1.

Supplementing Materials

Information technologies have radically altered what might be asked of researchers to disclose about the backstage of their work. This is especially apparent in the rise of online "supplemental materials" to journal articles. Articles have traditionally offered additional results or materials as "available upon request," and many professional ethical codes state explicitly what authors are expected to provide on request by others. However, studies report abysmal success rates for such requests in practice (LeBel et al. 2013). Epistemic activists have pressed to replace vague ethical conventions about how researchers should respond to requests after publication with explicit incentives to post all relevant information publicly online at the time of publication.

Posting supplemental materials is already commonplace for journals like *Science*. In some cases, the supplemental materials may be far longer than the actual article (e.g., Rietveld et al. 2013 is a 3-page article with a 172-page supplement). *Psychological Science*, along with some other psychology journals, now allows articles to display an "Open Materials" badge if they publicly share sufficient material about their experimental procedures to permit other researchers to attempt to replicate the article's findings by collecting new data for new subjects.

The push for additional materials has also included calls for public disclosure of the quantitative raw data on which findings are based. In Simonsohn's (2013) paper describing his statistical analyses that resulted in the resignation of two psychologists for alleged data fabrication, he notes that he had also found similar irregularities in work of a third, unnamed psychologist, but pursuit had been stymied by the author simply claiming to have lost the original data. Information technologies have greatly simplified the process of making data publicly available, enabling the possibility of requiring data availability as a condition of publication.

At present, journals that have changed policy in response to pressure for greater data "openness" evince the same three levels of reform articulated earlier, allowing us to review these as summary here. First is *requiring explicit disclosure about whatever is done*: Some have introduced checklist-style forms requiring authors to affirm that the paper discloses explicitly all relevant experimental and analytic decisions. Second is *certification and endorsement of a virtuous practice*, to provide a noncompulsory incentive toward adoption of the practice. *Psychological Science* and other journals offer researchers the opportunity to display badges certifying virtuous behavior. Third is *mandating virtuous practices*: While so far in psychology, requiring data deposit as a condition of publication is presently limited to a few, lower profile journals (e.g., *Archives of Scientific Psychology*; Cooper and VandenBos 2013), it is now standard policy of some leading economics journals (American Economic Association 2016).

STATISTICAL OBJECTIVITY II: CULTIVATING POPULATIONS OF STUDIES

As we have shown, threats to the objectivity of literatures have been made visible through the aggregate analysis of multiple studies. In response, incentivizing disclosure can improve confidence in individual studies. Increasing and standardizing the information available about a study also increases the capacity to recast individual studies as mere data points in larger and potentially more authoritative data sets. This ramping up of statistical thinking in fields already dominated by statistical methods produces a second-order quantification in which even a quantitative scientific finding only finds value to the degree it contributes to eventual meta-analysis. Rather than a rhetorical object with arguments and conclusions, the study is recast as simply a set of quantitative inputs from which, once aggregated, more objective conclusions may be derived. Statistical objectivity seeks to replace the logic of the "crucial experiment" with a logic of ongoing accumulation and assessment.

This transformation toward thinking in terms of populations of studies rather than individual findings has been a growing concern in the area of "evidence-based medicine" (EBM) (Timmermans and Berg 2003). Advocates of EBM have argued that medical decisions should be based on syntheses of the literature based on a "hierarchy of evidence" in which unsystematic methods like case reports are given little weight compared to more systematic methods like randomized control trials (Knaapen 2013). At the top of the hierarchy, however, are meta-analyses of randomized control trials. This is a method that aggregates multiple studies on the same topic into a single data set, which is taken to reduce bias and limited generalizability of potentially idiosyncratic individual studies.

This demotion of findings to mere data points is evidenced in two closely related developments: (1) growing calls for replication studies that follow practices of original studies as closely as possible and (2) a reorientation toward "cumulative estimation" in which researchers attempt to draw defensibly objective conclusions from populations of studies using the conceptual and methodological tools of meta-analysis.

Mechanical Replication

Grounding objectivity in populations of studies requires, first and foremost, that a population of studies exists. Cultivating this population of studies thus entails "replication" studies. As Collins (1985) made clear, however, whenever results of an intended replication diverge from those of an original study, an essential interpretive ambiguity is posed: Should the divergence be understood as evidence against the credibility of the original study, or should the divergence be explained by the differences in how the two studies were conducted?

Some psychologists use the term *conceptual replication* to refer to a study that employs a deliberately dissimilar research design to address the same hypothesis (Yong 2012b). Successful conceptual replications can be interpreted as strengthening an initial result by showing it to be robust to alternative operationalizations. Yet for purposes of cumulative estimation, "conceptual replications" are problematic because a "failed conceptual replication" is a practical oxymoron: Since study practices were deliberately intended to be dissimilar, any difference in outcome can easily be attributed to those dissimilarities. "Conceptual replications" can thus be dismissed by critics as intrinsically incapable of speaking to the credibility of the original study as they typically can only be published when successful (Freese and Peterson 2017; LeBel and Peters 2011; Pashler and Harris 2012).

As a result, the possibility of cumulative estimation entails replications that are as similar to the original study as can be logistically achieved. These have been referred to as "exact," "direct," or "close" replications, reflecting different levels of authorial optimism about the level of similarity (Cooper 2016; Finkel, Eastwick, and Reis 2014). We remain agnostic and call such studies *mechanical replications* to highlight their grounding in the basic logic of mechanical objectivity. The key principle is that researchers subordinate their own judgments to those of the authors of the original study, being as self-consciously *noncreative* as possible. By maximizing similarity, the study maximizes its commensurability for a cumulative estimation and thus also the extent to which results of the second study may be used to evaluate the credibility of the first. The desire to increase the number of mechanical replications motivated the largest replication attempt in history, the OSC's (2015) attempt to replicate 100 recent psychology studies. The project was hailed by *Science* as one of the top 10 scientific breakthroughs of 2015 (Stokstad et al. 2015).

However, the deliberate lack of creativity in mechanical replications presents an especially acute incentive problem in fields that prize novelty. Many esteemed psychology journals have simply refused to consider direct replication studies (Aldhous 2011), making incentives for mechanical replication very low. One study found a 1 percent rate of replication in psychological research since the year 1900 (Makel et al. 2012).

The lack of incentives for mechanical replications also reduces their credibility by increasing the plausibility of the interpretations for failed replications that focus on the competence or motivation of investigators. When a replication failed to substantiate one of the seminal experiments in social priming (Bargh, Chen, and Burrows 1996; Doyen et al. 2012), the original author lambasted the replicators for shoddy work and the journal for inadequate oversight (Satel 2013; Yong 2012a). The aforementioned effort to replicate 100 psychology studies was criticized for "permitting considerable infidelities that almost certainly biased the replication studies toward failure" (Gilbert et al. 2016:1037). Mechanical replications are often posed as training exercises for students who work in a lab, but when results diverge, their inexperience provides an easy rejoinder. For example, social psychologist Dijksterhuis

(2013) attributed the null results of a replication study of his findings as flawed by inclusion of "student projects" replete with "beginners' mistakes."

Thus, although forensic demonstrations provoke a push for more mechanical replications, this push confronts several incentive problems. To increase the benefits to doing replications, activists have successfully pressed both the *Journal of Personality and Social Psychology* and *Social Psychology and Personality Science* to change policies and entertain submission of mechanical replication studies (Cooper 2016; Vazire 2016).

For reducing the cost to doing replications, activists have encouraged authors of original studies to publicly deposit materials at the time of publication. For increasing the credibility of replications that are done, numerous strategies have been offered, many of which illustrate how activists have conceptualized the problem and its solution in terms of configurations of incentives. As one example, when Kahneman (2012) warned investigators working on social priming that there was a "trainwreck looming" in regards to their replicability, he recommended that they set up a system in which each participating lab would commit resources to conducting mechanical replications of the original findings of another participating lab. The goal of this system would be to strengthen the credibility of those findings by providing a record of their independent mechanical replication by another lab that had demonstrated expertise in conducting the type of experiment in question. As a second example, a special issue of the journal *Social Psychology* pioneered "registered reports" (Nosek and Lakens 2014), in which investigators provided proposals that detailed data collection and analysis plans for the mechanical replication of an important published finding. These proposals were then peer reviewed and then conditionally accepted for publication *before* any data were collected and thus irrespective of their results.

Meta-analytic Fundamentalism

"Cumulative science patiently awaits the meta-analysis," write Moffitt, Caspi, and Rutter (2006). Because replication attempts so regularly yield inconsistent results, epistemic activists have strongly urged against placing much confidence in new studies before success in replication studies is demonstrated. As one writer explains, "The problem isn't that many studies fail to replicate. It's that we believe in them before they've been thoroughly vetted" (Adler 2014).

Literature reviews cultivated by experts are the traditional means of performing this "vetting," but of course, such expert judgments are subject to the same perceived dangers of subjective bias that often prompt more formal quantitative research designs in the first place (Hunt 1997; Light and Pillemer 1984). "Meta-analysis" encompasses several increasingly prevalent quantitative techniques that seek to draw objective conclusions about real-world relationships by combining results of multiple studies (Egger, Davey Smith, and O'Rourke 2001).

Meta-analysis is neither new nor new to psychology. Important to the history of the method were efforts to draw objective conclusions from the inconsistent literatures both in parapsychology experiments and assessments of the efficacy of psychotherapy (Chalmers, Hedges, and Cooper 2002; O'Rourke 2007; Pratt et al. 1940; Smith and Glass 1977).

Yet, the prospective shift toward "meta-analytic thinking" is novel enough to serve as a cornerstone of what has been called psychology's "new statistics" (Cumming 2013:17). What is putatively new about "new statistics" is not the formal tools of meta-analysis but rather its reconceptualization of how results from individual studies are to be understood. The reconceptualization changes the locus of objectivity to the collective analysis of multiple studies and in doing so, seeks to alter the reporting of individual studies so that they may be brought in line with the ultimate authority of meta-analysis.

Meta-analysis is made more powerful by the changes described previously, especially by (1) standardizing analysis details reported by studies and (2) increasing and improving mechanical replication by making study materials publicly available on publication. In its approach to data analysis, meta-analytic fundamentalism promotes changes that displace the traditional emphasis on null hypothesis testing with a Bayesian-style approach to evidence. Basic ideas of Bayesian statistics are old—Bayes's theorem dates to 1763—but awareness of Bayesian methods has exploded in the past two decades as computational advances have greatly increased their practical availability (on the growth of Bayesian methods in psychology, see e.g., Andrews and Baguley 2013).

Regardless of whether researchers explicitly use Bayesian methods, a philosophically Bayesian style of reasoning motivates this emergent interpretation of individual experiments. Rather than each study providing a "finding," results merely increase or decrease the likelihood of some hypotheses being true versus others, with stronger evidence changing these likelihoods more. Meta-analysis can then be understood as simply extending this principle, articulating the likelihood of hypotheses being true and updating for the separate contributions of each study that is included.

Once conceived as such, meta-analysis becomes the apex of objectivity (Stegenga 2011). By combining three distinct, escalating virtues, meta-analysis is argued to be the most rational conclusion that may be drawn given available evidence. First, meta-analytic conclusions are necessarily based on more information than the conclusions of any single study included in the meta-analysis. For this reason, meta-analyses have played an important role in overall judgments of evidence strength for the purposes of implementation of clinical practice recommendations. Second, since they aggregate studies from different investigators, meta-analyses can be seen as potentially transcending individual biases that afflict particular investigations. Third—and bringing us back to the logic of forensic analytics presented at the outset—meta-analysis allows for the possibility that collective analysis of results may produce evidence of systematic biases, such as the "file drawer problem" of unreported studies (described in Table 1). This in turn raised the possibility of adjustments that attempt to "correct" these problems, as with methods that attempt to correct for the exaggeration of effect sizes revealed by biased funnel plots (Stanley and Doucouliagos 2013).

Putting matters together, then, meta-analysis introduces the prospect of shared accountability for literatures in which bad collective properties may undermine credibility even without revealing specific flaws of specific studies. Once this threat is posed, meta-analysis may be seen as the fundamental tool by which literatures can be collectively assessed, interpreted, and perhaps even rescued.

Meta-analysis and Objectivity

Statistical objectivity attempts to overcome subjectivity by aggregating individual knowledge claims. While this may seem like a triumph of mechanical procedure over expert judgment, in fact, this only moves the locus of expert subjectivity to a different plane: the choices and interpretations inevitably involved in carrying out meta-analyses.

In perhaps the earliest example, when a meta-analysis called into doubt work by Hans Eysenck (Smith and Glass 1977), he lashed out in familiar terms against what he labeled "meta-silliness":

> [Smith and Glass] advocate and practice the abandonment of critical judgments of any kind. A mass of reports—good, bad, and indifferent—are fed into the computer in the hope that people will cease caring about the quality of the material on which the

> conclusions are based. If their abandonment of scholarship were to be taken seriously . . . it would mark the beginning of a passage into the dark age of scientific psychology. (Eysenck 1978, P. 517)

In seeking to avoid bias by including all relevant studies, Smith and Glass opened themselves to the critique that they were abdicating their role as experts to evaluate the quality of studies and in so doing, produced biased results.

We find the tension between objectivity and expertise recapitulated even more fully in more recent controversies. A high-profile example from psychology concerns a finding that a specific genetic variant moderates the relationship between stressful life events and depression (Caspi et al. 2003). Subsequent replication attempts were a confusing mix of successes, partial successes, and failures. When a research team reported a meta-analysis that yielded negative findings (Risch et al. 2009), the primary authors of the original study responded with their own meta-analysis, arguing that the null finding was the result of overly selective criteria regarding which studies ought to be included (Caspi et al. 2010). Additional meta-analyses both strengthening and undermining the original finding were reported by others (Duncan and Keller 2011; Karg et al. 2011). More recently, in what may be considered an attempt to settle these arguments through a move to regulatory objectivity, experts were brought together to formulate a consensus document about how an authoritative meta-analysis should proceed (Culverhouse et al. 2013). This was immediately criticized as biased from the outset by the original authors (Moffitt and Caspi 2014).[3]

Meta-analysis is not a foolproof method of marginalizing subjectivity because populations of studies do not build themselves. Impactful choices are made regarding study similarity ("Are these studies actually testing the same hypothesis?") and quality ("Should higher quality studies count more? Should some studies be omitted entirely?") (Eysenck 1994; Knaapen 2013; Moreira 2007; Stegenga 2011; Will 2009). These decisions involve "meta-expertise" (Collins and Evans 2007) that cannot be adjudicated through purely objective criteria. This has led to the phenomenon of "dueling meta-analyses" in which experts using different selection or analytic criteria produce meta-analyses of the same body of studies that arrive at opposing conclusions. Describing one such episode, Tabery (2014:87) laments, "The meta-analyses were supposed to provide the meta-solution, but instead they only elevated it to a meta-problem."

Such developments highlight how objectivity is like an ouroboros, the snake that eats its own tail. Statistical objectivity neither summarizes nor supersedes earlier developments in the history of objectivity. Rather, it creates a new level on which prior developments may repeat themselves, only in the analysis of sets of studies rather than the analysis of primary data. Ioannidis, so influential in raising concerns about false positive findings, has more recently complained about an "epidemic" of flawed meta-analyses driven by author biases (Chawla 2016). This is not to say newer inquiries do not represent progress. They may, but the logic behind these products and the methods used to produce them invoke strategies of disciplining subjectivity that are at once novel and familiar.

DISCUSSION

We have described a set of developments that together constitute a newly coherent epistemic virtue that is emerging in various scientific fields. With its emphasis on transparency and the ongoing aggregation of findings, statistical objectivity offers researchers a potent package that includes a set of analytic tools, philosophy of science, and ethic of scientific disclosure. These elements are not only complementary but mutually reinforcing (Figure 3).

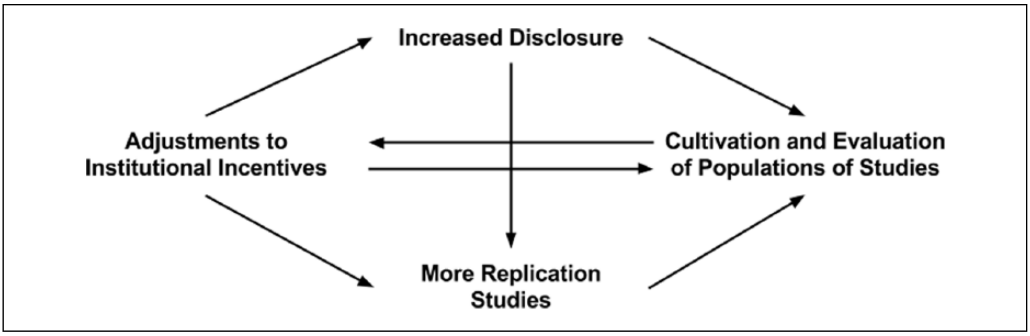

Figure 3. The reinforcing character of statistical objectivity.

Forensic demonstrations reveal problems of bias, cast doubt on entire literatures, and suggest the need for new epistemic virtues. An economic understanding of the problem then motivates efforts to adjust institutional incentives. Adjusted institutional incentives promote greater disclosure of experimental and analytic detail, which both makes it easier for other researchers to conduct replication studies and improves the information available when aggregating studies (e.g., meta-analyses). The possibility of meta-analyses in turn affords the view that individual studies are inherently tentative and so promotes adjusting incentives further to increase the potential scope and power of meta-analytic methods.

The reinforcing character of statistical objectivity shown in Figure 3 is significant because it implies that half-measures anywhere can result in weakness throughout. Thus, the call for a wide-ranging overhaul of prevailing practices is a frequent refrain. Everything from graduate training to journal reviewing is implicated.

Like other experts, scientists must often issue abstract judgments in ambiguous conditions (Abbott 1988), and the difficulty of evaluating expert judgments can lead to challenges to their authority. Fields often meet this challenge by setting up internal methods to govern behavior and maintain standards (Fligstein and McAdam 2012). In scientific communities, peer review is often restricted to a "core set" of scientists who are deemed the sole legitimate arbiters of research in a specific domain (Collins 1981; Collins and Evans 2002). Although academic fields are structured by the same competitive dynamics that structural all fields, there is also a shared belief in the "rules of the game," including common conceptual tools (Fleck 1981), tastes (Shapin 2012), and evidentiary standards (Collins 1998).

Bourdieu (2004) has labeled this shared culture that gives fields their patina of fairness the *illusio*. Although scientific fields are composed around different concepts, tastes, and standards, Bourdieu argues that they all share a purported belief in the value of disinterestedness (see also, Grundmann 2013; Merton 1973). Scientific experts are trusted to the degree that they are neutral arbiters of the real. When their neutrality is doubted, the fragile illusio falls away and reveals what appears to be a naked fight for power and resources. Distrust of expertise often leads to social movements within science that attempt to reformulate the constitution of scientific capital (Bourdieu 2004; Frickel and Gross 2005). Although the specifics of these movements differ, they often frame themselves as "returning to the sources, the origin, the spirit, the authentic essence of the game, in opposition to the banalization and degradation which is has suffered" (Bourdieu 1993:74).

For scientific social movements, this "return" is often formulated as a recommitment to the epistemic virtues of objectivity. However, as Daston and Galison (2007) have argued, the meaning of objectivity has been articulated differently throughout the history of science, changing in response to some newly perceived threat of subjectivity. Activists advocating statistical objectivity have framed the threat in terms of the systematic bias that can occur in fields when incentives are not properly aligned. This provides a new framing of the scientific self that breaks sharply from the narratives of heroic self-discipline and self-correcting institutions that have defined the self-presentation of scientists through the modern era (Shapin 2010). In its place, statistical objectivity poses an image of the scientist as a flawed character driven by worldly incentives into destructive epistemic vices.

Beyond framing, advocates of statistical objectivity have proposed various reforms focused on disclosure, mechanical replication, and meta-analysis. Quantification, which has often been viewed as a key aspect of objectivity (Espeland and Stevens 2008; Porter 1995), is again at the center of the discussion. Although statistical methods are dominant across the areas we discussed, statistical objectivity represents a new layer of quantification: the quantification of quantitative literatures. Where experts were expected to translate their quantitative findings into narratives, in statistical objectivity, those findings are simply absorbed into ever-growing datasets.

The emergence of statistical objectivity presents new questions for social scientists. For instance, the goal of preventing false positives may seem like an unambiguous good; when trying to decide whether treatments work or what criteria should be used in screening or prediction, the costs of false negatives are hardly an idle concern. Moreover, evaluating the success of replication is far more complex in fields that require high levels of local or embodied knowledge (e.g., Doing 2004), and "transparency" may be difficult to conceptualize in fields in which a single article may involve bringing together the work of a hundred or more co-authors. And returning to the tension between objectivity and expertise, the reforms detailed in this article constrain expert autonomy. How will rank-and-file researchers respond to constraints on their expert authority to do things like edit their own data (Leahey 2008) or increasing demands for a level of transparency that may seem unnecessary or invasive? Can the ethic of projective disclosure be integrated into everyday research, or will it be viewed as another layer of bureaucracy to be completed with perfunctory formalism (Smith-Doerr and Vardi 2015; Zimmerman 2008)?

NOTES

1. Like many scientific and intellectual social movements (Frickel and Gross 2005), the scholar "activists" we investigate are a heterogeneous group. As their efforts represent an attempt to transform the epistemic cultures of scientific fields, we use the term *epistemic activists* throughout the article.
2. Examples of other social psychology journals that have adopted checklists include *Social Psychology and Personality Science* and the *Journal of Experimental Social Psychology*.
3. Similar battle lines are being drawn in reaction to the Open Science Collaboration's (2015) ambitious replication attempt. Rather than accept the article's conclusion of widespread methodological problems, a prominent group of social scientists (Gilbert et al. 2016:1037) reanalyzed the replications using different statistical assumptions and argued that nothing contradicts the conclusion that "the reproducibility of psychological science is quite high." This sparked a series of critiques about the critique (for a summary and links, see Srivastava 2016).

REFERENCES

Abbott, Andrew. 1988. *The System of Professions: An Essay on the Division of Expert Labor*. Chicago, IL: University of Chicago Press

Adler, Jerry. 2014. "The Reformation: Can Social Sciences Save Themselves?" *Pacific Standard*. Retrieved January 6, 2015 (http://www.psmag.com/navigation/health-and-behavior/can-social-scientists-save-themselves-human-behavior-78858/).

Aldhous, Peter. 2011. "Journal Rejects Studies Contradicting Precognition." *New Scientist*. Retrieved July 24, 2018 (http://www.newscientist.com/article/dn20447-journal-rejects-studies-contradicting-precognition.html).

American Economic Association. 2016. "Data Availability Policy." Retrieved April 20, 2016 (https://www.aeaweb.org/journals/policies/data-availability-policy).

American Psychological Association. 2017. "APA Journals Program Collaborates with Center for Open Science to Advance Open Science Practices in Psychological Research." Retrieved August 25, 2017 (http://www.apa.org/news/press/releases/2017/08/open-science.aspx).

Andrews, Mark, and Thom Baguley. 2013. "Prior Approval: The Growth of Bayesian Methods in Psychology." *British Journal of Mathematical and Statistical Psychology* 66(1):1–7.

Baker, Monya. 2015. "First Results from Psychology's Largest Reproducibility Test." *Nature*. Retrieved August 18, 2018 (https://www.nature.com/news/first-results-from-psychology-s-largest-reproducibility-test-1.17433).

Bargh, John A., Mark Chen, and Lara Burrows. 1996. "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action." *Journal of Personality and Social Psychology* 71(2):230–44.

Begley, C. Glenn, and Lee M. Ellis. 2012. "Drug Development: Raise Standards for Preclinical Cancer Research." *Nature* 483(7391):531–33.

Bem, Daryl J. 2011. "Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect." *Journal of Personal and Social Psychology* 100(3):407–25.

Benford, Robert D., and David A. Snow. 2000. "Framing Processes and Social Movements: An Overview and Assessment." *Annual Review of Sociology* 26:611–39.

Benjamin, Daniel J., James Berger, Magnus Johannesson, Brian A. Nosek, Eric-Jan Wagenmakers, et al. 2017. "Redefine Statistical Significance." *Nature Human Behaviour* 2(6):6–10.

Berg, Marc, Klasien Horstman, Saskia Plass, and Michelle van Heusden. 2000. "Guidelines, Professionals and the Production of Objectivity: Standardisation and the Professionalization of Insurance Medicine." *Sociology of Health and Illness* 22(6):765–91.

Borsboom, Denny, Han van der Maas, and Eric-Jan Wagenmakers. 2014. "Questions and Answers about the Förster Case." Retrieved July 24, 2018 (http://osc.centerforopenscience.org/2014/05/29/forster-case/).

Bourdieu, Pierre. 1993. *Sociology in Question*. London: Sage.

Bourdieu, Pierre. 2004. *Science of Science and Reflexivity*. New York: Polity.

Burri, Regula V., and Joseph Dumit. 2008. "Social Studies of Scientific Imaging and Visualization." Pp. 297–317 in *The Handbook of Science and Technology Studies*. 3rd ed., edited by E. J. Hackett, O. Amsterdamska, M. Lynch, and J. Wajcman. Cambridge, MA: The MIT Press.

Button, Katherine S., John P. A. Ioannidis, Clair Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. 2013. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews: Neuroscience* 14(5):365–76.

Cambrosio, Alberto, Peter Keating, Thomas Schlich, and George Weisz. 2006. "Regulatory Objectivity and the Generation and Management of Evidence in Medicine." *Social Science and Medicine* 63(1):189–99.

Cambrosio, Alberto, Peter Keating, Thomas Schlich, and George Weisz. 2009. "Biomedical Conventions and Regulatory Objectivity: A Few Introductory Remarks." *Social Studies of Science* 39(5):651–64.

Caspi, Avshalom, Ahmad R. Hariri, Andrew Holmes, Rudolf Uher, and Terrie E. Moffitt. 2010. "Genetic Sensitivity to the Environment: The Case of the Serotonin Transporter Gene and Its Implications for Studying Complex Diseases and Traits." *American Journal of Psychiatry* 167(5):509–27.

Caspi, Avshalom, Karen Sugden, Terrie E. Moffitt, Alan Taylor, Ian W. Craig, HonaLee Harrington, Joseph McClay, Jonathan Mill, Judy Martin, Antony Braithwaite, and Richie Poulton. 2003. "Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-Htt Gene." *Science* 301(5631):386–89.

Center for Open Science. 2013. "Badges to Acknowledge Open Practices." Retrieved January 6, 2015 (https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/).

Center for Open Science. 2015. "Transparency and Openness Promotion (TOP) Guidelines." Retrieved March 11, 2016 (https://cos.io/top/#signatories).

Chalmers, Iain, Larry V. Hedges, and Harris Cooper. 2002. "A Brief History of Research Synthesis." *Evaluation and the Health Professions* 25(1):12–37.

Chambers, Chris. 2014. “Physics Envy: Do ‘Hard’ Sciences Hold the Solution to the Replication Crisis in Psychology?” *The Guardian*. Retrieved February 12, 2015 (http://www.theguardian.com/science/head-quarters/2014/jun/10/physics-envy-do-hard-sciences-hold-the-solution-to-the-replication-crisis-in-psychology).

Chawla, Dalmeet Singh. 2016. “We Have an Epidemic of Deeply Flawed Meta-analysis, says John Ioannidis.” *Retraction Watch*. Retrieved September 5, 2017 (http://retractionwatch.com/2016/09/13/we-have-an-epidemic-of-deeply-flawed-meta-analyses-says-john-ioannidis/).

Cole, Simon A. 1998. “Witnessing Identification: Latent Fingerprinting Evidence and Expert Knowledge.” *Social Studies of Science* 28(5–6):687–712.

Collins, Harry M. 1981. “The Place of the Core Set in Modern Science: Social Contingency with Methodological Propriety.” *History of Science* 19(1):6–19.

Collins, Harry M. 1985. *Changing Order: Replication and Induction in Scientific Practice*. London: Sage Publications.

Collins, Harry M. 1998. “The Meaning of Data: Open and Closed Evidential Cultures in the Search for Gravitational Waves.” *American Journal of Sociology* 104(2):293–338.

Collins, Harry M. 2004. *Gravity’s Shadow: The Search for Gravitational Waves*. Chicago: University of Chicago Press.

Collins, Harry M., and Robert Evans. 2002. “The Third Wave of Science Studies: Studies of Expertise and Experience.” *Social Studies of Science* 32(2):235–96.

Collins, Harry M., and Robert Evans. 2007. *Rethinking Expertise*. Chicago: University of Chicago Press.

Cooper, Harris, and Gary R. VandenBos. 2013. “Archives of Scientific Psychology: A New Journal for a New Era.” *Archives of Scientific Psychology* 1(1):1–6.

Cooper, M. Lynne. 2016. “Editorial.” *Journal of Personality and Social Psychology* 110(3):431–34.

Culverhouse, Robert C., Lucy Bowes, Naomi Breslau, John I. Nurnberger, Jr., Margit Burbeister, David M. Fergusson, Marcus Munafó, Nancy L. Saccone, and Laura J. Beirut. 2013. “Protocol for a Collaborative Meta-analysis of 5-HTTLPR, Stress, and Depression.” *BMC Psychiatry* 13(304):1–11.

Cumming, Geoff. 2013. “The New Statistics: Why and How.” *Psychological Science* 25(1):7–29.

Danziger, Kurt. 1990. *Constructing the Subject: Historical Origins of Psychological Research*. Cambridge, UK: Cambridge University Press.

Daston, Lorraine. 1992. “Objectivity and the Escape from Perspective.” *Social Studies of Science* 22(4):597–618.

Daston, Lorraine, and Peter Galison. 1992. “The Image of Objectivity.” *Representations* 40(Autumn):81–128.

Daston, Lorraine, and Peter Galison. 2007. *Objectivity*. New York: Zone Books.

Dijksterhuis, Ap. 2013. “Replication Crisis or Crisis in Replication? A Reinterpretation of Shanks et al.” *PLoS One*. Retrieved January 6, 2015 (http://www.plosone.org/annotation/listThread.action?root=64751).

Doing, Park. 2004. “‘Lab Hands’ and the ‘Scarlet O’ Epistemic Politics and (Scientific) Labor.” *Social Studies of Science* 34(3):299–323.

Doyen, Stéphane, Olivier Klein, Cora-Lise Pichon, and Axel Cleeremans. 2012. “Behavior Priming: It’s All in the Mind, but Whose Mind?” *PLoS One* 7(1):e29081.

Duncan, Laramie E., and Matthew C. Keller. 2011. “A Critical Review of the First 10 Years of Candidate Gene-by-environment Interaction Research in Psychiatry.” *American Journal of Psychiatry* 168(10):1041–49.

Durante, Kristina M., Ashley Rae, and Vlasdas Griskevicius. 2013. “The Fluctuating Female Vote: Politics, Religion, and the Ovulatory Cycle.” *Psychological Science* 24(6):1007–16.

Durtschi, Cindy, William Hillison, and Carl Pacini. 2004. “The Effective Use of Benford’s Law to Assist in Detecting Fraud in Accounting Data.” *Journal of Forensic Accounting* 5(1):17–34.

Edwards, Paul N., Matthew S. Mayernik, Archer L. Batcheller, Geoffrey C. Bowker, and Christine L. Borgman. 2011. “Science Friction: Data, Metadata, and Collaboration.” *Social Studies of Science* 41(5):667–90.

Egger, Matthias, George Davey Smith, and Keith O’Rourke. 2001. “Introduction: Rationale, Potentials, and Promise of Systematic Reviews.” Pp. 1–19 in *Systematic Reviews in Health Care: Meta-Analysis in Context*, edited by M. Egger, G. Davey Smith, and D. G. Altman.

Eich, Eric. 2014. “Business Not as Usual.” *Psychological Science* 25(1):3–6.

Espeland, Wendy N., and Mitchell L. Stevens. 2008. "A Sociology of Quantification." *European Journal of Sociology* 49(3):401–36.

Everett, Jim A. C., and Brian D. Earp. 2015. "A Tragedy of the (Academic) Commons: Interpreting the Replication Crisis in Psychology as a Social Dilemma for Early-career Researchers." *Frontiers in Psychology* 6(August):1–4.

Eysenck, Hans. 1978. "An Exercise in Mega-Silliness." *American Psychologist* 33(5):517.

Eysenck, Hans. J. 1994. "Meta-analysis and Its Problems." *BMJ* 309(6957):789–92.

Fleck, Ludwik. 1981. *The Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press.

Fligstein, Neil, and Doug McAdam. 2012. *A Theory of Fields*. Oxford, UK: Oxford University Press.

Finkel, Eli J., Paul W. Eastwick, and Harry T. Reis. 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):275–97.

Freese, Jeremy, and David Peterson. 2017. "Replication in Social Science." *Annual Review of Sociology* 43:147–65.

Frickel, Scott, and Neil Gross. 2005. "A General Theory of Scientific/Intellectual Movements." *American Sociological Review* 70(2):204–32.

Frow, Emma K. 2012. "Drawing a Line: Setting Guidelines for Digital Image Processing in Scientific Journal Articles." *Social Studies of Science* 42(3):369–92.

Gelman, Andrew, and John Carlin. 2014. "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors." *Perspectives on Psychological Science* 9(6):641–51.

Gelman, Andrew, and Eric Loken. 2014. "The Statistical Crisis in Science." *American Scientist* 102(6): 460–65.

Gilbert, Daniel T., Gary King, Stephen Pettigrew, and Timothy D. Wilson. 2016. "Comment on 'Estimating the Reproducibility of Psychological Science.'" *Science* 351(6277):1037.

Gilbert, G. Nigel, and Michael Mulkay. 1984. *Opening Pandora's Box: A Sociological Analysis of Scientists' Discourse*. Cambridge, UK: Cambridge University Press.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor Books.

Greene, Casey S., Nadia M. Pendrod, Scott M. Williams, and Jason H. Moore. 2009. "Failure to Replicate a Genetic Association May Provide Important Clues about Genetic Architecture." *PLoS One* 4(6):e5639.

Grundmann, Reiner. 2013. "'Climategate' and the Scientific Ethos." *Science, Technology, & Human Values* 38(1):67–93.

Healy, Kieran. 2017. "Fuck Nuance." *Sociological Theory* 35(2):118–27.

Holton, Gerald. 1978. "Subelectrons, Presuppositions, and the Millikan-Ehrenhaft Dispute." *Historical Studies in the Physical Sciences* 9:161–224.

Hughes, Virginia. 2014. "Simple Steps Aim to Solve Science's 'Reproducibility Problem.'" *Spectrumnews.com*. Retrieved August 14, 2019 (https://www.spectrumnews.org/news/simple-steps-aim-to-solve-sciences-reproducibility-problem/).

Hunt, Morton. 1997. *How Science Takes Stock: The Story of Meta-analysis*. New York: Russell Sage Foundation.

Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2(8):e124.

John, Leslie K., George Loewenstein, and Drazen Prelec. 2012. "Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling." *Psychological Science* 23(5):524–32.

Kahneman, Daniel. 2012. "A Proposal to Deal with Questions about Priming Effects." Retrieved July 24, 2018 (http://www.nature.com/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman%20Letter.pdf).

Karg, Katja, Margit Burmeister, Kerby Shedden, and Srijan Sen. 2011. "The Serotonin Transporter Variant (5-HTTLPR), Stress, and Depression Meta-analysis Revisited: Evidence of Genetic Moderation." *JAMA Psychiatry* 68(5):444–54.

Klein, Richard A., Kate A. Ratliff, Michelangelo Vianello, Reginald B. Adams, Jr., Štěpán Bahník, et al. 2014. "Investigating Variation in Replicability: A 'Many Labs' Replication Project." *Social Psychology* 45(3):142–52.

Knaapen, Loes. 2013. "Being 'Evidence-based' in the Absence of Evidence: The Management of Non-evidence in Guideline Development." *Social Studies of Science* 43(5):681–706.

Knorr Cetina, Karin. 1983. "The Ethnographic Study of Scientific Work: Towards a Constructivist Interpretation of Science." Pp. 115–40 in *Science Observed: Perspectives on the Social Studies of Science*, edited by K. Knorr-Centina. London: Sage.

Knorr Cetina, Karin. 1995. "Laboratory Studies: The Cultural Approach to the Study of Science." Pp. 140–67 in *The Handbook of Science and Technology Studies*, edited by S. Jasanoff, G. E. Markle, J. C. Petersen, and T. Pinch. Thousand Oaks, CA: Sage.

Krippner, Greta R. 2007. "The Making of US Monetary Policy: Central Bank Transparency and the Neoliberal Dilemma." *Theory and Society* 36(6):477–513.

Kruse, Corinna. 2012. "The Bayesian Approach to Forensic Evidence: Evaluating, Communicating, and Distributing Responsibility." *Social Studies of Science* 43(5):657–80.

Kuhn, Thomas Samuel. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Leahey, Erin. 2008. "Overseeing Research Practice: The Case of Data Editing." *Science, Technology and Human Values* 33(5):605–30.

LeBel, Etienne P., Denny Borsboom, Roger Giner-Sorolla, Fred Hasselman, Kurt R. Peters, Kate A. Ratliff, and Colin Tucker Smith. 2013. "PsychDisclosure.org: Grassroots Support for Reforming Reporting Standards in Psychology." *Perspectives on Psychological Science* 8(4):424–32.

LeBel, Etienne P., and Kurt R. Peters. 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(4):371–79.

Light, Richard J., and David B. Pillemer. 1984. *Summing up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press.

Love, Jessica. 2013. "The Allure of the Counterintuitive." *The American Scholar*. Retrieved July 24, 2018 (https://theamericanscholar.org/the-allure-of-the-counterintuitive/).

Lynch, Michael, Simon A. Cole, Ruth McNally, and Kathleen Jordan.2008. *Truth Machine: The Contentious History of DNA Fingerprinting*. Chicago: University of Chicago Press.

Lyon, David. 2001. *Surveillance Society: Monitoring Everyday Life*. Buckingham, UK: Open University Press.

Makel, Matthew C., Jonathan A. Plucker, and Boyd Hegarty. 2012. "Replications in Psychology Research: How Often Do They Really Occur?" *Perspectives on Psychological Science* 7(6):537–42.

Merton, Robert K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago press.

Meyer, John. W., and Patricia Bromley. 2014. "The Worldwide Expansion of 'Organization.'" *Sociological Theory* 31(4):366–89.

Miguel, Edward, Colin Camerer, Katherine Casey, Joshua Cohen, Kevin M. Esterling, Alan Gerber, Rachel Glennerster, D. P. Green, M. Humphreys, G. Imbens, and D. Laitin. 2014. "Promoting Transparency in Social Science Research." *Science* 343(6166):30–31.

Moffitt, Terrie E., and Avshalom Caspi. 2014. "Bias in a Protocol for a Meta-analysis of 5-HTTLPR, Stress, and Depression." *BMC Psychiatry* 14(1):179.

Moffitt, Terrie E., Avshalom Caspi, and Michael Rutter. 2006. "Measured Gene-environment Interactions in Psychopathology: Concepts, Research Strategies, and Implications for Research, Intervention, and Public Understanding of Genetics." *Perspectives on Psychological Science* 1(1):5–27.

Montgomery, Paul, Sean Grant, Sally Hopewell, Geraldine Macdonald, David Moher, Susan Michie, and Evan Mayo-Wilson. 2013. "Protocol for CONSORT-SPI: An Extension for Social and Psychological Interventions." *Implementation Science* 8(99):1–7.

Moreira, Tiago. 2007. "Entangled Evidence: Knowledge Making in Systematic Reviews in Healthcare." *Sociology of Health and Illness* 29(2):180–97.

Nosek, Brian A., and Daniel Lakens. 2014. "Registered Reports: A Method to Increase the Credibility of Published Results." *Social Psychology* 45(3):137–41.

Nosek, Brian A., Jeffrey R. Spies, and Mathew Motyl. 2012. "Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth of Publishability." *Perspectives on Psychological Science* 7(6):615–31.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349(6251):943.

O'Rourke, Keith. 2007. "An Historical Perspective on Meta-analysis: Dealing Quantitatively with Varying Study Results." *Journal of the Royal Society of Medicine* 100(12):579–82.

Pashler, Harold, and Christine R. Harris. 2012. "Is the Replicability Crisis Overblown? Three Arguments Examined." *Perspectives on Psychological Science* 7(6):531–36.

Peterson, David. 2015. "All That Is Solid: Bench-building at the Frontiers of Two Experimental Sciences." *American Sociological Review* 80(6):1201–25.

Porter, Theodore M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Posner, Eric. 2014. "The Lewisization/Gladwellization of Social Science." Retrieved January 5, 2015 (http://ericposner.com/the-lewisizationgladwellization-of-social-science/).

Pratt, Joseph Gaither, Joseph Banks Rhine, Burke M. Smith, Charles Edward Stuart, and Joseph Albert Greenwood. 1940. *Extra-sensory Perception after Sixty Years: A Critical Appraisal of the Research in Extra-sensory Perception*. New York: Henry Holt and Company.

Prinz, Florian, Thomas Schlange, and Khusru Asadullah. 2011. "Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets?" *Nature*. Retrieved January 6, 2015 (http://www.nature.com/nrd/journal/v10/n9/full/nrd3439-c1.html).

Reinhart, Martin. 2016. "Reproducibility in the Social Sciences." Pp. 407–23 in *Reproducibility: Principles, Problems, Practices, and Prospects*, edited by H. Atmanspacher and S. Maasen. Hoboken, NJ: Wiley.

Rietveld, Cornelius A., Sarah E. Medland, Jaime Derringer, Jian Yang, Tõnu Esko, Nicolas W. Martin, Harm-Jan Westra, et al. 2013. "GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment." *Science* 340(6139):1467–71.

Risch, Neil, Richard Gerrell, Thomas Lehner, Kung-Yee Liang, Lindon Eaves, Josephine Hoh, Andrea Griem, Maria Kovacs, Jurg Ott, and Kathleen Ries Merikangas. 2009. "Interaction Between the Serotonin Transporter Gene (5-HTTLPR), Stressful Life Events, and Risk of Depression." *JAMA* 301(23):2462–71.

Rosental, Claude. 2013. "Toward a Sociology of Public Demonstrations." *Sociological Theory* 31(4):343–65.

Satel, Sally L. 2013. "Primed for Controversy." *New York Times*, February 24. Retrieved February 26, 2013 (http://www.nytimes.com/2013/02/24/opinion/sunday/psychology-research-control.html?_r=1and).

Schnall, S. 2014. "Social Media and the Crowd-Sourcing of Social Psychology." Retrieved January 5, 2015 (http://www.psychol.cam.ac.uk/cece/blog).

Schulz, Kenneth F., Douglas G. Altman, and David Moher for the CONSORT Group. "CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials." Retrieved July 23, 2018 (http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000251).

Shapin, Steven. 2010. *Never Pure: Historical Studies of Science as if It Was Produced by People with Bodies, Situated in Time, Space, Culture, and Society, and Struggling for Credibility and Authority*. Baltimore, MD: Johns Hopkins University Press.

Shapin, Steven. 2012. "The Sciences of Subjectivity." *Social Studies of Science* 42(2):170–84.

Shapin, Steven, and Simon Shaffer. 1985. *Leviathan and the Air Pump*. Princeton, NJ. Princeton University Press.

Simera, Iveta, David Moher, John Hoey, Kenneth F. Schulz, and Douglas G. Altman. 2010. "A Catalogue of Reporting Guidelines for Health Research." *European Journal of Clinical Investigation* 40(1):35–53.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22(11):1359–66.

Simonsohn, Uri. 2013. "Just Post It: The Lesson from Two Cases of Fabricated Data Detected by Statistics Alone." *Psychological Science* 24(10):1875–88.

Simonsohn, Uri, Leif D. Nelson, and Joseph P. Simmons. 2014. "P-curve: A Key to the File-drawer Problem." *Journal of Experimental Psychology: General* 143(2):534–47.

Singal, Jesse. 2016. "Inside Psychology's 'Methodological Terrorism' Debate." *New York Magazine*. Retrieved August 26, 2017 (http://nymag.com/scienceofus/2016/10/inside-psychologys-methodological-terrorism-debate.html).

Smith, Mary L., and Gene V. Glass. 1977. "Meta-analysis of Psychotherapy Outcome Studies." *American Psychologist* 32(9):752–60.
Smith-Doerr, Laurel, and Itai Vardi. 2015. "Mind the Gap: Formal Ethics Policies and Chemical Scientists' Everyday Practices in Academia and Industry." *Science, Technology, and Human Values* 40(2):176–98.
Srivastava, Sanjay. 2016. "Evaluating a New Critique of the Reproducibility Project." *The Hardest Science*. Retrieved March 17, 2016 (https://hardsci.wordpress.com/2016/03/03/evaluating-a-new-critique-of-the-reproducibility-project/).
Stanley, Tom D., and Hristos Doucouliagos. 2013. "Meta-regression Approximations to Reduce Publication Selection Bias." *Research Synthesis Methods* 5(1):60–78.
Stapel, Diederik. 2014. *Faking Science: A True Story of Academic Fraud*. Translated by Nicholas J. L. Brown. Retrieved July 24, 2018 (http://beinspired.no/wp-content/uploads/2016/03/FakingScience-20141214.pdf).
Stegenga, Jacob. 2011. "Is Meta-analysis the Platinum Standard of Evidence?" *Studies in History and Philosophy of Biological and Biomedical Sciences* 42(4):497–507.
Stokstad Erik, Elizabeth Pennisi, Jocelyn Kaiser, Jon Cohen, Jennifer Couzin-Frankel, Paul Voosen, Ann Gibbons, and Adrian Cho. "The Runners-up." *Science* 350(6267):1458–63.
Strang, David, and Kyle Siler. 2015. "Revising as Reframing: Original Submissions Versus Published Papers in Administrative Science Quarterly, 2005 to 2009." *Sociological Theory* 33(1):71–96.
Tabery, James. 2014. *Beyond Versus: The Struggle to Understand the Interaction of Nature and Nurture*. Cambridge, MA: The MIT Press.
Timmermans, Stefan, and Marc Berg. 2003. *The Gold Standard: The Challenge of Evidence-based Medicine and Standardization in Health Care*. Philadelphia, PA: Temple University Press.
Timmermans, Stefan, and Steven Epstein. 2010. "A World of Standardization but not a Standard World: Toward a Sociology of Standards and Standardization." *Annual Review of Sociology* 36:69–89.
van der Heijden, A. J., P. J. F. Groenen, and R. Zeelenberg. 2014. *Report of the Smeesters Follow-up Investigation Committee*. Rotterdam: Erasmus University.
Vazire, Simine. 2016. "Editorial." *Social Psychological and Personality Science* 7(1):3–7.
Wakeham, Joshua. 2017. "Bullshit as a Problem of Social Epistemology." *Sociological Theory* 35(1):15–38.
Will, Catherine M. 2009. "Identifying Effectiveness in 'The Old Old': Principles and Values in the Age of Clinical Trials." *Science, Technology, and Human Values* 34(5):607–28.
Woolgar, Steve. 1982. "Laboratory Studies: A Comment on the State of the Art." *Social Studies of Science* 12(4):481–98.
Woolston, Chris. 2015. "Psychology Journal Bans *P* Values." *Nature* 519(7541):9.
Yong, Ed. 2012a. "A Failed Replication Draws a Scathing Personal Attack from a Psychology Professor." *Discover Magazine*. Retrieved July 26, 2018 (http://blogs.discovermagazine.com/notrocket-science/2012/03/10/failed-replication-bargh-psychology-study-doyen/#.UVMv9JMQZvl.
Yong, Ed. 2012b. "Replication Studies: Bad Copy." *Nature* 485(7398):298–300.
Zimmerman, Ann S. 2008. "New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data." *Science, Technology and Human Values* 33(5):631–52.

AUTHOR BIOGRAPHIES

Jeremy Freese is a professor of sociology at Stanford University. He is co–principal investigator of the Time Sharing Experiments in Social Sciences program and the General Social Survey. His research includes work on social science genomics and health disparities. He is also coauthoring a book about methods for improving reproducibility and transparency in social science.

David Peterson recently graduated from Northwestern and will start as a postdoctoral fellow at the Institute for Society and Genetics at UCLA in the fall. His research focuses on how practice, evaluation, and ethics differ across scientific fields. He has previously published a comparative study looking at practices in molecular biology and psychology in the *American Sociological Review*, the rhetorical strategies scientists use to avoid theoretical confrontations in *Social Studies of Science*, and the struggle to maintain environmental control in laboratories studying infant cognition in *Socius*. He is currently researching the development of methodological activism across science.