Industrial and Organizational Psychology
PERSPECTIVES ON SCIENCE AND PRACTICE

Industrial and Organizational Psychology, **6** (2013), 252–268.

FOCAL ARTICLE

How Trustworthy Is the Scientific Literature in Industrial and Organizational Psychology?

SVEN KEPES AND MICHAEL A. McDANIEL
Virginia Commonwealth University

Abstract
The trustworthiness of research findings has been questioned in many domains of science. This article calls for a review of the trustworthiness of the scientific literature in industrial–organizational (I–O) psychology and a reconsideration of common practices that may harm the credibility of our literature. We note that most hypotheses in I–O psychology journals are confirmed. Thus, we are either approaching omniscience or our journals are publishing an unrepresentative sample of completed research. We view the latter explanation as more likely. We review structural problems in the publication process and in the conduct of research that is likely to promote a distortion of scientific knowledge. We then offer recommendations to make the I–O literature more accurate and trustworthy.

> "False facts are highly injurious to the progress of science for they often endure long."
> Charles Darwin (1981/1871, p. 385)

In recent years, trust in many scientific areas has come under scrutiny (Stroebe, Postmes, & Spears, 2012). The medical sciences have received particular attention. For example, some pharmaceutical companies appear to have withheld data to hide results that cast their products in a negative light (e.g., Berenson, 2005; Dickersin, 2005; Goldacre, 2012; Krimsky, 2006; Sarawitz, 2012; Saul, 2008; Whalen, Barrett, & Loftus, 2012). The situation is not much different in other scientific areas, ranging from the medical sciences, chemistry, and physics to the social sciences (e.g., Braxton, 1999; Gallup Organization, 2008; Goldstein, 2010; LaFollette, 1992; Park, 2001; Reich, 2010; Stroebe, Postmes, & Spears, 2012).

Psychology has not escaped concerns about scientific credibility. For instance, previously "established" effects in many areas of psychology such as the Mozart effect on spatial task performance (Rauscher, Shaw, & Ky, 1993), social priming (Bargh, Chen, & Burrows, 1996), psychic effects (Bem, 2011), verbal overshadowing (Schooler & Engstler-Schooler, 1990), and "practical" intelligence measures purported to yield a general factor that is substantially different from *g* (Sternberg et al., 2000) have been questioned or refuted (e.g., Doyen, Klein, Pichon, & Cleeremans, 2011; Francis, 2012; Lehrer, 2010; McDaniel & Whetzel, 2005; Ritchie, Wiseman, & French, 2012; Rouder & Morey, 2011; Yong, 2012a). Recently, substantial attention has been focused on allegations of scientific misconduct in

Correspondence concerning this article should be addressed to Sven Kepes.
E-mail: skepes@vcu.edu
Address: Virginia Commonwealth University, 301 West Main Street, PO Box 844000, Richmond, VA 23284
This focal article has benefited from feedback provided by several colleagues.

Stapel's social psychology research (see Stroebe et al., 2012, for a discussion of Stapel's multiple publications asserted to be based on fabricated or inappropriately manipulated data). As a result of these and other controversies, some have concluded that the reputation of science in general, including the scientific principles researchers purport to uphold, has suffered severe damage (e.g., Krimsky, 2006).

Most commonly, "established" effects are not totally discredited, but the magnitude of the effects (e.g., bilateral body symmetry and health; a test validity; mean racial differences in job performance) may be different, sometimes substantially, from the purported values (Banks, Kepes, & McDaniel, 2012; McDaniel, McKay, & Rothstein, 2006; Van Dongen & Gangestad, 2011). Such findings are not likely the result of fraud but may reflect suppression of some research findings and other nonoptimal practices of authors and editors or reviewers.

When we suggest that the magnitude of effects may be different from the purported values, we rely on evidence from publication bias analyses that estimates the extent to which the purported values are likely to be distorted because they are not representative of all research completed on a particular topic (e.g., Banks, Kepes, & Banks, 2012; Banks, Kepes, & McDaniel, 2012; Kepes, Banks, McDaniel, & Whetzel, 2012; Kepes, Banks, & Oh, in press; McDaniel, Rothstein, & Whetzel, 2006; Renkewitz, Fuchs, & Fiedler, 2011; Sutton, 2005). We refer to missing studies as being "suppressed," a term drawn from the publication bias literature (Kepes, Banks, McDaniel, & Whetzel, 2012; Rothstein, Sutton, & Borenstein, 2005b; Sutton, 2009). This suppression is not necessarily the result of a purposeful intent to hide results. Rothstein, Sutton, and Borenstein (2005a, p. 3) mention several potential data or results suppression mechanisms that can distort our cumulative knowledge. These mechanisms include the outcome bias (the selective reporting of some outcomes but not others in primary studies, depending on the direction and statistical significance of the results), the language bias (the selective inclusion of samples from studies published in the English language in meta-analytic reviews), the availability bias (the selective inclusion of samples from studies in a meta-analytic review that are readily available to a researcher), and the familiarity bias (the selective inclusion of samples from studies in meta-analytic reviews from the researcher's own scientific discipline). Thus, most forms of suppression are not likely to be a function of explicit intent to distort or hide research results.

We offer that it is time to consider the trustworthiness of research in industrial–organizational (I–O) psychology. We came to this position in two steps. First, we note that hypotheses in I–O journals are almost always supported, which is consistent with other disciplines of psychology (e.g., Sterling, 1959; Sterling & Rosenbaum, 1995; Yong, 2012a).[1] Although substantial support for hypotheses is a phenomenon across all sciences, the percentage of significant results reported in psychology journals is noticeably higher than that in most scientific disciplines (Fanelli, 2010b; Sterling, 1959; Sterling & Rosenbaum, 1995; Yong, 2012a). Furthermore, although there is an increase over time in the percentage of confirmed hypotheses, particularly in U.S. journals, the growth is strongest in psychology (Fanelli, 2012). Second, we considered reasons for why hypotheses are confirmed so consistently. We suggest that I–O researchers are either approaching omniscience or there are forces at work

1. Some might note that a rejected null hypothesis does not necessarily constitute a confirmed alternative hypothesis. In this article, we use phrases such as "confirmed hypothesis" and "supported hypothesis" to indicate that there was an inference that the alternative hypothesis was supported primarily based on a finding of statistical significance. In addition, some may assert that the reliance on statistical significance is waning. We disagree with this assertion. For example, Leung (2011) reported that all 45 quantitative articles published in the *Academy of Management Journal* in 2009 involved null hypothesis testing.

that cause I–O journal articles to be unrepresentative of all completed I–O research. We favor the latter explanation.

We acknowledge that the issues we raise about the I–O psychology literature are not unique to I–O psychology. Rather, the evidence that we offer is drawn from research in several scientific disciplines, and we assert that our concerns and recommendations are applicable to all disciplines of psychology. As well, our article has relevance to a broad range of scientific disciplines. Still, our primary interest is in the I–O literature and its scientific credibility, and we frame the article with respect to our literature.

Research-active I–O psychologists are drawn to the field, at least in part, because they wish to advance scientific knowledge. These scholars typically invest substantial time and effort in this pursuit. Similarly, editors and reviewers are well-accomplished researchers who devote considerable time to evaluating research and offering suggestions for improvement, typically on a voluntary basis. These scholars deserve substantial respect because of their efforts to advance scientific knowledge. Yet, we argue that many practices common to the scientific process, including the editorial review process, no matter how well intentioned and no matter how common, could be damaging to the accuracy of our cumulative scientific knowledge. We offer this article to highlight our concerns, to suggest a reexamination of common practices, and to encourage constructive debate. We seek to offer light rather than heat, but some commentary will necessarily invoke both.

In this article, we discuss the chase for statistical significance (Ferguson & Heene, 2012) that results in a strong tendency to obtain confirmed hypotheses. We then review structural issues in the publication process that may drive this chase for significance. We identify issues in the conduct of research that contribute to our specious omniscience. Finally, we offer recommendations for improving the trustworthiness of science in I–O psychology.

Structural Problems in the Scientific Process

For decades, we have known that studies with statistically significant results are more likely to get published in our journals than studies with nonsignificant results (Greenwald, 1975; Orlitzky, 2012; Porter, 1992; Rosenthal, 1979; Sterling, 1959). Articles with statistically significant, large magnitude effect sizes tend to dominate our journals (Greenwald, 1975; Orlitzky, 2012; Porter, 1992; Rosenthal, 1979; Sterling, 1959; Sterling & Rosenbaum, 1995). In a review of psychology journals, Sterling (1959) found that 97% of the articles rejected the null hypothesis. In a replication more than 36 years later, Sterling and Rosenbaum (1995) reported essentially identical results. We offer these findings as evidence supporting the inference of structural problems in the psychology scientific process that should also affect I–O psychology.

Academic journals are in a competitive market (Jefferson, 1998). They seek to publish ''hot new findings'' (Witten & Tibshirani, 2012) because articles with such findings get cited more often than other articles, and these citations enhance the reputation of a journal as judged by impact factors and related indices. In many scientific areas, the most heavily cited articles are concentrated in a small number of journals (Ioannidis, 2006). Relative to articles with unsupported hypotheses, articles with supported hypotheses tend to be judged more interesting, capture greater attention, and boost the reputation of a journal. This is likely to be particularly true for prestigious journals (Murtaugh, 2002). Therefore, the reward structure for academic journals may inadvertently encourage the acceptance of articles with supported hypotheses. Relatedly, the reward structure could also encourage the rejection of studies that do not find support for hypotheses.

Academic departments in I–O psychology are also in a competitive market. The highest ranked programs are motivated to stay highly ranked. The lesser ranked programs are motivated to improve their ranking. Thus, programs tend to encourage their faculty and graduate students to publish in high impact journals (Giner-Sorolla, 2012). Some faculty and academic programs may even hold the norm that only publications in elite journals are worthwhile, and researchers are encouraged to abandon research studies that have been rejected from such journals. Even in the absence of such norms, faculty publications and the quality of the journals are emphasized in faculty employment actions, including tenure, promotion, salary increases, and discretionary funding (e.g., Gomez-Mejia & Balkin, 1992). I–O psychology graduate students pursuing academic careers typically seek to find a job in the best-ranked academic program that they can. They recognize that their academic placement is substantially driven by their publications. Thus, academic researchers seek to publish in the most elite journals and recognize that the probability of such publications relies in part on having hypotheses that are supported.

Given the scientist–practitioner emphasis of I–O psychology, many I–O psychologists employed in consulting, industry, and government also seek to publish. The consultant's research often relates to their organization's products or services. Industry and governmental researchers often wish to document the efficacy of an organization's I–O practices (e.g., selection systems or team practices). These I–O psychologists strive for publications with supported hypotheses in part to serve the commercial or reputational interests of their organizations (e.g., ''our products and services work well;'' ''our team-based workforce is effective'').

Because of the reward structure (Koole & Lakens, 2012; Nosek, Spies, & Motyl, 2012), researchers are motivated to ''chase the significant'' (Ferguson & Heene, 2012, p. 558) in order to find support for hypotheses (Fanelli, 2010a; Jasny, Chin, Chong, & Vignieri, 2011; Koole & Lakens, 2012; Nosek et al., 2012; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).[2] This chase of the significant can lead to a high rate of false-positive published findings and thus the distortion of science (Ioannidis, 2005, 2012; Sarawitz, 2012; Simmons, Nelson, & Simonsohn, 2011). False-positive results are very ''sticky'' (i.e., they rarely get disconfirmed) because null results have many possible causes. In addition, failures to replicate false positives are seldom conclusive (Ferguson & Heene, 2012; Simmons et al., 2011). Note that we are not arguing that our literature is entirely composed of zero-magnitude population effect sizes that are falsely presented as nonzero. Rather, we are arguing that it is common for the magnitude of our population effect sizes to be misestimated, often overestimated (Kepes, Banks, McDaniel, & Whetzel, 2012). Thus, although authors and journals seek to improve science, our actions in the ''chase for significance'' may actually damage it. It seems as if the academic community could be ''rewarding A, while hoping for B'' (Kerr, 1975, p. 769).

Issues in the Conduct of Research and the Editorial Process That Contribute to our Specious Omniscience

Researchers have substantial methodological flexibility (Bedeian, Taylor, & Miller, 2010; Simmons et al., 2011; Stroebe et al., 2012) that can be marshaled to obtain supported hypotheses. In part because our field tends to develop several measures for the same construct, researchers have the flexibility to use multiple operationalizations of constructs in order to identify the measurement approaches that best support their hypotheses (e.g., ''throw lots of variables against the wall and talk about the variables that stick''). Researchers can stop data collection, tweak the design or

2. In the next section of this article (Issues in the Conduct of Research and the Editorial Process That Contribute to our Specious Omniscience), we describe specific actions authors can take to facilitate the finding of statistical significance.

measurement, and discard the original data and collect new data. Researchers can drop outliers or other observations that diminish the magnitude of obtained effects. They can collect additional data needed to increase sample size to move a marginally statistically significant effect size (e.g., $p < .06$) to a significance level that is more acceptable to journals (e.g., $p < .05$).[3] Researchers can also terminate data collection once the preferred p value is obtained. They can alter the analysis method to identify the analysis approach that best supports the hypothesis. They can also abandon the hypotheses that were not supported. Researchers can then report the results that best support the retained hypotheses in a nice, neat publishable package and never mention the discarded variables, analyses, observations, and hypotheses.

If one is to trust Bedeian et al.'s (2010) survey of management faculty[4] that examined knowledge of methodological flexibility in their colleagues during the previous year, instances of methodological flexibility are not rare events. For example, 60% of faculty knew of a colleague who ''dropped observations or data points from analyses based on a gut feeling that they were inaccurate.'' Fifty percent of faculty knew of a colleague who ''withheld data that contradicted their previous research.'' Other questionable practices are also fairly common (see Bedeian et al., 2010).

If methodological flexibility does not yield the desired support of a hypothesis, researchers can simply change the hypotheses to match the results (HARKing: hypothesizing after the results are known; Kerr, 1998). The Bedeian et al. (2010) survey reported that approximately 92% of faculties know at least one colleague who has engaged in HARKing in the last year. Thus, HARKing is likely to be a common practice in I–O psychology.

But it is not only researchers who engage in practices such as HARKing. Reviewers and editors sometimes encourage these practices as part of well-intentioned efforts to improve submitted manuscripts (Leung, 2011; Rupp, 2011). In essence, editors and reviewers function as gatekeepers (Crane, 1967; Rupp, 2011). Thus, it is not unheard of for editors and reviewers to suggest hypotheses or ask authors to drop (or add) analyses, results, or hypotheses to push the editor or reviewer's perspective or research agenda (Leung, 2011; Rupp, 2011), or to simply make the paper more ''interesting.'' In the most extreme case, if HARKing cannot yield a supported hypothesis, authors can fabricate their data to be consistent with the hypotheses (Carey, 2011; Crocker & Cooper, 2011; LaFollette, 1992; Stroebe et al., 2012; Vogel, 2011; Yong, 2012b).

In summary, there is a growing body of evidence indicating that I–O-related research is affected by publication bias (e.g., Ferguson & Brannick, 2012; Kepes, Banks, McDaniel, & Whetzel, 2012; Rothstein, 2012; Rothstein et al., 2005a). That is, our published or readily accessible research is not representative of all completed I–O research on a particular relation of interest. Thus, the I–O research literature may likely contain an uncomfortably high rate of false-positive results (i.e., the incorrect rejection of a null hypothesis) and other misestimated effect sizes. This can have a long-lasting distorting effect on our scientific knowledge, particularly because *exact* replications tend to be discouraged by our journals (Makel, Plucker, & Hegarty, 2012; Neuliep & Crandall, 1990, 1993; Simmons et al., 2011; Yong, 2012a). Because of the lack of such replication studies, some scholars believe that our sciences have lost the ability to self-correct and generate accurate cumulative knowledge (Giner-Sorolla, 2012; Ioannidis,

3. We are not arguing against increasing one's sample size. We also note that increasing one's sample size would not, on average, cause the resulting effect size to be overestimated. We do argue that increasing one's sample size solely to move a marginally statistically significant effect size into a state of statistical significance is an example of methodological flexibility (Simmons et al., 2011).
4. We note that many I–O psychologists are employed in management departments due to the generally higher salaries than can be obtained in psychology departments.

2005, 2012; Koole & Lakens, 2012; Nosek et al., 2012). The lack of exact replication studies prevents the opportunity to disconfirm research results and thus to falsify theories (Kerr, 1998; Leung, 2011; Popper, 1959). This has led to a ''vast graveyard of undead theories'' (Ferguson & Heene, 2012, p. 555). We note that we have many theories, and only a small number have been refuted through replication.

Furthermore, even if replication studies are conducted, it may take a considerable amount of time before such studies become publically available (a time lag bias; Banks, Kepes, & McDaniel, 2012; Trikalinos & Ioannidis, 2005). This situation has caused Lehrer (2010) to question the scientific methods in the social sciences (see also Giner-Sorolla, 2012; Ioannidis, 2005, 2012; Koole & Lakens, 2012; Nosek et al., 2012; Stroebe et al., 2012; Yong, 2012a). In Lehrer's (2010, p. 52) view, ''the truth wears off'' because facts are ''losing their truth: claims that have been enshrined . . . are suddenly unprovable.'' Relatedly, Ioannidis (2012) asserted that most scientific disciplines may be producing wrong or distorted information on a massive scale (see also Ioannidis, 2005).

Results from multiple studies have documented likely suppression bias (e.g., declining to report effect sizes in a published study or not publishing a study in its entirety). In the most comprehensive review of publication bias in psychology, Ferguson and Brannick (2012) reported that publication bias was present in approximately 40% of meta-analyses and that the degree of this bias was worrisome in about 25% of meta-analyses. This bias has resulted in misestimating the magnitude of population effects in several I–O research domains. For example, results were found consistent with the inference that the validities of some commercially available employment tests are overestimated, sometimes substantially (McDaniel, Rothstein, & Whetzel, 2006). In addition, Renkewitz et al. (2011) reported that publication bias exists in the literature on judgment and decision making. Furthermore, Banks, Kepes, and McDaniel (2012) found evidence consistent with an inference of publication bias in the literature on conditional reasoning tests. Similarly, Kepes, Banks, and Oh (in press) analyzed four datasets from previously published I–O meta-analyses and reported that three of these datasets (work experience and performance, gender differences in transformational leadership, and Pygmalion interventions) are likely to have been affected by publication bias. Together, these and other studies suggest that at least some literature areas in our field have a bias against the publishing of null, contrarian (i.e., effect sizes that are counter to already published findings), and small magnitude effect sizes.[5]

Having offered evidence that concerns are likely warranted about the trustworthiness of the I–O psychology research literature, we now offer recommendations on improving the credibility and accuracy of the I–O psychology research literature.

Recommendations

Create research registries. We offer that one of the most effective ways to overcome trustworthiness concerns in I–O psychology is the creation and mandatory use of research registries (Ferguson & Brannick, 2012; Kepes, Banks, McDaniel, & Whetzel, 2012). A research registry is a database in which researchers register studies that they plan to conduct (Banks & McDaniel, 2011; Berlin & Ghersi, 2005). Such registries exist across scientific disciplines, particularly in the medical sciences (Berlin & Ghersi, 2005). When study registration occurs prior

5. We acknowledge that Dalton, Aguinis, Dalton, Bosco, and Pierce (2012) concluded that publication bias is not worrisome in our literature. However, that paper stands alone in that conclusion. We note that the Dalton et al. (2012) effort differed from all other published publication bias analyses in that did not examine any specific research topic in the literature (Kepes, McDaniel, Brannick, & Banks, 2013). As such, we do not find it an informative contribution to the publication bias literature (i.e., publication bias is concerned with the availability of effect sizes on a particular relation of interest).

to data collection and, thus, before the results are known, the possibility that results are being fabricated or suppressed is minimized (Chalmers et al., 1987; Dickersin, 1990; Easterbrook, Gopalan, Berlin, & Matthews, 1991). In addition, when conducting a review on a particular relation of interest, researchers can search the registry for information on unpublished samples, which can then be included in the review (Kepes et al., 2012; Rothstein, 2012). Thus, research registries would promote greater comprehensibility and credibility in the reviews and summaries of our literatures.

Although many registries exist in other scientific areas, none exist in I–O psychology and related disciplines in the organizational sciences (Kepes, Banks, McDaniel, & Whetzel, 2012).[6] Registries in the medical and related sciences are typically linked to a particular organization. For example, the ClinicalTrials.gov registry is provided by the U.S. National Institutes of Health. Another registry is available through the Cochrane Library (Dickersin et al., 2002), and even the U.S. Food and Drug Administration operates a registry (Turner, 2004). To encourage the use of registries, the *International Committee of Medical Journal Editors*, in 2005, required registration before a clinical trial study was conducted as a condition for publication in the participating journals. This has led to studies being routinely registered (Laine et al., 2007). We recommend that journals in I–O psychology and related disciplines require registration of studies prior to the collection of data. Leading associations and organizations within our field, such as the American Psychological Association, the Society for Industrial Organizational Psychology, or the Academy of Management could take a leading role in the establishment of registries (Kepes, Banks, McDaniel, & Whetzel, 2012). Alternatively, individual journals or consortiums of journals could establish and maintain registries.[7] In establishing registries, the Open Science Framework could assist (Open Science Collaboration, 2012). Universities might also require study registration as many do now for institutional review board (IRB) approval when the registry is to contain data from individuals. Regardless of whether individual data are to be released into a registry, IRB policies could require registration of a planned study.

If the goal of registering a study prior to data collection is to minimize problems associated with HARKing and methodological flexibility, one can identify a preliminary list of data elements for a given planned study to be included in a registry. These data elements would include hypotheses, all variables and their operational definitions, a power analysis, a statement of the anticipated sample size, and a description of the planned analyses. One would also need data elements to identify the topic areas associated with the research as well as contact information for the individual(s) who registered the planned study.

Changes in editorial review process. The current editorial review system is likely part of the reason for the bias against the publishing of null, contrarian, and small magnitude effect sizes. Journal reviewers and editors may inadvertently promote HARKing and

6. We note, however, the existence of some attempts to create research registries. For instance, PsycFileDrawer (http://psychfiledrawer.org) is an archive of replication attempts in experimental psychology. The Campbell Collaboration (http://www.campbellcollaboration.org) covers the fields of education, crime and justice, and social welfare, and is modeled after the very successful and prestigious Cochrane Collaboration in the medical sciences. One of the most interesting attempts is the Open Science Framework (http://openscienceframework.org), currently in its beta version, which allows for the documentation and archiving of studies (Open Science Collaboration, 2012).

7. We note that, due to the interrelatedness of I–O psychology with disciplines such as management and social psychology, registration requirements for only I–O psychology journals could have adverse effects on I–O researchers and the field of I–O psychology. Thus, we recommend that journal editors from the leading journals in I–O psychology and related scientific disciplines coordinate their efforts and act in concert (similar to the effort in the medical sciences with the International Committee of Medical Journal Editors).

other questionable practices (Leung, 2011; Rupp, 2011), such as requesting authors to change hypotheses or to delete hypotheses that are not supported (e.g., ''we recommend that you revisit Hypothesis 1 by considering the research related to xxx;'' ''The lack of support for Hypothesis 3 harms the flow of the paper and we recommend its deletion, particularly because we seek a reduction in manuscript length'').[8] Reviewers and editors are able to engage in these nonoptimal behaviors because they see a study's result when reviewing a manuscript. Although well intentioned, such practices can lead to the suppression of research results.

To prevent or reduce the frequency of such behaviors, we recommend the implementation of a two-step review process (Liberati, 1992). In the first stage, authors would submit only part of their manuscript for review, specifically the introduction, method section, and the analysis approach. During this stage of the review process, editors and reviewers can make decisions regarding soundness of the study (e.g., theory development, study design, measurement, and analysis) and the likely contribution to the literature. Manuscripts that pass this stage would advance to the second stage in which the complete manuscript, including the results and discussion, would be provided. The editor and reviewers could then assess whether the results and conclusion sections are aligned with the introduction, theory, and method sections (i.e., verify that the authors actually did what they proposed during the initial submission). Such a two-stage review process could minimize post hoc rationalizing during the editorial review process. In addition, authors could have less motivation to engage in HARKing and related inappropriate practices because the results and degree of hypothesis confirmation would have less impact on the publication decision.

As part of the second stage of the review process, we also recommend that authors be required to submit the raw data and relevant documentation, including syntax and a summary of all measures. This practice benefits research in three ways. First, it gives reviewers and the editor the opportunity to check the data and syntax for potential mistakes, a practice that currently is almost never done (Schminke, 2009; Schminke & Ambrose, 2011). In addition, the research protocol and the list of available variables could be reviewed. If particular variables are left out of the analysis, reviewers could ask for their inclusion or for a clear statement concerning why they were excluded.

Second, this practice would ensure that the data are securely stored so that, at some future time (e.g., 5 years after the publication date), the journal may publically release the data. Some would argue that journals have the duty to assist the scientific community by permitting the reproduction of published results, an important aspect of the scientific process (Baggerly, 2010). If the raw data and syntax are made available, other researchers could also use it for secondary analyses, thereby likely advancing our scientific knowledge. Unfortunately, psychologists are very reluctant to share their data (Wicherts, Borsboom, Kats, & Molenaar, 2006; Wolins, 1962).[9] We note that the data from these cited studies are not specific to I–O psychology. Although this is anecdotal evidence, the authors of this article have often requested I–O related datasets and many times our requests were refused. Thus, we do not believe that I–O psychologists tend to be more generous with their data than psychologists in general.

Furthermore, in their reanalysis of articles published in two major psychology journals, Wicherts, Bakker, and Molenaar

8. The quotes are offered as illustrative of comments editors have made to the authors or to colleagues of the authors. They are not direct quotes from editorial reviews.

9. The *American Psychological Association* (2002, p. 1071) requires the sharing of data with ''other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose.'' We assert that this is a very restrictive policy that does not actively promote data sharing.

(2011) found that the reluctance to share data is associated with weak statistical evidence against the null hypothesis and errors in the reporting of statistical results. Finally, evidence from other scientific disciplines indicates that the reproducibility of findings from published articles is shockingly low (e.g., Begley & Ellis, 2012; Ioannidis et al., 2009), likely due to incomplete reporting of relevant statistical information in the published articles (Baggerly, 2010). We suspect that the situation is not much different in I–O psychology. According to editors of the journal *Science*, who promote the public release of scientific data, ''we must all accept that science is data and that data are science'' (Hanson, Sugden, & Alberts, 2011, p. 649). The sharing of data is thus important for the advancement and credibility of our literature.

Third, authors would be aware that their data are subject to audit immediately by the editor and reviewers and possibly later by other researchers. This should cause authors to be more careful in their analysis and may reduce the number of inappropriate research practices. In addition, because statistical analyses are prone to error (Murphy, 2004; Strasak, Zaman, Marinell, Pfeiffer, & Ulmer 2007; Wolins, 1962), reanalyses can ensure the accuracy of our scientific knowledge. Otherwise, our cumulative knowledge can be adversely affected. We recognize the increased labor demands on the editors and reviewers that a data audit would entail. However, with access to raw data and related documentation, editors and reviewers could use guidelines (Simmons et al., 2011) or checklists (Nosek et al., 2012) and inspect the raw data if they feel that it is warranted. With access to the raw data, reviewers may also become more tolerant toward ''imperfect'' results because they can actually check them.

We encourage the release of data for meta-analyses in addition to data from primary studies. We note that past contentious debates concerning differing meta-analytic results in personality test validity (debates in the 1990s) and, more recently, in integrity test validity, would have been more clearly and reasonably resolved if the authors had released their data and related documentation. Note that we do not wish to underestimate the controversies associated with a requirement to release data.

In addition, we recommend that all journals that publish I–O psychology research accept and encourage supplemental information to be submitted. Although several medical journals and journals in many other scientific disciplines permit the submission of supplemental material that is stored indefinitely on the journal websites, I–O journals tend not to do this. Supplemental information can contain material that does not fit in a journal article given the page constraints of the journal. If null effects or unsupported hypotheses were removed from a paper, the deleted results and analyses could be placed in the supplemental materials. If the web has room for a million cat videos, it certainly has room for supplemental information for our research studies.

Encouragement of null effect publications and replications. The suppression of small or null effect sizes is likely the primary cause of research distortion due to publication bias. Although there have been a few attempts by some journals to address this effect size suppression (e.g., the *Journal of Business and Psychology* has a forthcoming special issue on the topic of ''Nothing, zilch, nil: Advancing organizational science one null result at a time''), one special issue is insufficient to curtail data suppression. Instead, all of our journals, particularly our most prestigious journals, should actively promote the publication of scientific findings regardless of the magnitude of the effect sizes.[10] We recommend that journals reserve space in each issue for the publication of null results. We note that since 2000, *Cancer Epidemiology, Biomarkers & Prevention*, one of the

10. We are not arguing that journals accept all submitted papers. We are arguing that the magnitude of effect sizes should not be a criterion in acceptance decisions.

top oncology and public health journals, contains a section on null results (Shields, 2000; Shields, Sellers, & Rebbeck, 2009).

As mentioned previously, in the social sciences, including I–O psychology, there is a severe lack of exact replications (e.g., Makel et al., 2012; Pashler & Harris, 2012; Yong, 2012a). For the entire field of psychology, Makel et al. (2012) estimated that between 1900 and today only around 1% of all published articles in psychology journals are replication studies. This is unfortunate because a scientifically ''true'' effect is one ''which can be regularly reproduced by anyone who carries out the appropriate experiment in the way prescribed'' (Popper, 1959, p. 23). Thus, *exact* replications are necessary to determine whether an observed effect is ''true.'' We therefore recommend the publication of exact replications in our I–O journals, including our elite journals, regardless of the results. Replications are considered by many to be the ''scientific gold standard'' (Jasny et al., 2011, p. 1225) because they are essential for the ability of a scientific field to self-correct, which is one of the hallmarks of science (Merton, 1942, 1973). As noted by, Schmidt (2009, p. 90):

> Replication is one of the central issues in any empirical science. To confirm results or hypotheses by a repetition procedure is at the basis of any scientific conception. A replication experiment to demonstrate that the same findings can be obtained in any other place by any other researcher is conceived as an operationalization of objectivity. It is the proof that the experiment reflects knowledge that can be separated from the specific circumstances (such as time, place, or persons) under which it was gained.

Studies in I–O journals typically follow the confirmatory or verification strategy, which can impair theoretical progress (Leung, 2011; Popper, 1959; Uchino, Thoman, & Byerly, 2010). This strategy seeks to provide confirmatory rather than disconfirmatory research evidence for researcher's beliefs (i.e., confirmation bias; Nickerson, 1998). The publication of exact replication studies regardless of their results could uncover the potential prevalence of the confirmation bias and the publication of false-positive or otherwise erroneous results (Yong, 2012a) as well as assist the generation of accurate cumulative knowledge (Schwab, 2005).

We thus recommend the publication of exact replication studies regardless of the results. Some journals, such as *Personnel Psychology*, already have a special section for book reviews. These contributions are relatively short. We suggest that exact replication studies would not need more space because the rationale for the study and most details of the method and analysis appear in the original article that is the subject of the replication.

Strengthen the methods-related belief system. One can divide our scientific knowledge into theory-relevant and method-relevant beliefs (LeBel & Peters, 2011). Although the distinction is not sharp, theory-relevant beliefs concern the how and why (Sutton & Staw, 1995); that is, the theoretical mechanisms that cause behaviors and other outcomes. By contrast, method-relevant beliefs concern the procedures and processes by which data are measured, collected, and analyzed (LeBel & Peters, 2011). The centrality of these beliefs affects how obtained results are interpreted. Because psychology tends to be driven by a theory-relevant belief system, much more than a method-relevant belief system, researchers tend to interpret confirmatory results as theory relevant and disconfirmatory results as method relevant, ''with the result that the researcher's hypothesis is artificially buffered from falsification'' (LeBel & Peters, 2011, p. 372).[11]

11. We suggest that I–O psychology is more theory-oriented than methods-oriented based on research and assertions in the psychology and management literatures (e.g., Hambrick, 2007; LeBel & Peters, 2011; Leung, 2011).

Our elite I–O journals emphasize the importance of ''theoretical contributions,'' and papers are often rejected because they do not make a strong enough theoretical contribution. Consider the *Journal of Applied Psychology (JAP)*. In an editorial on the journal's content and review process 24 years ago, Schmitt (1989, p. 844) mentioned the word ''theory'' one time, and in a context that is unrelated to theory development (to account for the term ''theoretical'' and related words, the search string ''theor*'' yielded seven results):

> We will publish replications, sometimes as short notes. In certain situations, there should be replication, such as a replication of a counterintuitive result or of a result that is inconsistent with theory or past research. [12]

By contrast, the most recent editorial by Kozlowski (2009) on the mission and scope as well as the review process at *JAP* mentions the word ''theory'' 36 times (25 times excluding the references; ''theor*'' was detected 54 times), often in contexts that relate to theory development, theory extension, and so on. The increase in theory-related language has thus increased over the years. The situation in other journals is similar. For instance, all articles published in the *Academy of Management Journal* ''must also make **strong theoretical contributions**'' (the phrase ''strong theoretical contribution'' is bold in the original; *Academy of Management Journal*, 2012; see Leung, 2011).

As Leung (2011, p. 471) phrased it, ''the zeitgeist of top-notch . . . journals mandates a huge premium on theoretical contributions in determining what papers will see the light within their pages.'' Hambrick (2007, p. 1346) noted that the ''blanket insistence on theory, or the requirement of an articulation of theory in everything we write, actually retards our ability to achieve our end: understanding.'' Still, within the last 6 years since Hambrick's remarks, the ''theory fetish'' (2007, p. 1346) may have become stronger instead of weaker in our elite journals. Researchers follow the requirements of these journals, resulting in an emphasis on theory development at the potential expense of methodological rigor and the unfortunate introduction of HARKing and other inappropriate practices to fit the data to a theory (Bedeian et al., 2010; Ferguson & Heene, 2012; Hambrick, 2007; LeBel & Peters, 2011; Leung, 2011; Rupp, 2011; Simmons et al., 2011).

Although we agree with Lewin (1952, p. 169) that ''there is nothing more practical than a good theory,'' researchers have developed much theory in the past decades without properly evaluating the theories (Ferguson & Heene, 2012). HARKing and related inappropriate practices could have been encouraged by the strong emphasis on theory. We recommend that our journals lessen their ''theory fetish'' (Hambrick, 2007, p. 1346) and free up the journal space that this stance requires. We suggest that this journal space could be more appropriately used for replication studies of already promulgated theories.

Consistent with a greater emphasis on a method-focused orientation, we recommend that all submitted manuscripts include a section that assesses the robustness of the obtained results. For example, if the results section contains analyses with various covariates and without outliers, this separate section should contain the results with outliers as well as without covariates. The reporting of these supplemental results would allow a detailed assessment of the robustness of the ''main'' results. Similarly, results using different operationalizations of particular variables should be included in this section. For instance, dispersion, often used in unit-level research to assess constructs such as team diversity, climate strength, or pay variation, can be operationalized in multiple ways (Allison, 1978;

12. Interestingly, Neal Schmitt's (1989) editorial explicitly mentions that the *Journal of Applied Psychology* will publish replications. More recent editorials have not always made this explicit, highlighting the predisposition against replications.

Harrison & Klein, 2007), and the chosen operationalization can affect the obtained results (Roberson, Sturman, & Simons, 2007). However, our journals rarely report multiple operationalizations of a particular construct and the potentially differing results.

Relatedly, different scales for the same construct may yield differing results (e.g., personality constructs; Pace & Brannick, 2010). Thus, the section on the robustness of the results should contain the obtained results when using alternative operationalizations, scales, or constructs. Sensitivity analyses may also consider different analysis strategies and approaches. Thus, one would not want to offer a conclusion that is not replicable using other appropriate analysis methods. The reporting of a rigorous power analysis should also become more common in our published journal articles (Simmons et al., 2011). These practices could greatly reduce concerns related to the methodological flexibility problem (Simmons et al., 2011). Although there may not be anything more practical than a good theory (Lewin, 1952), the use of sound methods as well as the development of new methods and statistical techniques could be even more important (Greenwald, 2012), particularly if we are interested in assessing the robustness of our results and improving the accuracy and trustworthiness of our literature. To the extent that sensitivity analyses make an article excessively long, the key findings of such analyses could be mentioned in the article and the details placed in the supplemental materials, just as they are in other scientific disciplines (Evangelou, Trikalinos, & Ioannidis, 2005).

The aforementioned recommendations pertain largely to primary studies. However, analogous recommendations apply to review articles, particularly meta-analytic (systematic) reviews (Kepes et al., 2013). Most meta-analytic reviews fail to assess the robustness of the obtained results, whether this assessment involves outliers, publication bias, or related phenomena (Aguinis, Dalton, Bosco, Pierce, & Dalton 2011; Aytug, Rothstein, Zhou, & Kern, 2012; Banks, Kepes, & McDaniel, 2012; Kepes, Banks, McDaniel, & Whetzel, 2012). In concurrence with previous calls for more rigorous assessments of meta-analytic results (Banks, Kepes, & McDaniel, 2012; Ferguson & Heene, 2012; Kepes, Banks, McDaniel, & Whetzel, 2012; Kepes, Banks, & Oh, in press; Kepes et al., 2013; McDaniel, McKay, & Rothstein, 2006) and the Meta-analytic Reporting Standards (MARS) of the *American Psychological Association* (2008, 2010), we recommend that journals require such analyses in all meta-analytic reviews.

Implementation Considerations

We note that there are many challenges facing those seeking to adopt the recommendations presented in this article. Some will argue that we have overstated the problems and their opinions will need to be addressed. Others will agree that change would benefit our discipline but have low expectations that change will actually happen. Many of our recommendations, when implemented, are likely to increase the labor demands during the editorial review process. Change is hard but is sorely needed.

We offer that better practices can best be obtained through the coordinated action of multiple journals.[13] For example, change occurred in medical research when a committee of journal editors required registration before a clinical trial study was conducted as a condition for publication in the participating journals (Laine et al., 2007). As a result, studies began to be routinely registered. In addition, most of the commercial publishers of our journals (i.e., publishers not owned by scientific organizations) have experience in maintaining supplemental information

13. We acknowledge http://editorethics.uncc.edu as a step in this direction. We also refer readers to http://psychdisclosure.org/, a website which, on a volunteer basis, provides important methodological details about recently published articles that are not included in the journal publication.

because they also publish journals in fields in which supplemental information is a common practice. We note that APA journals can accept supplemental information, although the practice does not appear to be widely used by the *Journal of Applied Psychology* or other I–O related journals. Although we argue that changes are best initiated at the journal level, we are encouraged by the survey findings of Fuchs, Jenny, and Fiedler (2012), which indicate that research psychologists are generally supportive of changes in the scientific process. We do not seek to underestimate the challenges faced, but we note that some disciplines have made substantial progress in rectifying some of the structural problems in the scientific process.

Although we have made a case for the improvement of science in I–O psychology through the implementation of our recommendations, there are clear costs. The work load of journal editors and the editorial boards would increase due to the need for new policies and procedures. For instance, the article review process may entail more labor. In addition, there is also substantial labor involved in starting and maintaining a registry. Moreover, there are costs involved in implementing changes to the way we conduct research. If we stop HARKing and manipulating our data and research designs to enhance the likelihood of publication, it may become more difficult to publish. We do not wish to underestimate any of these costs. However, the question becomes whether we wish to have a trustworthy science that entails some additional costs, or should we continue with our problematic scientific practices?

Conclusion

We offer that a review of the trustworthiness of the scientific literature in I–O psychology is warranted. Concern was expressed over the fact that most hypotheses in our journals are confirmed. We offered that our journals are likely to be publishing an unrepresentative sample of completed research in certain literature areas and that this is damaging our scientific knowledge in that population effects are misestimated, typically overestimated. We reviewed structural problems in the publication process as well as in the conduct of research that could be promoting distortions of our scientific knowledge. We provided recommendations to make the I–O literature more trustworthy and accurate.

Just as we doubt that the researchers who publish in I–O journals are omniscient, we offer that we do not have all the answers to resolving issues in the scientific process of I–O psychology. However, we are quite sure that researcher practices and our editorial review process need to be altered in order to build better science and enhance the trustworthiness of our literature. Although we respect our researcher colleagues, reviewers, and journal editors, and recognize that they seek to promote good science, we offer that some of well-intentioned policies, practices, and behaviors that are common during the research and editorial review processes could be damaging to the accuracy of our scientific knowledge.

References

Academy of Management Journal. (2012). *Information for Contributors*. Retrieved from http://aom.org/Publications/AMJ/Information-for-Contributors.aspx

Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2011). Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. *Journal of Management*, *37*, 5–38. doi: 10.1177/0149206310377113

Allison, P. D. (1978). Measures of inequality. *American Sociological Review*, *43*, 865–880. doi: 10.2307/2094626

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, *57*, 1060–1073. doi: 10.1037/0003-066x.57.12.1060

American Psychological Association. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, *63*, 839–851. doi: 810.1037/0003-1066X.1063.1039.1839

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

Aytug, Z. G., Rothstein, H. R., Zhou, W., & Kern, M. C. (2012). Revealed or concealed? Transparency of procedures, decisions, and judgment calls in meta-analyses. *Organizational Research Methods*, *15*, 103–133. doi: 10.1177/1094428111403495

Baggerly, K. (2010). Disclose all data in publications. *Nature*, *467*, 401. doi: 10.1038/467401b

Banks, G. C., Kepes, S., & Banks, K. P. (2012). Publication bias: The antagonist of meta-analytic reviews and effective policy making. *Educational Evaluation and Policy Analysis*, *34*, 259–277. doi: 10.3102/0162373712446144

Banks, G. C., Kepes, S., & McDaniel, M. A. (2012). Publication bias: A call for improved meta-analytic practice in the organizational sciences. *International Journal of Selection and Assessment*, *20*, 182–196. doi: 10.1111/j.1468-2389.2012.00591.x

Banks, G. C., & McDaniel, M. A. (2011). The kryptonite of evidence-based I–O psychology. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *4*, 40–44. doi: 10.1111/j.1754-9434.2010.01292.x

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230–244. doi: 10.1037/0022-3514.71.2.230

Bedeian, A. G., Taylor, S. G., & Miller, A. N. (2010). Management science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning & Education*, *9*, 715–725. doi: 10.5465/amle.2010.56659889

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*, 531–533. doi: 10.1038/483531a

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407–425. doi: 10.1037/a0021524

Berenson, A. (2005, May 31). Despite vow, drug makers still withhold data. *New York Times*, p. A1. Retrieved from http://www.nytimes.com/2005/05/31/business/31trials.html.

Berlin, J. A., & Ghersi, D. (2005). Preventing publication bias: Registries and prospective meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta analysis: Prevention, assessment, and adjustments*. West Sussex, England: Wiley.

Braxton, J. M. (1999). *Perspectives on scholarly misconduct in the sciences*. Columbus, OH: Ohio State University Press.

Carey, B. (2011, November 3). Fraud case seen as a red flag for psychology research. *New York Times*, p. 3. Retrieved from http://www.nytimes.com/2011/11/03/health/research/noted-dutch-psychologist-stapel-accused-of-research-fraud.html

Chalmers, T. C., Levin, H., Sacks, H. S., Reitman, D., Berrier, J., & Nagalingam, R. (1987). Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Statistics in Medicine*, *6*, 315–325. doi: 10.1002/sim.4780060320

Crane, D. (1967). The gatekeepers of science: Some factors affecting the selection of articles for scientific journals. *The American Sociologist*, *2*, 195–201. doi: 10.2307/27701277

Crocker, J., & Cooper, M. L. (2011). Addressing scientific fraud. *Science*, *334*, 1182. doi: 10.1126/science.1216775

Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in a meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology*, *65*, 221–249. doi: 10.1111/j.1744-6570.2012.01243.x

Darwin, C. (1981/1871). *The descent of man and selection in relation to sex*. London, England: Princeton University Press.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association*, *263*, 1385–1389. doi: 10.1001/jama.1990.03440100097014

Dickersin, K. (2005). Publication bias: Recognizing the problem, understandings its origins and scope, and preventing harm. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 11–34). West Sussex, England: Wiley.

Dickersin, K., Manheimer, E., Wieland, S., Robinson, K. A., Lefebvre, C., McDonald, S., & Group, C. D. (2002). Development of the Cochrane Collaboration's central register of controlled clinical trials. *Evaluation & the Health Professions*, *25*, 38–64. doi: 10.1177/016327870202500104

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2011). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE*, *7*, e29081. doi: 10.1371/journal.pone.0029081

Easterbrook, P. J., Gopalan, R., Berlin, J. A., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, *337*, 867–872. doi: 10.1016/0140-6736(91)90201-y

Evangelou, E., Trikalinos, T. A., & Ioannidis, J. P. (2005). Unavailability of online supplementary scientific information from articles published in major journals. *The FASEB Journal*, *19*, 1943–1944. doi: 10.1096/fj.05-4784lsf

Fanelli, D. (2010a). Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS ONE*, *5*, e10271. doi: 10.1371/journal.pone.0010271

Fanelli, D. (2010b). ''Positive'' results increase down the hierarchy of the sciences. *PLoS ONE*, *5*, e10068. doi: 10.1371/journal.pone.0010068

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

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, *17*, 120–128. doi: 10.1037/a0024445

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*, 555–561. doi: 10.1177/1745691612459059

Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, *19*, 151–156. doi: 10.3758/s13423-012-0227-9

Fuchs, H. M., Jenny, M., & Fiedler, S. (2012). Psychologists are open to change, yet wary of rules. *Perspectives on Psychological Science*, *7*, 639–642. doi: 10.1177/1745691612459521

Gallup Organization. (2008). Observing and reporting suspected misconduct in biomedical research. Retrieved from http://ori.hhs.gov/sites/default/files/gallup_finalreport.pdf

Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, *7*, 562–571. doi: 10.1177/1745691612457576

Goldacre, B. (2012). *Bad pharma: How drug companies mislead doctors and harm patients*. London, England: Fourth Estate.

Goldstein, D. (2010). *On fact and fraud: Cautionary tales from the front lines of science*. Princeton, NJ: Princeton University Press.

Gomez-Mejia, L. R., & Balkin, D. B. (1992). Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal*, *35*, 921–955. doi: 10.2307/256535

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*, 1–20. doi: 10.1037/h0076157

Greenwald, A. G. (2012). There is nothing so theoretical as a good method. *Perspectives on Psychological Science*, *7*, 99–108. doi: 10.1177/1745691611434210

Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, *50*, 1348–1352. doi: 10.5465/amj.2007.28166119

Hanson, B., Sugden, A., & Alberts, B. (2011). Making data maximally available. *Science*, *331*, 649. doi: 10.1126/science.1203354

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *The Academy of Management Review*, *32*, 1199–1228. doi: 10.2307/20159363

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

Ioannidis, J. P. A. (2006). Concentration of the most-cited papers in the scientific literature: Analysis of journal ecosystems. *PLoS ONE*, *1*, e5. doi: 10.1371/journal.pone.0000005

Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, *7*, 645–654. doi: 10.1177/1745691612464056

Ioannidis, J. P. A., Allison, D. B., Ball, C. A., Coulibaly, I., Cui, X., Culhane, A. C., … van Noort, V. (2009). Repeatability of published microarray gene expression analyses. *Nature Genetics*, *41*, 149–155. doi: 10.1038/ng.295

Jasny, B. R., Chin, G., Chong, L., & Vignieri, S. (2011). Again, and again, and again…. *Science*, *334*, 1225. doi: 10.1126/science.334.6060.1225

Jefferson, T. (1998). Redundant publication in biomedical sciences: Scientific misconduct or necessity? *Science and Engineering Ethics*, *4*, 135–140. doi: 10.1007/s11948-998-0043-9

Kepes, S., Banks, G. C., McDaniel, M. A., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, *15*, 624–662. doi: 10.1177/1094428112452760

Kepes, S., Banks, G., & Oh, I.-S. (in press). Avoiding bias in publication bias research: The value of ''null'' findings. *Journal of Business and Psychology*. doi: 10.1007/s10869-012-9279-0

Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic reviews in the organizational sciences: Two meta-analytic schools on the way to MARS (the Meta-analytic Reporting Standards). *Journal of Business and Psychology*, *28*, 123–143. doi: 10.1007/s10869-013-9300-2

Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *Academy of Management Journal*, *18*, 769–783. doi: 10.2307/255378

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*, 196–217. doi: 10.1207/s15327957pspr0203_4

Koole, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science*, *7*, 608–614. doi: 10.1177/1745691612462586

Kozlowski, S. W. J. (2009). Editorial. *Journal of Applied Psychology*, *94*, 1–4. doi: 10.1037/a0014990

Krimsky, S. (2006). Publication bias, data ownership, and the funding effect in science: Threats to the integrity of biomedical research. In W. Wagner, & R. Steinzor (Eds.), *Rescuing science from politics: Regulation and the distortion of scientific research* (pp. 61–85). New York, NY: Cambridge University Press.

LaFollette, M. C. (1992). *Stealing into print: Fraud, plagiarism, and misconduct in scientific publishing*. Los Angeles, CA: University of California Press.

Laine, C., Horton, R., DeAngelis, C. D., Drazen, J. M., Frizelle, F. A., Godlee, F., … Verheugt, F. W. A. (2007). Clinical trial registration: Looking back and moving ahead. *New England Journal of Medicine*, *356*, 2734–2736. doi: 10.1056/NEJMe078110

LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology*, *15*, 371–379. doi: 10.1037/a0025172

Lehrer, J. (2010). The truth wears off. *New Yorker*, *86*, 52–57.

Leung, K. (2011). Presenting post hoc hypotheses as a priori: Ethical and theoretical issues. *Management and Organization Review*, *7*, 471–479. doi: 10.1111/j.1740-8784.2011.00222.x

Lewin, K. (1952). *Field theory in social science: Selected theoretical papers by Kurt Lewin*. London, England: Tavistock.

Liberati, A. (1992). Publication bias and the editorial process. *Journal of the American Medical Association*, *267*, 2891. doi: 10.1001/jama.1992.03480210049017

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

McDaniel, M. A., McKay, P., & Rothstein, H. R. (2006). *Publication bias and racial effects on job* *performance*: *The elephant in the room*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

McDaniel, M. A., Rothstein, H. R., & Whetzel, D. L. (2006). Publication bias: A case study of four test vendors. *Personnel Psychology*, *59*, 927–953. doi: 10.1111/j.1744-6570.2006.00059.x

McDaniel, M. A., & Whetzel, D. L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence*, *33*, 515–525. doi: 10.1016/j.intell.2005.02.001

Merton, R. K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology*, *1*, 115–126.

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

Murphy, J. R. (2004). Statistical errors in immunologic research. *The Journal of allergy and clinical immunology*, *114*, 1259–1263.

Murtaugh, P. A. (2002). Journal quality, effect size, and publication bias in meta-analysis. *Ecology*, *83*, 1162–1166. doi: 10.1890/0012-9658(2002)083[1162:jqesap]2.0.co;2

Neuliep, J. W., & Crandall, R. (1990). Editorial bias against replication research. *Journal of Social Behavior & Personality*, *5*, 85–90.

Neuliep, J. W., & Crandall, R. (1993). Reviewer bias against replication research. *Journal of Social Behavior & Personality*, *8*, 21–29.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220. doi: 10.1037/1089-2680.2.2.175

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*, 615–631. doi: 10.1177/1745691612459058

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*, 657–660. doi: 10.1177/1745691612462588

Orlitzky, M. (2012). How can significance tests be deinstitutionalized? *Organizational Research Methods*, *15*, 199–228. doi: 10.1177/1094428111428356

Pace, V. L., & Brannick, M. T. (2010). How similar are personality scales of the ''same'' construct? A meta-analytic investigation. *Personality and Individual Differences*, *49*, 669–676. doi: 10.1016/j.paid.2010.06.014

Park, R. L. (2001). *Voodoo science: The road from foolishness to fraud*. Oxford, England: Oxford University Press.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*, 531–536. doi: 10.1177/1745691612463401

Popper, K. R. (1959). *The logic of scientific discovery*. Oxford, England: Basic Books.

Porter, T. M. (1992). Quantification and the accounting ideal in science. *Social Studies of Science*, *22*, 633–652. doi: 10.1177/030631292022004004

Rauscher, F. H., Shaw, G. L., & Ky, C. N. (1993). Music and spatial task performance. *Nature*, *365*, 611. doi: 10.1038/365611a0

Reich, E. S. (2010). *Plastic fantastic: How the biggest fraud in physics shook the scientific world*. New York, NY: Palgrave Macmillan.

Renkewitz, F., Fuchs, H. M., & Fiedler, S. (2011). Is there evidence of publication biases in JDM research? *Judgment and Decision Making*, *6*, 870–881.

Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: Three unsuccessful attempts to replicate Bem's 'retroactive facilitation of recall' effect. *PLoS ONE*, *7*. doi: 10.1371/journal.pone.0033423

Roberson, Q. M., Sturman, M. C., & Simons, T. L. (2007). Does the measure of dispersion matter in multilevel research? A comparison of the relative performance of dispersion indexes. *Organizational Research Methods*, *10*, 564–588. doi: 10.1177/1094428106294746

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638–641. doi: 10.1037/0033-2909.86.3.638

Rothstein, H. (2012). Accessing relevant literature. In H. M. Cooper (Ed.), *APA handbook of research methods in psychology: Vol. 1. Foundations, planning, measures, and psychometrics* (pp. 133–144). Washington, DC: American Psychological Association.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005a). Publication bias in meta-analyses. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta analysis: Prevention, assessment, and adjustments* (pp. 1–7). West Sussex, England: Wiley.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005b). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. West Sussex, England: Wiley.

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review*, *18*, 682–689. doi: 10.3758/s13423-011-0088-7

Rupp, D. E. (2011). Ethical issues faced by editors and reviewers. *Management and Organization Review*, *7*, 481–493. doi: 10.1111/j.1740-8784.2011.00227.x

Sarawitz, D. (2012). Beware the creeping cracks of bias. *Nature*, *485*, 149. doi: 10.1038/485149a

Saul, S. (2008, October 8). Experts conclude Pfizer manipulated studies. *New York Times*, p. 4. Retrieved from http://www.nytimes.com/2008/10/08/health/research/08drug.html

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*, 90–100. doi: 10.1037/a0015108

Schminke, M. (2009). Editor's comments: The better angels of our time-Ethics and integrity in the publishing process. *The Academy of Management Review*, *34*, 586–591. doi: 10.5465/amr.2009.44882922

Schminke, M., & Ambrose, M. L. (2011). Ethics and integrity in the publishing process: Myths, facts, and a roadmap. *Management and Organization Review*, *7*, 397–406. doi: 10.1111/j.1740-8784.2011.00248.x

Schmitt, N. (1989). Editorial. *Journal of Applied Psychology*, *74*, 843–845. doi: 10.1037/h0092216

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*, 36–71. doi: 10.1016/0010-0285(90)90003-m

Schwab, D. P. (2005). *Research methods for organizational studies* (2nd ed.). Mahwah, NJ: Erlbaum.

Shields, P. G. (2000). Publication bias is a scientific problem with adverse ethical outcomes: The case for a section for null results. *Cancer Epidemiology Biomarkers & Prevention*, *9*, 771–772.

Shields, P. G., Sellers, T. A., & Rebbeck, T. R. (2009). Null results in brief: Meeting a need in changing times. *Cancer Epidemiology Biomarkers & Prevention*, *18*, 2347. doi: 10.1158/1055-9965.epi-09-0684

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. doi: 10.1177/0956797611417632

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association*, *54*, 30–34. doi: 10.2307/2282137

Sterling, T. D., & Rosenbaum, W. L. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician*, *49*, 108–112. doi: 10.1080/00031305.1995.10476125

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., . . . Grigorenko, E. L. (2000). *Practical intelligence in everyday life*. New York, NY: Cambridge University Press.

Strasak, A. M., Zaman, Q., Marinell, G., Pfeiffer, K. P., & Ulmer, H. (2007). The use of statistics in medical research. *The American Statistician*, *61*, 47–55. doi: 10.1198/000313007x170242

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

Sutton, A. J. (2005). Evidence concerning the consequences of publication and related biases. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 175–192). West Sussex, England: Wiley.

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 435–452). New York, NY: Russell Sage Foundation.

Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, *40*, 371–384. doi: 10.2307/2393788

Trikalinos, T. A., & Ioannidis, J. P. A. (2005). Assessing the evolution of effect sizes over time. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 241–259). West Sussex, England: Wiley.

Turner, E. H. (2004). A taxpayer-funded clinical trials registry and results database. *PLoS Medicine*, *1*, e60. doi: 10.1371/journal.pmed.0010060

Uchino, B. N., Thoman, D., & Byerly, S. (2010). Inference patterns in theoretical social psychology: Looking back as we move forward. *Social and Personality Psychology Compass*, *4*, 417–427. doi: 10.1111/j.1751-9004.2010.00272.x

Van Dongen, S., & Gangestad, S. W. (2011). Human fluctuating asymmetry in relation to health and quality: A meta-analysis. *Evolution and Human Behavior*, *32*, 380–398. doi: 10.1016/j.evolhumbehav.2011.03.002

Vogel, G. (2011). Psychologist accused of fraud on ‘‘astonishing scale.’’ *Science*, *334*, 579. doi: 10.1126/science.334.6056.579

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

Whalen, J., Barrett, D., & Loftus, P. (2012, July 3). Glaxo sets guilty plea, $3 billion settlement. *Wall Street Journal*, B1. Retrieved from http://online.wsj.com/article/SB10001424052702304299704577502642401041730.html

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, *6*, e26828. doi: 10.1371/journal.pone.0026828

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*, 726–728. doi: 10.1037/0003-066x.61.7.726

Witten, D. M., & Tibshirani, R. (2012). Scientific research in the age of omics: The good, the bad, and the sloppy. *Journal of the American Medical Informatics Association*, *1–3*. doi: 10.1136/amiajnl-2012-000972

Wolins, L. (1962). Responsibility for raw data. *American Psychologist*, *17*, 657–658. doi: 10.1037/h0038819

Yong, E. (2012a). Replication studies: Bad copy. *Nature*, *485*, 298–300. doi: 10.1038/485298a

Yong, E. (2012b). Uncertainty shrouds psychologist’s resignation. *Nature*. doi: 10.1038/nature.2012.10968