2023

The Trustworthiness of the Cumulative Knowledge in Industrial/Organizational Psychology: The Current State of Affairs and a Path Forward

Sheila K. Keener
Old Dominion University, skeener@odu.edu

Sven Kepes
Virginia Commonwealth University

Ann-Kathrin Torka
TU Dortmund University

Original Publication Citation
Keener, S. K., Kepes, S., & Torka, A. K. (2023). The trustworthiness of the cumulative knowledge in industrial/organizational psychology: The current state of affairs and a path forward. *Acta Psychologica*, *239*, 1-15, Article 104005. https://doi.org/10.1016/j.actpsy.2023.104005

Acta Psychologica 239 (2023) 104005

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

The trustworthiness of the cumulative knowledge in industrial/organizational psychology: The current state of affairs and a path forward

Sheila K. Keener[a,*,1], Sven Kepes[b,1], Ann-Kathrin Torka[c,1]

[a] Department of Management, Old Dominion University, Norfolk, VA, United States of America
[b] Department of Management and Entrepreneurship, Virginia Commonwealth University, Richmond, VA, United States of America
[c] Department of Social, Work, and Organizational Psychology, TU Dortmund University, Dortmund, Germany

ARTICLE INFO

Keywords:
Reproducibility
Replicability
Scientific misconduct
Questionable research practices
Trustworthiness of our scientific knowledge
Open science practices

ABSTRACT

The goal of industrial/organizational (IO) psychology, is to build and organize trustworthy knowledge about people-related phenomena in the workplace. Unfortunately, as with other scientific disciplines, our discipline may be experiencing a “crisis of confidence” stemming from the lack of reproducibility and replicability of many of our field's research findings, which would suggest that much of our research may be untrustworthy. If a scientific discipline's research is deemed untrustworthy, it can have dire consequences, including the withdraw of funding for future research. In this focal article, we review the current state of reproducibility and replicability in IO psychology and related fields. As part of this review, we discuss factors that make it less likely that research findings will be trustworthy, including the prevalence of scientific misconduct, questionable research practices (QRPs), and errors. We then identify some root causes of these issues and provide several potential remedies. In particular, we highlight the need for improved research methods and statistics training as well as a re-alignment of the incentive structure in academia. To accomplish this, we advocate for changes in the reward structure, improvements to the peer review process, and the implementation of open science practices. Overall, addressing the current “crisis of confidence” in IO psychology requires individual researchers, academic institutions, and publishers to embrace system-wide change.

1. Introduction

Science is a systematic endeavor that builds and organizes knowledge in the form of testable explanations and predictions about nature and the universe (Heilbron, 2003). In the scientific discipline of industrial/organizational (IO) psychology, this endeavor concerns the study of the human mind and behavior at work. Thus, IO psychologists develop predictions and collect and analyze data to test their predictions with the objective to better understand how people think, feel, and behave at work and to help solve problems in the workplace[2]. To generate accurate and trustworthy cumulative knowledge on these people-related phenomena, the scientific method must be followed, and the editorial review process ought to ensure that deviations and errors are identified and corrected. In cases where misleading or erroneous results are published and enter the cumulative knowledge, science itself ought to be self-correcting. That is, new research should test and confirm previously published findings. Indeed, it is generally believed that published erroneous findings will get detected as these events are rare and “occur in a system [i.e., the scientific method] that operates in an effective, democratic and self-correcting mode” (Broad & Wade, 1982, p. 11–12).

Unfortunately, over the past decade or so, there has been growing concern about the state of this self-correcting mechanism in the sciences generally (e.g., Stroebe et al., 2012) and in IO psychology and related fields, such as social psychology and management, specifically (e.g., Bergh et al., 2017; Byington & Felps, 2017; Earp & Trafimow, 2015; Hensel, 2021; Kepes & McDaniel, 2013; Pashler & Wagenmakers, 2012). For the last decade or more, articles in the popular press (e.g., Carey, 2011; Lehrer, 2010; Yong, 2018) and academic journals (e.g., Ioannidis, 2005; Kepes & McDaniel, 2013; O'Boyle et al., 2017; Simmons et al.,

* Corresponding author at: Strome College of Business, Old Dominion University, 2037 Constant Hall, Norfolk, VA 23529, United States of America.
E-mail addresses: skeener@odu.edu (S.K. Keener), skepes@vcu.edu (S. Kepes), ann-kathrin.torka@tu-dortmund.de (A.-K. Torka).

[1] The order of authorship is alphabetical as they contributed equally to this work.

[2] We recognize that IO psychologists may also engage in qualitative and/or exploratory research. While our primary emphasis is on quantitative research involving predictions, it is important to note that qualitative and exploratory research is not immune from the challenges discussed in our paper.

https://doi.org/10.1016/j.actpsy.2023.104005
Received 12 April 2023; Received in revised form 13 July 2023; Accepted 4 August 2023
Available online 23 August 2023

2011) have repeatedly questioned the robustness and trustworthiness of scientific knowledge. This can lead to reputational damage and the loss of public trust (e.g., Chopik et al., 2018; Simmons et al., 2011; Wingen et al., 2020). Indeed, Broomell and Kane (2017) found that uncertainty regarding a scientific field's evidence leads to the field as being perceived as less valuable. This can have drastic consequences. In addition to shaping future research agendas by misleading or erroneous findings, the misallocation of limited resources for scientific research, and the widening of the often-lamented science-practice gap (e.g., Aguinis et al., 2020; Kepes et al., 2014; Rynes et al., 2018), the loss of public trust has the potential to increase the already growing anti-science movement (Hotez, 2021; Philipp-Muller et al., 2022). This, in turn, is likely to lead to less public support and, therefore, funding for important scientific endeavors. Indeed, drastic cuts in science spending in countries across the world, ranging from the U.S. (Ledford et al., 2019) and U.K. (Patten, 2021) to Brazil (Kowaltowski, 2021) and India (Nair, 2019), have been proposed or enacted.

The social sciences are often especially affected by these (proposed) cuts. In the U.S., there have been and continue to be proposals to cut federal support for the National Science Foundation's funding for the social sciences (e.g., Matthews, 2014; Ross, 2017; Sides, 2015). It seems likely that these attempts will only grow if the public continues to lose trust in our scientific endeavors and published findings. As the House Science Committee told Ed Yong recently, “there's a lot of sloppy science that's out there - irreproducible science” (Yong, 2017). Therefore, in this review, we discuss the current state of affairs in IO psychology, and related fields such as social psychology and management, as it relates to its trustworthiness and credibility. Where we lack IO psychology-specific evidence (or social psychology/management), we highlight this and relate findings from other fields. As part of our review, we discuss factors that affect our field's credibility and the trustworthiness of its cumulative knowledge. We then provide several suggestions for how the identified problems can be addressed, including highlighting some ways in which we are already making progress.

2. The current state of affairs

As previously stated, science should be self-correcting. Two of the important pillars of the self-correction process are reproducibility and replicability. That is, published studies should be reproducible and replicable. *Reproducibility* denotes the ability of researchers to obtain the same results when they reanalyze the data of a published study. Once a particular research finding has been reproduced, the finding's *replicability* should be assessed to determine whether the original study's findings can be obtained using other random samples (Asendorpf et al., 2013; Kepes et al., 2014). As such, reproducibility is often viewed as a necessary but insufficient condition for replicability (Aguinis et al., 2018; Asendorpf et al., 2013). Reproducibility and replicability have been called the “cornerstone of science” (Moonesinghe et al., 2007, p. 218; Simons, 2014, p. 76) as they are methodological approaches to confirm or disconfirm as well as build on previously published results and, therefore, generate cumulative scientific knowledge. In other words, under the premise of “trust but verify,” they provide the necessary proof that the cumulative knowledge on a particular phenomenon (e.g., the relation between *X* and *Y*) is accurate and trustworthy. In addition, replications help to identify boundary conditions, which is necessary to establish the generalizability of published findings. Although evidence indicates that published findings are rarely reproduced or replicated in psychology (Makel et al., 2012; Neuliep & Crandall, 1990), prompting some to call this a “crisis of confidence” (e.g., Earp & Trafimow, 2015, p. 1), there also have been more sanguine (e.g., Maxwell et al., 2015) or even contradictory voices (e.g., Gilbert et al., 2016; Schmidt & Oh, 2016). Generally, these dissenting perspectives claim that changes in the underlying protocol of a study are a reason for being unable to replicate its findings (e.g., Gilbert et al., 2016). Alternatively, they suggest that issues related to sampling error and statistical power in replication studies explain the lack of successful replications (e.g., Gilbert et al., 2016; Maxwell et al., 2015; Schmidt & Oh, 2016), or state that replications do occur frequently (e.g., Schmidt & Oh, 2016).

As Köhler and Cortina (2019) illustrated, different perceptions regarding the extent of the replication crisis may be due to a lack of precision and clarity when we use terms such as reproducibility and, in particular, replicability as there are different forms of replication, including literal or exact, quasi-random, constructive, confounded, and regressive replications. Not distinguishing between these terms and, instead, using them interchangeably tends to muddy the waters. Furthermore, the Open Science Collaboration's (OSC's) replication attempts (discussed in more detail in the coming sections) generally used replications that were higher-powered than the initial studies (OSC, 2015). Lastly, the argument that “the large number of meta-analyses in our literatures shows that replication studies are in fact being conducted in most areas of research” (Schmidt & Oh, 2016, p. 32) fails to account for the large and substantial degrees of heterogeneity published meta-analytic mean estimates often entail (Kepes, Wang, Cortina, 2023), which explicitly indicates that research findings do not necessarily replicate.

Next, we review the current evidence regarding the reproducibility and replicability of research findings. Although there is little to no research on these issues explicitly in the field of IO psychology, there is ample evidence in related disciplines, including general psychology, social psychology, management, and economics, and there is little reason to assume that the situation is different in IO.

2.1. Reproducibility

As noted previously, reproducible findings are results that can be verified by a third party, typically an independent researcher, using the same data and the same methodological approaches and steps. Recently, Artner et al. (2021) examined the reproducibility of major statistical conclusions drawn from 46 articles in 2012 by three journals from the *American Psychological Association* for which the raw data were available. The researchers identified 232 key statistical claims and attempted to reproduce the underlying statistical results (185 of these claims were associated with statistically significant results). They were only able to successfully reproduce 163 (70.3 %) of the 232 claims following the analytical approach outlined in the original articles. An additional 18 (7.8 %) could be verified by deviating from the methodological description in the respective articles (the remaining 51 [22 %] could not be replicated). Notably, 13 (7 %) of the 185 claims deemed statistically significant by the authors of the original studies (out of the 232 total claims) were no longer significant upon reproduction. Other studies in psychology (e.g., Bakker & Wicherts, 2011; Hardwicke et al., 2018; Wolins, 1962), strategic management (Bergh et al., 2017; Goldfarb & King, 2016), economics (Chang & Li, 2022), and the medical sciences (e.g., Bergeat et al., 2022; Ioannidis et al., 2009; Naudet et al., 2018) reported similar findings, suggesting that the relatively low levels of reproducibility are not limited to psychology but, instead, are present in virtually all of the social and medical sciences.

To determine the reproducibility of meta-analytic studies in psychology, Maassen et al. (2020) conducted an interesting study with two parts. First, they selected 33 meta-analytic studies that included a data table with the primary studies that contributed data to the respective meta-analysis. Overall, the 33 meta-analytic studies included 1978 primary study effect sizes. Next, they tried to reproduce (i.e., re-calculate) 500 randomly selected primary study effect sizes and found that they could only do so without any issues in 276 (55.20 %) of the cases. Then, in part 2 of their study, Maassen et al. (2020) estimated the effect of non-reproducible primary study effect size data on meta-analytic results. Unsurprisingly, the authors found that the meta-analytic results (e.g., mean effect size estimate, confidence interval, heterogeneity statistics) of 13 (39.39 %) meta-analytic studies were adversely affected. In sum, it seems that reproducibility is a serious concern in the sciences overall and, particularly, psychology.

2.2. Replicability

Examinations regarding the replicability of research findings have taken two major forms. First, reviews of the published literature have examined how many of the published studies in journals are replications. For instance, in their review of the 100 most prestigious psychology journals, Makel et al. (2012) found that only about 1.6 % of all articles published since 1900 used the term 'replication' in the text. In a more extensive examination of 500 randomly selected articles, the authors estimated an overall replication rate of 1.07 %. More recently, using a sample of articles in three of the most prestigious IO psychology and management journals, Köhler and Cortina (2019) showed that some types of replications are fairly common, while others are rare.

In their study, Köhler and Cortina (2019) first distinguished between five different types of replications, literal replications, quasirandom replications, constructive replications, confounded replications, and regressive replications. In addition, all types of replications can be dependent (i.e., the same researchers that conducted the original study are conducting a replication) or independent (i.e., different researchers are conducting a replication independently).[3]

Each of these types of replications have different purposes. For instance, while the purpose of a literal replication is to exactly replicate the original study, including the sample, research design, measures, and statistical procedures, the purpose of a constructive replication is to improve upon the original study by, for instance, using a more carefully defined sample, more valid measures, or more sophisticated statistical techniques. Quasirandom replications fall between these two types and confounded and regressive replications tend to contain methodological compromises that can weaken the rigor of the original study. Thus, constructive replications may be the most valuable as they explicitly attempt to improve the original study, allowing one to gain new insights.

Köhler and Cortina (2019) found that quasirandom replications are quite common in IO psychology and management, which may be responsible for the perception that plenty of replications exist (e.g., Schmidt & Oh, 2016). However, these types of replications are generally not conducted with the express purpose of improving upon the original study. Instead, factors such as convenience or familiarity with a particular methodological aspect of the study motivates these replications. As such, it is typically unclear whether the incorporated modifications actually strengthen or weaken the original study. Therefore, these types of replications tend not be particularly valuable when trying to better understand a phenomenon of interest and make scientific progress. Unfortunately, the most beneficial type of replication, the constructive replication, which allows one to gain new insights, was virtually absent from Köhler and Cortina's sample. This, specifically, is what they labeled the "replication crisis" (p. 510).

The second type of study regarding the replicability of research findings are direct replication attempts, often including several replications in one large-scale project. Many of these are conducted with the goal of being literal replications. However, they often fall short of this objective as, for instance, the methodological descriptions in the original studies are not clear enough. Similarly, sometimes, these replication efforts try to improve upon the original studies, often by using larger samples to increase the statistical power. The probably most influential empirical investigation of this type stems from the OSC. In the early 2010s, the OSC started to conduct replications of 100 experimental and correlational studies published in three psychology journals, typically with higher-powered designs when compared to the original published studies. Using three different criteria to determine replicability, the OSC concluded that "a large portion of replications produced weaker evidence for the original findings despite using materials provided by the original authors, review in advance for methodological fidelity, and high statistical power to detect the original effect sizes" (OSC, 2015, p. 943). As an example, only 35 (36.08 %) of the 97 originally reported significant effect sizes were also significant in the replication, which is a substantial and statistically significant reduction. These replicability rates were lower in social psychology (14/55; 25 %) than in cognitive psychology (21/42; 50 %).

Many other large-scale replication efforts have been published in psychology, particularly social psychology (e.g., Ebersole et al., 2016; Ebersole et al., 2020; Klein et al., 2014; Klein et al., 2022; Ritchie et al., 2012). Although replicability rates in these studies have varied (e.g., 3/10 [30 %] in Ebersole et al., 2016; 10/13 [77 %] in Klein et al., 2014), they generally suggest replication rates that are less than desirable. Unfortunately, there are no studies focusing explicitly on the replicability of findings in IO psychology. This is possibly due to the shunning of exact replication in most of our journals (Kepes & McDaniel, 2013; Martin & Clarke, 2017; Tipu & Ryan, 2021). Indeed, in Köhler and Cortina's (2019) sample of IO psychology and management journals, literal (or direct) replications were virtually absent. However, given the similarity of IO and social psychology, which shows generally low rates of successful replication, we have no reason to believe the situation is any different in our field.

In sum, the evidence regarding the reproducibility and replicability of research in psychology, particularly social psychology, is disheartening, which also bodes poorly for our discipline. Furthermore, recent evidence indicates that failed replications of published psychological studies have little effect on the citation rates of the originally published studies (von Hippel, 2022). As such, it seems as if replication failure does not affect the influence of non-replicated findings, which means self-correction may not be occurring.

3. Causes of the replication crisis and the untrustworthiness of our cumulative knowledge

There are several factors that make it less likely that research findings will be reproducible or replicable and ultimately contribute to the untrustworthiness of the cumulative knowledge in IO psychology. Three specific issues include scientific misconduct, the use of questionable research practices (QRPs), and errors in scientific studies. Scientific misconduct includes behaviors such as fabricating or falsifying data or results, plagiarism, or otherwise mischaracterizing a study's research method, such that the stated approach and findings do not represent the true way in which a study or its results was conducted (e.g., Stroebe et al., 2012). There have been several high-profile cases of individuals who conduct research in the areas of IO psychology, social psychology, and management engaging in scientific misconduct. For instance, David Degeest, admitting to falsifying results, leading to four retractions (Retraction Watch, 2018). As a more extreme example, Diederik Stapel was found to have fabricated data for several studies, resulting in 58 retractions (e.g., Callaway, 2011; Retraction Watch, 2015). Notably, in these cases, many of the retracted articles had been published at prestigious journals (e.g., *Science, Organizational Behavior and Decision Processes, Psychological Science, Personality and Social Psychology Bulletin, Journal of Organizational Behavior, Journal of Management*) and, thus, likely had an outsized influence on cumulative knowledge and future research agendas.

Besides these cases, there are other instances of misconduct that can be identified by examining the stated reasons for retractions. For instance, Stricker and Günther's (2019) analysis of retractions of IO psychology-related articles from PsycINFO found that misconduct was

[3] Theoretically, independent replications tend to be favored as this type addresses potential conflicts of interests and confirmation biases (e.g., the original study authors may have a vested interested in replicating the originally obtained results). However, recent studies by Landy et al. (2020) and Schweinsberg et al. (2021) illustrate that such replication attempts can face substantial practical hurdles (e.g., different subjective design choices or operationalizations can yield vastly different results).

identified as the cause of retractions at a rate of 0.77 per 10,000 articles published between 1860 and 2017. Focusing on a more recent sample (1998–2017) of 160 retractions published in psychology journals, Craig et al. (2020) found that, across psychology disciplines, data fabrication, falsification, and fraud accounted for 48 % of the retractions and plagiarism, including self-plagiarism, accounted for an additional 13 %. In a similar investigation of retractions of business and management studies, Tourish and Craig (2020) found that misconduct was also a frequent cause for retractions. Specifically, data fraud accounted for 33% of the 154 reasons provided for the 131 retractions included in their analysis; plagiarism and self-plagiarism accounted for an additional 25 %. Together, these findings suggest that when articles are retracted, misconduct is a common contributing factor; yet retractions overall, and scientific misconduct specifically, are, fairly rare.

Of course, these cases of misconduct were eventually identified. Unfortunately, other research suggests that there may be additional instances of misconduct that have not been caught. Indeed, in the field of management, Bedeian et al. (2010) found that 26.8 % of the surveyed faculty stated they were aware of at least one faculty member fabricating their data within the past year. It is difficult to say how much misconduct this actually translates to, but it suggests that misconduct is likely to occur and may not be detected. That being said, across two studies of management researchers, Banks et al. (2016) found fairly low rates of data falsification – only 0.4 % of their sample admitted to ever falsifying data. This is consistent with the rates suggested in John et al.'s (2012) survey of American psychologists (i.e., 0.6 % of those surveyed indicated that they had falsified data). Estimates from other studies of psychologists have been a bit higher, however. Specifically, 2.3 % of Italian psychologists admitted to falsifying data (Agnoli et al., 2017). Yet any estimate from the literature is likely to underestimate the true prevalence of scientific misconduct, as researchers may be unlikely to admit to such behavior and unaware, or unwilling to believe, that their colleagues do. Thus, it is reasonable to suspect that there is more scientific misconduct occurring than suggested by these estimates, and certainly more than is identified and labeled as such through retractions.

Scientific misconduct is not the only behavior that affects the trustworthiness of our cumulative knowledge, however. QRPs, including selectively reporting hypotheses, hypothesizing after the results are known (HARKing), adding or dropping data, and modifying scales (e.g., removing items) after data are collected (see Table 1 for examples of QRPs), characterize a set of behaviors that skirt the line between ethical and unethical research practice. For instance, dropping data from a data set may be perfectly acceptable in some instances (e.g., if the datapoints are identified as outliers using a priori decision rules). Yet, if datapoints are dropped purely because doing so moves a marginally significant result to below the magical $p < .05$ threshold, many researchers would conclude that there is no legitimate justification for dropping these data.

The frequency of engagement in QRPs varies depending on the specific type of QRP. For instance, Banks et al. (2016) asked two samples of management researchers if they had ever engaged in a selection of behaviors: 11.1 % admitted to rounding off p-values, 49.7 % selectively reported hypotheses that "worked", 49.6 % engaged in HARKing, 28.5 % decided to drop data after looking at how doing so would impact their results, and 33.3 % had selectively included/excluded control variables based on statistical significance. Similar rates of QRP engagement have been found in studies of American (John et al., 2012) and Italian (Agnoli et al., 2017) psychologists.

As discussed earlier, these percentages may underestimate the true prevalence of QRPs as they rely on researchers freely admitting to the behavior. Thus, there have been other attempts to examine these behaviors using methods that do not rely on self-admission. One approach has been to compare published and unpublished studies to uncover differences between them that may suggest QRP engagement. For instance, Mazzola and Deuling (2013) compared journal articles and dissertations in IO psychology to determine which had more supported and unsupported hypotheses. Consistent with expectations, journal articles had significantly more supported hypotheses and significantly fewer unsupported hypotheses as compared to dissertations, which suggests that IO researchers are engaging in selective outcome reporting and/or HARKing. A similar study of management researchers at top universities compared different versions of the same study (i.e., dissertations to the published version of the dissertation; Kepes et al., 2022). Evidence showed that unsupported dissertation hypotheses were dropped from published articles at a higher rate than supported dissertation hypotheses. Further evidence indicated that newly created hypotheses were more likely to be supported than unsupported (suggesting HARKing). They also found that researchers engaged in a variety of QRPs, such as adding data and changing covariates to change unsupported dissertation hypotheses to supported hypotheses in published articles. Similar findings have been observed in other studies in management (O'Boyle et al., 2017) and psychology (Franco et al., 2016).

Another approach to determine how common QRP engagement occurs is examine trends in published studies. For instance, in two studies (one examining articles published in two top IO journals and one examining a meta-analysis on the relation between job satisfaction and job performance), Bosco et al. (2016), showed that effect sizes were larger for hypothesized relations than non-hypothesized relations. After ruling out several other potential explanations, they concluded that these findings were consistent with HARKing. As another example, O'Boyle et al. (2019), reviewed six top journals in the fields of IO psychology and management and found that, despite low statistical power, most reported moderated multiple regression models were statistically significant. Further examination uncovered factors (e.g., an increase in p-values just below the 0.05 cutoff) that suggested outcome reporting bias was a contributing factor. Similar findings have been observed in psychology for mediation effects (Götz et al., 2021). Taken together, these studies provide evidence that QRP engagement among IO and management researchers is quite common and much more common than outright scientific misconduct like completely fabricating one's data.

Besides intentional unethical or questionable behaviors, errors in scientific studies also contribute to potential untrustworthiness. Though true errors are unintentional and the reasons for their occurrence may be different from those for misconduct or QRPs, they still make it less likely that research findings will replicate.[4] Errors are not exactly rare either. Bakker and Wicherts (2011) found in two studies that between 9.7 % (Study 1) and 12.8 % (Study 2) of results in a sample of psychology articles were reported incorrectly. Furthermore, between 55 % (Study 1) and 35 % (Study 2) of articles contained at least one error. Regarding specific types of errors, in their sensitivity analyses, Nuijten et al. (2016)

Table 1
Examples of questionable research practices.

Questionable research practices	Example articles for more information
Selective reporting of hypotheses	Fanelli (2010, 2012), Greenwald (1975), Sterling et al. (1995)
Selective reporting of conditions	Agnoli et al. (2017), John et al. (2012)
Selective reporting of studies	Agnoli et al. (2017), Fanelli (2010, 2012), John et al. (2012)
Adding/dropping data	Kepes et al. (2022), O'Boyle et al. (2017)
Rounding p-values	Hartgerink et al. (2016), Nuijten et al. (2016)
HARKing	Kepes et al. (2022), Kerr (1998)
Modifying measures (e.g., scales) after data collection	Kepes et al. (2022)
Adding/dropping DVs	Agnoli et al. (2017), John et al. (2012)
Adding/dropping covariates	Kepes et al. (2022)

4 Although these are described here as "errors", it is difficult to determine if the identified issues are due to real errors. It is also possible that the information was intentionally misreported to distort study findings and misrepresent the conclusions that could be drawn from the study. This could ultimately enhance the chances of publication.

found that almost half of the psychology articles they examined had at least one inconsistency between true *p*-values and the reported *p*-values and roughly 13 % of articles had "at least one gross inconsistency" (p. 1209). On a positive note, however, the *Journal of Applied Psychology* (*JAP*) had the lowest rate of overall inconsistencies (only 33.6 % of articles had at least one inconsistency; 12.4 % had at least one gross inconsistency). Similar findings regarding inconsistent *p*-values in *JAP* were reported in other studies (i.e., Veldkamp et al., 2014).

A variety of errors related to CFA and SEM models have also been reported. In two studies examining articles published in top IO and management journals (i.e., Cortina, Green et al., 2017; Harms et al., 2018), inconsistencies between reported degrees of freedom and models were noted between 38 % and 57 % of the time. Moreover, evidence suggests that some of the higher-order CFA models reported in top IO and management journals contain "demonstrably incorrect analyses" (Credé & Harms, 2015, p. 866), including chi-square values that are mathematically impossible given the models presented.

Taken together, considering the evidence provided, it appears that IO psychology is not immune from the problems that plague other disciplines. Specifically, there is ample evidence that published IO literature contains many findings that are likely affected by errors and QRPs. Furthermore, although outright scientific misconduct appears to be rare, it does still occur and there have likely been instances where it has not been caught. Thus, although the evidence regarding reproducibility and replicability in IO literature specifically is lacking, it is clear that the underlying issues that contribute to problems with reproducibility and replicability are present in our field (see also Efendic & Van Zyl, 2019).

4. Why are misconduct, QRPs, and errors problematic?

Scientific misconduct, QRPs, and errors have a variety of negative effects that contribute to issues related to reproducibility, replicability, and ultimately an untrustworthy cumulative knowledge in our field. First, high-profile cases of misconduct resulting in retractions may decrease trust in science (Stroebe et al., 2012). That being said, cases of misconduct, QRPs, and instances of errors that are *not* identified and retracted have a more insidious effect, as they appear to be legitimate but are actually inaccurate and, therefore, misleading.

While it is clear why any scientific misconduct would have a negative impact on the credibility and accuracy of the cumulative knowledge in IO psychology – the data and/or results are falsified – the impact of QRPs may be less immediately clear. As the goal of engaging in QRPs is to make one's research look "better," they usually involve manipulations that result in the hypothesized effects being statistically significant and effect sizes being inflated. Stated differently, hypotheses that would be unsupported without QRP engagement become supported. On the other hand, hypotheses that remain unsupported are likely to be dropped from the research paper. Together, this means that QRPs produce false positive results that are unlikely to replicate and, furthermore, hide evidence related to unsupported results (Rupp, 2011; Simmons et al., 2011). Thus, QRPs clearly contribute to the replication crisis (Earp & Trafimow, 2015; Maxwell et al., 2015).

Errors, regardless of their direction (in favor of, or against, anticipated results) also make it unlikely that findings will replicate. Confirmation bias suggests, however, that errors aligning with initial suppositions may be less likely to be caught than errors that go against them (see e.g., Nickerson, 1998 for a discussion of various ways that confirmation bias affects researchers). Thus, as with QRPs and misconduct, errors that make it into published articles likely indicate false positives or inflated effect sizes.

It is important to note that the issues caused by scientific misconduct, QRPs, and errors are not isolated to the specific articles in which they occur. Publication bias (PB) occurs when the published literature on a topic is not representative of all existing evidence related to that topic (Banks et al., 2015; Kepes et al., 2012). Thus, PB occurs when researchers suppress specific hypotheses that are unsupported or choose not to publish a study at all because it did not "work" (i.e., putting these results into their *file-drawer*). While it is possible that suppressed results are suppressed for good reason (e.g., there was not enough statistical power to detect the hypothesized effect), it is notable that unsupported and supported hypotheses are typically not suppressed at the same rate (see e.g., Kepes et al., 2022). Thus, even among underpowered studies, which are very common in psychology and management (e.g., Maxwell, 2004; Paterson et al., 2016), supported hypotheses are still more likely to be published as compared to unsupported hypotheses. Ultimately, when meta-analysts attempt to summarize the literature, these unpublished findings are more difficult to locate and often end up not being included. This suppression of non-significant effect sizes, coupled with the inclusion of effect sizes that are affected by misconduct, QRPs, and errors, likely results in misleading (typically inflated) meta-analytic mean effect size estimates. It is possible that effect sizes that are influenced by misconduct, QRPs, and/or errors represent outliers in a meta-analytic dataset and, therefore, could be identified and removed. Unfortunately, many meta-analyses do not include comprehensive sensitivity analyses, such as PB or outlier analyses (Aguinis et al., 2011; Banks et al., 2012; Siegel et al., 2021). Therefore, the impact of PB and outliers on a particular meta-analysis may be unknown. However, large scale sensitivity analyses in the field of IO psychology clearly suggest that PB likely influences the trustworthiness of meta-analyses in our field. For instance, depending on the method used to detect bias, Siegel et al. (2021) found evidence of non-negligible bias in between 60 and 70 % and 70–80 % of meta-analytic datasets examined. Thus, one can conclude that roughly 70 % of meta-analytic datasets examined contained non-negligible bias. When meta-analytic results are significantly impacted by PB or outliers, their conclusions are likely to be misleading. This is particularly troublesome as meta-analyses often have a great impact on cumulative knowledge in a particular area (Borenstein et al., 2009; Kepes et al., 2013). This impact, in turn, affects future research agendas and funding (Ioannidis, 2012), and contributes to the widening of the science-practice gap (Kepes et al., 2014).

5. Why do misconduct, QRPs, and errors occur?

The likelihood of individuals engaging in scientific misconduct or QRPs and making errors may be influenced by various internal and external factors (Hoole, 2019). First, some personal characteristics may increase the likelihood that an individual engages in misconduct and/or QRPs or makes errors. Specifically, ample evidence shows that conscientiousness, agreeableness, and morality are negatively related to unethical behaviors, such as academic dishonesty, while impulsivity and narcissism are linked to higher instances of these behaviors (Lee et al., 2020). Furthermore, highly conscientious individuals may be less likely to make errors (e.g., accidentally mis-specifying a model or incorrectly copying results output) as they tend to be detail-oriented (Fong & Tosi, 2007). Taken together, individuals with these characteristics may be more resistant, or particularly susceptible, to the system-wide factors discussed next.

The second factor that influences misconduct, QRPs, and errors is that the training many receive in psychology graduate programs may be insufficient, particularly when it comes to advanced statistical methods (Aiken et al., 2008; Tonidandel et al., 2014), sampling issues (Fisher & Sandell, 2015) and research ethics (e.g., Byrne et al., 2014; Swift et al., 2022). Errors may occur due to lack of knowledge. For instance, a researcher may be unaware that they are choosing an inappropriate statistical test, conducting a statistical analysis incorrectly, or reporting and/or interpreting results incorrectly (Hardwicke et al., 2019). Indeed, studies have shown that many psychology students and researchers misunderstand core concepts in research methodology and statistics, including the assumptions associated with regression analyses (Ernst & Albers, 2017), how to interpret *p*-values (e.g., Gigerenzer, 2018), and the importance of statistical power (e.g., Gigerenzer, 2018). Additionally, due to technological advances in statistics software, researchers do not need to fully understand statistical analyses to run them (Cortina, Aguinis et al., 2017), making it more likely mistakes will be made. Furthermore, HARKing and related QRPs have actually been advocated by some well-known and influential researchers and publications (e.g., Bem, 1987; Dane, 1990; see more detailed discussions in, e.g., Kerr, 1998; Leung, 2011; Maxwell, 2004). Thus, some research methods training may directly encourage the use of QRPs. Relatedly, although many programs include *formal* training related to research ethics and students state that they feel prepared to behave ethically (Fisher et al., 2009), graduate students are more likely to engage in QRPs when their mentors do (Swift et al., 2022). Given how common QRP engagement seems to be among IO and management researchers (e.g., Banks et al., 2016; Bedeian et al., 2010), it is reasonable to suspect that the *informal* training many students receive also supports QRP engagement. This informal training is especially harmful, because it may not only “serve to reinforce or undermine what is taught in the classroom” (Swift et al., 2022, p. 21), but also because it can influence subsequent students by passing down questionable behavior from one generation to another.

Third, researchers operate in a system that emphasizes the importance of publishing, particularly in top journals, for tenure, promotion, individual prestige, and academic rankings (Ball, 2005; Gomez-Mejia & Balkin, 1992; Nosek et al., 2012; Ostriker et al., 2009; Podsakoff et al., 2008). Given that journals, especially top journals, prefer to publish articles with statistically significant results (e.g., Kepes & McDaniel, 2013), and given that “environment and psychological processes can lead us to engage in ethically questionable behavior even if it violates our own values” (Bazerman, 2020, p. 93), researchers are motivated to engage in a variety of behaviors to increase the likelihood of publication, including misconduct and QRPs. Supporting this idea, Diederik Stapel stated the following after his extensive fraud was uncovered: “In the past years the pressure became too much for me. I have not been able to withstand the pressure to score, to publish, to be better and better” (Stapel, 2011, cited in Stroebe et al., 2012). However, it is also reasonable that we behave unethically in these situations out of self-interest but might be unaware of our unethical behavior (i.e., *motivated blindness;* Bazerman, 2020) and the damage it may cause to science. Furthermore, when individuals are under significant pressure, their performance is likely to suffer (Lepine et al., 2005), which could result in them making more errors.

Lastly, there is ample opportunity to engage in misconduct and QRPs, as these behaviors, along with errors, are unlikely to be caught. Specifically, it is often difficult to verify findings during the peer-review process or post-publication because data sharing does not yet appear to be the norm (e.g., Gabelica et al., 2022; Tenopir et al., 2015; Wicherts et al., 2006). Indeed, in one study, psychologists appeared to be less willing to share their data than researchers in many of the hard sciences, such as biology and engineering (Tenopir et al., 2015). Furthermore, many have identified problems with the peer-review process (see e.g., Hardwicke et al., 2019; Miller, 2006; Suls & Martin, 2009), including poor inter-rater agreement in reviews (see e.g., Starbuck, 2005). As noted earlier, many reviewers may be insufficiently prepared to critique the statistical techniques used in articles, and thus, unlikely to identify problematic issues or even errors (Hardwicke et al., 2019). Furthermore, reviewers and editors sometimes encourage engagement in QRPs, such as dropping unsupported hypotheses (e.g., Rupp, 2011). To provide a particularly extreme example of potential issues with the peer-review process, Hindawi and Wiley (who owns Hindawi) announced plans to retract more than 1200 articles that appear to have been published as part of peer-review rings, where authors, reviewers, and editors conspired to publish papers without adequate peer review (e.g., Retraction Watch, 2023). Although such peer-review rings are (hopefully) rare, this scandal highlights what could go wrong if we do not take the peer review process seriously.

Importantly, our most prestigious, or supposedly “high-quality,” journals do not seem to be immune from misconduct, QRPs, and errors. For instance, in a sample of IO and management journals, Kepes et al. (2022), found that QRPs were more common in highly prestigious journals when compared to less prestigious ones. This is consistent with research in other fields, which has found that top journals tend to publish more statistically significant results than lower-tier journals (Eisend & Tarrahi, 2014; Murtaugh, 2002). Furthermore, in their analysis of retractions, Craig et al. (2020) found that almost a quarter of retractions (37/160; 23 %), including those issues for misconduct and errors, were at journals with impact factors greater than five. Thus, it is not clear that articles published in our most prestigious journals are any more credible or trustworthy than articles published in journals that are not commonly considered top journals.

6. A path forward

It is likely not possible to address every potential reason why researchers may engage in misconduct and QRPs and make errors. For instance, as a field, we are probably not going to start assessing individuals' morality prior to admitting them to graduate programs. However, there are several changes that we can make, which should help reduce the frequency and impact of misconduct, QRPs, and errors. Specifically, we can improve training and address the incentive structure in academia by, for example, using open science practices (OSPs). Below, we expand on each of these potential remedies and, where they exist, provide examples of positives changes that are already being made.

6.1. *Improving the quality of training*

In the field of psychology as a whole, and thus likely in IO as well, current training seems to be inadequate to ensure that researchers grasp important statistical concepts and new techniques (e.g., Gigerenzer, 2018; Tonidandel et al., 2014) as well as minimize researcher misconduct and QRPs (e.g., Byrne et al., 2014; Swift et al., 2022). Therefore, more extensive training and explicit discussions of inappropriate researcher behaviors, including misconduct and QRPs, during training for (under)graduates, doctoral students, and other researchers, is needed (Swift et al., 2022). Such information could, for instance, be provided in courses on ethics, research methods, and statistics, or be taught directly in the labs (Grand, Rogelberg, Allen, et al., 2018). Education about replicability and reproducibility might (at least temporally) shift undergraduates' attitudes towards these behaviors (e.g., Chopik et al., 2018). Similarly, Sacco and Brown (2019) tested the efficacy of an intervention to reduce the acceptance of QRPs in psychology graduate students from different graduate programs and found promising evidence, although additional training and education are necessary for long-term benefits. Furthermore, researchers (both in school and after obtaining their Ph.D.) should be encouraged to continue their professional education in statistics and research methods through programs such as the Consortium for the Advancement of Research Methods and Analysis (CARMA; https://carmattu.com). By receiving training on methods that they may be unfamiliar with, researchers can reduce the likelihood of making errors *and* improve their ability to identify problematic issues in papers when acting as a peer reviewer.

6.2. *Re-aligning the current incentive structure*

Even if future researchers receive better training, it is not likely to lead to the extinction of misconduct and QRPs until the academic incentive structure changes as well. As previously stated, individuals engage in misconduct and QRPs because, among other reasons, they face pressure to publish statistically significant results, and feel they can engage in these behaviors without getting caught. The retraction case of Hart (2013), from the IO related field of social psychology, illustrates this clearly. The graduate student who was responsible for the data noticed that he could achieve statistically significant results if he would fake some data, so he duplicated some cases while deleting others. As consequence, he "recognized that the system rewarded him, and his collaborators, not for interesting research questions, or sound methodology, but for significant results. When he showed his collaborators the findings, they were happy with them—and happy with [him]" (Tullett, cited in McCook, 2017). Furthermore, the more senior researchers did not question the student's data: "Hindsight's a b*tch…I wish we had treated our data with the skepticism of someone who was trying to determine whether they were fabricated, but instead we looked at them with the uncritical eye of scientists whose hypotheses were supported" (Tullett, cited in McCook, 2017). As Tullett (cited in McCook, 2017) sums up: "the incentive structures in the field are problematic … this is an extreme case of what the consequences of that can be." In other words, "the reward system in academia may be rewarding A (the use of QRPs) while hoping for B (the use of scientifically sound and rigorous processes and procedures) (Kerr, 1998)" (Kepes et al., 2022, p. 1192; see also Spector, 2022).

To address this, we need more system-wide change to re-align the incentive structure in academia. This could be achieved, for example, with changes to the reward system, including the tenure and promotion processes, more tolerance for the publication of "null" (i.e., statistically non-significant) results, a change in the current review process, and the broad and field-wide use of OSPs.

6.3. *Changing the reward system*

The promotion and the tenure processes are still strongly determined by the number of publications, especially in highly prestigious journals (Aguinis et al., 2020; Podsakoff et al., 2008). Therefore, it is not surprising that tenure status is associated with QRPs. For instance, in the field of management, Kepes et al. (2022) found that QRPs were more likely in early stages after dissertations were completed (when people are typically tenure-seeking) than in later stages. This suggests that changes in the tenure process may reduce the motivation to engage in misconduct and QRPs. Spector (2022) explains misconduct and QRPs as a product of perceived pressure (e.g., a need for publication), opportunity (e.g., having the skills to use QRPs), and rationalization (e.g., a willingness to justify QRPs). Changing the reward system would help to address the pressure-element. For instance, we recommend that promotion and tenure committees widen their focus from assessing only research quantity (mainly in prestigious journals) to also focusing on the *quality* of publications (i.e., the quality of the specific article rather than just the journal where it was published). Especially in IO, findings may have a valuable impact beyond academia, even though not published in an A journal. Rather than striving to publish a large amount of research in the most prestigious journals, regardless of the quality or topic of the research, it may be more beneficial to demand the development of a coherent research program (i.e., becoming an expert in a specific area). This includes formulating interesting research questions, using scientifically sound and rigorous scientific methods, and submitting a reasonable number of high-quality studies to journals that provide the best fit for them (for more examples how to evaluate academic performance in IO, see Spector, 2022; Grand, Rogelberg, Allen, et al., 2018). In such a changed reward system, researchers would not need to be doing everything they can (e.g., misconduct, QRPs) to increase the chances their article will get published in a "prestigious" journal.

6.4. *Emphasizing null-results*

Emphasizing the importance of "null" results could also add to the realignment of our problematic reward structure. It would clearly reduce the motivation to engage in misconduct and use QRPs to find statistically significant results if more journals would encourage the publication of "null" findings. Despite what many believe, such findings from rigorous research cannot only set an example for the importance of thorough research (regardless of its results), they also offer benefits as long as these findings inform the field. For instance, statistically non-significant results may call into question the generalizability of a theory and, therefore, encourage the investigation of boundary conditions. Like Landis et al. (2014) put it, "we need a complete picture of our phenomena of interest to truly advance our scientific knowledge" (p. 164). Recognizing this, the *Journal of Business and Psychology* dedicated a special issue to null results. In their editorial, the editorial team illuminated the advantages of publishing null results and offered suggestions for editors, reviewers, and authors (Landis et al., 2014). Some journals, like *Meta-Psychology* or *Academy of Management Discoveries* also recognized the importance of null results and welcome studies with them or even "negative" ones on their websites. Unfortunately, these types of journals are often not viewed as top-tier. Thus, the motivation to publish in them is comparatively lower.

6.5. *Changing the review process*

The review process functions as a bottleneck for scientific discoveries; it determines which studies get published. Although it is the purpose of peer-review to "maintain the integrity of science by filtering out invalid or poor-quality articles" (Wiley, 2000-2023), and things would certainly be worse without *any* form of peer-review, the current process is not ideal. As previously mentioned, editors and reviewers are more likely to recommend publishing an article with statistically significant results (Fanelli, 2012). This tendency against publishing null results has a long history (Sterling, 1959; see Sterling et al., 1995 for a replication); therefore, editors and reviewers need to change behavior that was common for over 50 years. They also occasionally encourage authors to engage in QRPs (e.g., dropping some of their insignificant findings or to exclude certain conditions) "to streamline manuscripts" (Franco et al., 2016, p. 8), at least in psychology experiments. Thus, even if individual researchers are trained in research ethics and are willing to provide a more complete report of their study, they are unlikely to do so if they fear negative consequences (i.e., a rejection).

Following this, some researchers have argued that the aforementioned improved training and re-alignment of the current incentive structure is only achievable with a change in the review process (Nosek et al., 2012; Wagenmakers et al., 2012). One proposal to strengthen the review process is the use of results-blind review (RBR; see also the section on *Registered Reports*, below), where manuscripts are reviewed in two stages (Grand, Rogelberg, Banks, et al., 2018; Kepes & McDaniel, 2013). During Stage 1, a manuscript is submitted and reviewed without the results and discussion. With this format, a paper is judged by the quality of its research question, hypotheses, design, and methodological approach, not by the statistical significance-level of its findings. Thus, a paper with an important research question and rigorous methods is likely to get an 'in-principle-acceptance,' and will be published regardless of its results. During Stage 2, the whole manuscript, including results and discussion, is reviewed to ensure that the authors have followed their research plan (Grand, Rogelberg, Banks, et al., 2018; Kepes & McDaniel, 2013). RBR "reduces the impact of statistical significance on acceptance decisions" (Kepes et al., 2014, p. 459) and thus, should reduce the motivation of individual researchers to engage in scientific fraud or QRPs.

Some journals already offer RBRs. For instance, in 2016, a couple of IO psychology/management journals (e.g., *Journal of Business and Psychology*; *Leadership Quarterly*) joined forces for a joint initiative towards more reliable research. They announced the launch of an optional results-blind review submission option (see https://jbp.charlotte.edu/). Beyond that, there are journals, such as the *Journal of Management Scientific Reports*, that offer results-masked submissions (see https://smgmt.org/jomsr/). Allowing RBR is a valuable first step in changing the review process. However, to realize the change towards RBR as the new standard, more authors need to submit papers on this track and reviewers need explicit instructions (and more feedback from their action editors) on how to review a Stage 1 and Stage 2 manuscript and to learn about the underlying values of the two-stage process (i.e., a shift from outcome focus to process focus; Grand, Rogelberg, Banks, et al., 2018).

Although reviewing is a cornerstone in our profession (Köhler et al., 2020), there are currently no requirements to follow formal guidelines or standards, for peer review, in general, and for RBR, in particular. However, some first training for peer reviewing has been developed, including top tips for reviewers (Chambers & Tzavella, 2022), the competency framework for reviewers (Köhler et al., 2020), and the entrance test for editors by the Peer Community in Registered Reports (https://rr.peercommunityin.org/). In addition, numerous reporting guidelines and standards are available (e.g., Appelbaum et al., 2018; Levitt et al., 2018; Moher et al., 2009). It should be the responsibility of reviewers and editors to ensure that researchers adhere to these guidelines and standards. To aid in this important endeavor, specific recommendations for editors and reviewers could be developed, just as they have been for meta-analytic studies or particular methodological or statistical approaches (e.g., DeSimone et al., 2021; Kepes, Wang, Cortina, 2022).

A second suggestion to strengthen the review process by increasing accountability is open peer review (OPR), where some to all aspects of the peer review process are made publicly available. For instance, reviewers can agree to sign their reviews that can be published alongside published paper. Thus, OPR holds reviewers accountable for their comments while they are also able to get credit for their work. However, there is a debate about the pros (e.g., potentially higher quality reviews) and cons (e.g., potentially biased reviews) of OPR: Some studies found no effect of OPR on review quality, recommendation regarding publication, and time to review (Van Rooyen et al., 1999; Van Rooyen et al., 2010). Interestingly, they *did* find an increase in decline to review. However, others found some positive effects of OPR (e.g., Kowalczuk et al., 2013; Walsh et al., 2000). Taken together, the findings are rather ambiguous (and partly outdated). Thus, to give a valuable suggestion regarding the use of ORP, we might need more research.

6.6. Using open science practices (OSPs)

In response to the replication crisis in social psychology, the Transparency and Openness Promotion (TOP) guidelines (Nosek et al., 2015) introduced author guidelines for journals that are commonly known as OSPs. These practices have been developed as a tool to address the replication crisis, in particular, and the low reproducibility and replicability of research, in general. OSPs carry the promise to reduce QRPs and foster a transparent research culture. The TOP guidelines (Nosek et al., 2015) include eight standards: citation standards; transparency of data, analytic methods (code), research materials, and design and analysis; preregistration of studies and analysis plans; and replication studies as publishing format. There are also recommendations and preregistration templates for reviews and meta-analyses (Moreau & Gamble, 2022; Van den Akker et al., 2020). Based on these eight standards and two other OSPs, registered reports as publishing formats (Chambers & Tzavella, 2022) and availability of open science badges in a journal (Kidwell et al., 2016; see Table 2 for more information on OSPs[5]), the Center for Open Science (COS) developed the TOP factor (Center for Open Science, 2020). The TOP factor assesses the degree to which a journal adopts each OSP and, thus, evaluates the degree to which journals support transparency and reproducibility (Kepes et al., 2020; Nosek et al., 2015). Notably, some IO psychology journals have signed on to the TOP guidelines (e.g., *JAP*, *Journal of Business and Psychology, Journal of Organizational Behavior, Human Resource Management Review*; https://www.cos.io/initiatives/top-guidelines). However, how successful are those practices in achieving the goal of transparent and rigorous research? Fortunately, there is initial evidence for the success of OSPs in psychology (Hardwicke et al., 2018; Obels et al., 2020). To illuminate the use of OSPs and their benefits, we provide more detail and some examples in the following sections.

Table 2
Overview of open science practices.

Open science practices	Example articles for more information
Citation standards	Cobb et al. (2023)
Transparency of data, analytic methods (code), research materials, and design and analysis	Grahe (2021); Nosek et al. (2015)
Preregistration of studies and analysis plans	Wagenmakers and Dutilh (2016); Nosek et al. (2019)
Replication studies and crowdsourcing	Asendorpf et al. (2013); Brandt et al. (2014); Köhler and Cortina (2019); Landy et al. (2020); Moshontz et al. (2018); Schweinsberg et al. (2021); Simons (2014); Uhlmann et al. (2019)
Registered reports	Chambers and Tzavella (2022)
Open science badges	Kidwell et al. (2016)

6.6.1. Preregistrations

Preregistrations are one means to reduce the use of QRPs (Banks et al., 2019). When preregistering, researchers commit to a research plan and/or analysis plan *before* they conduct their study and collect data. This plan is then submitted to an online repository like the Open Science Framework (OSF). A preregistration can include, for example, hypotheses, dependent variables, conditions, analyses, exclusion criteria, and determination of sample size (cf. https://aspredicted.org/). The reasoning behind preregistrations is that they limit the use of "researcher degrees of freedom" (Simmons et al., 2011) and thus the "methodological flexibility" of researchers when analyzing data (Kepes & McDaniel, 2013), which should minimize the display of QRPs and, therefore, lead to more reproducible and replicable results and more trustworthy cumulative knowledge. For instance, HARKing (Kerr, 1998), the selective reporting of outcomes (John et al., 2012), *p*-hacking (i.e., "report only […] analyses […] that 'work';" Simonsohn et al., 2014, p. 534), and false positive results (Nosek et al., 2019) should be minimized. Indeed, evidence indicates that preregistrations can help to decrease false positive findings (i.e., Type I error rates) as there are fewer statistically significant results in preregistered samples than in non-preregistered ones (Toth et al., 2021). Further, they are associated with greater transparency and more rigorous reporting of study methods and analyses (Toth et al., 2021). A survey among scientists revealed that implementing preregistration is perceived to improve the quality of their projects and research process (Sarafoglou et al., 2022).

6.6.2. Registered reports

Another OSP that is closely related to preregistrations are Registered Reports (RRs; Chambers, 2019; Chambers & Tzavella, 2022). RRs are a publication format in which peer review is conducted before data are analyzed, like with the results blind review process. The main difference between RRs and RBRs is the timing of data collection (Grand, Rogelberg, Banks, et al., 2018). While there are no specifications for RBR, RRs always undergo the Stage 1 review process before any data collection. This allows authors, reviewers, and editors to discuss and optimize hypotheses, methods, data collection, and analyses.

Against commonly heard doubts, research has shown that RRs are not only equally novel and creative, but they are also likely to be of higher rigor and quality than other studies (Soderberg et al., 2021). This

[5] To the list of OSPs provided in Table 2, we also added a note on replication studies and crowdsourcing. Recently, the idea emerged to pool our limited resources to conduct so called crowdsourced multisite replication research (Moshontz et al., 2018; Uhlmann et al., 2019) to study the replicability and generalizability of effects with high statistical power. In IO psychology, Castille, Kreamer, et al. (2022) suggested the creation of a multisite replication project that brings together practitioners and researchers alike – "ManyOrgs" (p. 548). Additional examples of these crowdsourcing initiatives, as well as references to articles that provide general guidance on conducting replication studies, are also provided.

is noteworthy, however, not surprising, because the review process before data collection allows improving the research design (Van't Veer & Giner-Sorolla, 2016). Further, RRs might prevent selective reporting and the file-drawer problem (Sterling et al., 1995) because they increase the chances of publishing null findings (Allen & Mehler, 2019). Indeed, there is a higher balance of the ratio of supported/unsupported hypotheses in RRs (44 % positive results) than in the standard literature (96 % positive results; Scheel et al., 2021). Thus, RRs are one means to reduce PB and Type I error inflation (cf. Franco et al., 2016). Some reservations regarding RRs include the fear that they might decrease a journal's impact factor, because mixed (or negative) results may be viewed as less interesting and, thus, less likely to be reported by the media and cited by the research community. Preliminary evidence from a comparison of citations between published RRs and comparable articles from the same journals dispelled these fears. Hummer et al. (2017) found similar to slightly greater numbers of citations for RRs.

6.6.3. *Open science badges*

Journals that want to incentivize the use of OSPs can award open science badges to articles that use OSPs. Currently, there are three open science badges that can be rewarded for an article that (a) has been preregistered, and/or made (b) its data or (c) its materials publicly available (Center for Open Science, n.d.). In psychological journals that offer these badges, the frequency and quality of data and materials sharing increased when compared to journals without such badges, the data and materials were not only more accessible and complete but also more correct and usable (Kidwell et al., 2016). In addition, Schneider et al. (2022) found that the use of badges increases the perceived trust in scientists by student teachers and social scientists. Badges can also further reduce the epistemic beliefs that scientific knowledge are subjective opinions. Thus, open science badges function as a signal for new norms in a field (Center for Open Science, n.d.) and may help science to be seen as objective and data-driven (instead of merely an opinion; Schneider et al., 2022). Taken together, open science badges are a simple and cost-effective method to foster the use of OSPs (Kidwell et al., 2016).

6.6.4. *Open science practices and IO psychology*

As OSPs have their origin in social psychology, some people argue that OSPs are not as suitable for IO psychology (Guzzo et al., 2022). Indeed, when talking about the implementation of OSPs in our field, IO specific characteristics need to be considered. Here, we discuss some of these specifics and illuminate whether they are compatible with OSPs. In particular, we discuss how OSPs can go hand in hand with (a) qualitative and inductive research, (b) the handling of sensitive data, and (c) studies in organizations or with other practitioners in the field.

First, Torka et al. (2023) identified a methods bias among editors from IO and management. When the editors were asked about their reasons to not include OSPs on their websites, they indicated a perceived inadequacy of available OSPs for certain research approaches (e.g., qualitative or inductive research). Unfortunately, some individual researchers also share (see Study 2 from Toth et al., 2021) and spread (Guzzo et al., 2022) this perspective. However, there is no need to fear that OSPs like preregistrations would devalue qualitative research over quantitative research or that OSPs will result in "methodological sameness" (Guzzo et al., 2022, p. 24). Instead, by being transparent with testing or developing hypotheses, and with which findings are (in-)consistent with expectations, using OSPs can build trust in inductive and exploratory research (Hüffmeier et al., 2022). Thus, preregistering a study is beneficial for all kinds of IO research (see also Torka et al., 2023) and, luckily, there are also preregistration templates available for qualitative research (e.g., Haven et al., 2020; Kern & Gleditsch, 2017) and preexisting data (Mertens & Krypotos, 2019).

Second, in IO, we use data from organizations more regularly than other related fields. Indeed, those data are often very sensitive and not as easy to share with others. For instance, the data may not be anonymized (e.g., HR data about employees) or crucial for the success of a company (e.g., financial data). However, OSPs like data sharing are never forcing researchers to disclose any sensitive data or to harm a company. On a side note, there is also no movement in the direction of full data sharing in IO (see results from Hüffmeier et al., 2022 and Torka et al., 2023). Moreover, even if the data are sensitive, there are still different ways to enact the philosophy of accessibility. For instance, personal identifiers (e.g., demographics) could be excluded from the dataset before making it available to others. Researchers could also only share data relevant to reproduce their analyses (e.g., descriptive statistics, intercorrelations, reliability estimates) instead of sharing individual-level data (Banks et al., 2019). Going even further, it is possible to simulate a dataset with similar statistical characteristics as the original dataset. "This retains the statistical properties of the original data, allowing other researchers to run confirmation analyses if desired, but protects individuals' responses and characteristics" (Morgan et al., 2022, p. 540)[6]. Finally, even if researchers are not allowed to share *any* data, they can still share relevant materials or analysis codes (see Transparency of data, analytic methods, research materials, and design and analysis, Table 2). Thus, using company data should not be an exclusion criterion for sharing data per se.

Besides data being sensitive, there are other challenges when working with companies or other practitioners in the field. Companies often pursue their own goals by participating in a study (e.g., collecting data for their HR department) and researchers have to include their wishes in the research process. It is also not unlikely that they request changes to the research plan while the study is already running. Hence, writing a preregistration where you should, at least ideally, specify in detail every part of your study in advance, can be challenging in applied settings. The good news is that deviations from preregistration are allowed and common (Claesen et al., 2021). It is only important to be transparent about what changed and why. Moreover, preregistering studies regularly may also help IO researchers to anticipate and make plans for necessary changes throughout the research process (Toth et al., 2021).

6.6.5. *Implementation of OSPs*

In general, OSPs can increase the perceived trustworthiness of research (Methner et al., 2022; Rosman et al., 2022; Schneider et al., 2022) and may lead to more trustworthy cumulative knowledge. To address (some of) the causes of untrustworthy cumulative knowledge, it would be desirable for researchers to use OSPs, and for journals to adopt them in their policies. Unfortunately, there is currently an implementation gap in OSPs: researchers often do not use OSPs (Aguinis et al., 2020; Ferguson, 2015; Tenney et al., 2021) and there are some reservations about them (Guzzo et al., 2022; Woznyj et al., 2018). As one possible explanation, journal policies may help to explain why individual researchers do not implement OSPs and still engage in misconduct and QRPs. If journal guidelines do not encourage the use of OSPs, individual researchers may see little reason to use them. Indeed, research on the implementation of OSPs in journals' policies found that these practices were hardly mentioned by IO and management journals (Torka et al., 2023; see also Feeney, 2018). A survey of editors further revealed some barriers to the adoption of OSPs. Journal editors named the perceived lower suitability of OSPs for qualitative research (see 6.6.4), missing authority, and missing familiarity with OSPs as a cause for the implementation gap (Torka et al., 2023).

Journals seem to be the key player for the broad implementation of OSPs. As mentioned before, the implementation of open science badges by a journal can increase data and materials sharing (Kidwell et al., 2016). However, there are also boundaries of these journal policies. For example, even with the requirement of data availability statements,

[6] These simulated/synthetic datasets can be produced using R packages such as 'synthpop' (Nowok et al., 2016).

many authors in the medical sciences are unwilling to share their data (Gabelica et al., 2022). It is also important to note that there are currently no incentives for journals to require the use of OSPs. In the medical sciences, preregistration requirements by journals led to increases in preregistration only after requirements were paired with fines for not preregistering clinical trials (De Angelis et al., 2004; Dickersin & Rennie, 2012; Laine et al., 2007). Thus, it appears that journal policies offer the possibility to make a substantial difference regarding the adoption and expansion of OSPs. However, it is important to explore exactly *how* OSPs need to be implemented in IO to be most effective (e.g., providing badges for open data is preferable to data availability statements; incentives for preregistering may complement requirements).

7. Conclusion

In this paper, we discussed the current state of affairs in IO psychology, and related fields such as social psychology and management, that affects the trustworthiness of our field's cumulative knowledge. Overall, the evidence regarding the reproducibility and replicability of research in our field is rather discouraging. Scientific findings are often neither reproducible (Artner et al., 2021; Bakker & Wicherts, 2011; Bergh et al., 2017; Hardwicke et al., 2018) nor replicable (Ebersole et al., 2016; Klein et al., 2022; Open Science Collaboration, 2015). Additionally, only a small number of studies deal with constructively replicating published effects (Köhler & Cortina, 2019; Makel et al., 2012). As a result, there is reason to doubt the credibility and self-correcting-ability (von Hippel, 2022) in our field. Thus, we discussed the causes of the replication crisis and untrustworthiness of our cumulative knowledge, as well as potential remedies to create a more credible scientific discipline – some of which (e.g., open science badges, preregistration) are already being adopted by some journals in our field.

Fig. 1 summarizes the internal and external factors that threaten our field's trustworthiness. On the right side of Box 1, we highlight some internal factors, such as personal characteristics, that can raise an individual's probability to engage in destructive behaviors (Fong & Tosi, 2007; Lee et al., 2020). Further, researchers are affected by external factors, like insufficient training (i.e., formal and informal), a reward system that incentivizes research-quantity, journals preferring statistically significant results, and the opportunity to engage in behaviors that increase the likelihood of publishing as many studies as possible (e.g., missing norms for data sharing). These factors, in turn, lead to a variety of researcher behaviors (Box 3), such as errors (e.g., inconsistency between reported degrees of freedom and described CFA and SEM models) and the use of QRPs (e.g., selectively reporting hypotheses) and outright scientific fraud (e.g., fabricating data). These types of behaviors then in turn adversely affect the trustworthiness of our cumulative scientific knowledge (Box 4) as they result in, for example, false positive results, low replicability, and publication bias. Thus, it currently remains uncertain what conclusion can be drawn from some of the findings in IO psychology.

We also outlined suggestions for addressing the harmful external factors (see Fig. 1). The left side of Box 1 contains a variety of proposed changes that would combat the current external factors identified on the right side of the box. One of these recommendations involves enhancing the quality of both formal and informal scientific training, with a particular focus on topics related to replicability and reproducibility. Additionally, it would be beneficial to implement a reward system that promotes research-quality (instead of research-quantity) and journals that publish null results, as well as to decreases the opportunity to engage in misconduct, QRPs, and errors. In the long run, these changes could mitigate the publication of false positive results and PB. Finally, it should be noted that OSPs (Box 2) offer the potential to moderate the effects of internal and external factors on researcher behaviors. By promoting and implementing OSPs, there is a possibility to reduce the use of misconduct and QRPs. This, in turn, can address some of the existing problems in our field.

Although the focus of this paper was on system-wide factors that could help address the credibility crisis in IO psychology (see also Hoole, 2019), this is not to say that individual researchers cannot act prior to

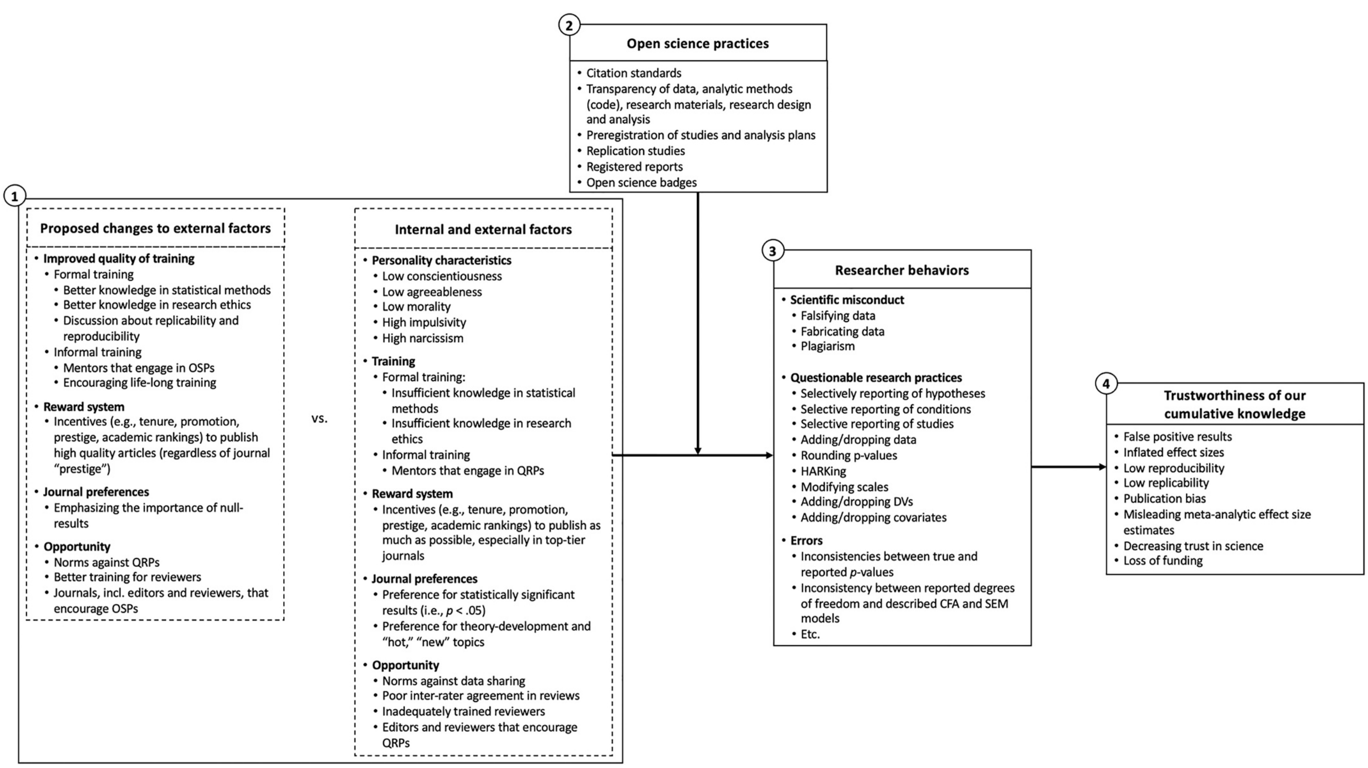

Fig. 1. Theoretical model: reasons for and examples of misconduct, QRPs, and errors.

systematic adoption of suggested changes by academic institutions and journals. Indeed, as noted by Castille, Köhler, et al. (2022), if each of us take small steps towards embracing open science principles, over time, we can change the field. Therefore, we recommend that all researchers (1) continue their professional education in research methods and statistics to reduce their likelihood of making errors, as well as increasing the likelihood of catching others' errors during the review process; (2) be as transparent as possible by, for instance, making data and code available and explaining the reasoning for various decisions made throughout the study design, analysis, and reporting process; and, (3) take advantage of alternative publishing avenues, such as results blind review. For more recommendations for different stakeholders in IO, see Table 2 in Grand, Rogelberg, Allen, et al. (2018).

In conclusion, we hope that our paper sparks debate surrounding the best ways to implement system-wide change and encourages researchers and journals to consider the adoption of OSPs, like preregistration, RRs and open science badges, as a viable path forward towards a trustworthy cumulative knowledge.

Declaration of competing interest

None.

Data availability

No data was used for the research described in the article.

References

Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PLoS One, 12*(3), Article e0172792. https://doi.org/10.1371/journal.pone.0172792

Aguinis, H., Banks, G. C., Rogelberg, S. G., & Cascio, W. F. (2020). Actionable recommendations for narrowing the science-practice gap in open science. *Organizational Behavior and Human Decision Processes, 158*, 27–35. https://doi.org/10.1016/j.obhdp.2020.02.007

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. https://doi.org/10.1177/0149206310377113

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What you see is what you get? Enhancing methodological transparency in management research. *Academy of Management Annals, 12*(1), 83–110. https://doi.org/10.5465/annals.2016.0011

Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist, 63*(1), 32–50. https://doi.org/10.1037/0003-066X.63.1.32

Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS Biology, 17(5), Article e3000246*. https://doi.org/10.1371/journal.pbio.3000246

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*, 3–25. https://doi.org/10.1037/amp0000191

Artner, R., Verliefde, T., Steegen, S., Gomes, S., Traets, F., Tuerlinckx, F., & Vanpaemel, W. (2021). The reproducibility of statistical results in psychological research: An investigation using unpublished raw data. *Psychological Methods, 26*(5), 527–546. https://doi.org/10.1037/met0000365

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., … Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*(2), 108–119. https://doi.org/10.1002/per.1919

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods, 43*(3), 666–678. https://doi.org/10.3758/s13428-011-0089-5

Ball, P. (2005). Index aims for fair ranking of scientists. *Nature, 436*(900). https://doi.org/10.1038/436900a

Banks, G. C., Field, J. G., Oswald, F. L., O'Boyle, E. H., Landis, R. S., Rupp, D. E., & Rogelberg, S. G. (2019). Answers to 18 questions about open science practices. *Journal of Business and Psychology, 34*, 257–270. https://doi.org/10.1007/s10869-018-9547-8

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*(2), 182–196. https://doi.org/10.1111/j.1468-2389.2012.00591.x

Banks, G. C., Kepes, S., & McDaniel, M. A. (2015). Publication bias: Understanding the myths concerning threats to the advancement of science. In C. E. Lance, & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (pp. 36–64). Routledge/Taylor & Francis Group. https://psycnet.apa.org/record/2015-01072-002.

Banks, G. C., O'Boyle, E. H., Jr., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., … Adkins, C. L. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management, 42*(1), 5–20. https://doi.org/10.1177/0149206315619011

Bazerman, M. H. (2020). A new model for ethical leadership. *Harvard Business Review, 98*(5), 90–97. https://hbr.org/2020/09/a-new-model-for-ethical-leadership.

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*(4), 715–725. https://journals.aom.org/doi/10.5465/amle.9.4.zqr715.

Bem, D. J. (1987). Writing the empirical journal article. In M. P. Zanna, & J. M. Darley (Eds.), *The compleat academic: A practical guide for the beginning social scientist* (pp. 171–201). Random House. https://psychology.yale.edu/sites/default/files/bemempirical.pdf.

Bergeat, D., Lombard, N., Gasmi, A., Le Floch, B., & Naudet, F. (2022). Data sharing and reanalyses among randomized clinical trials published in surgical journals before and after adoption of a data availability and reproducibility policy. *JAMA Network Open, 5*(6), Article e2215209. https://doi.org/10.1001/jamanetworkopen.2022.15209

Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. (2017). Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization, 15*(3), 423–436. https://doi.org/10.1177/1476127017701076

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. https://onlinelibrary.wiley.com/doi/book/10.1002/9780470743386.

Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. (2016). HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology, 69*(3), 709–750. https://doi.org/10.1111/peps.12111

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., … van 't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology, 50*, 217–224. https://doi.org/10.1016/j.jesp.2013.10.005

Broad, W. J., & Wade, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science*. Simon & Schuster.

Broomell, S. B., & Kane, P. B. (2017). Public perception and communication of scientific uncertainty. *Journal of Experimental Psychology: General, 146*(2), 286–304. https://doi.org/10.1037/xge0000260

Byington, E. K., & Felps, W. (2017). Solutions to the credibility crisis in management science. *Academy of Management Learning & Education, 16*(1), 142–162. https://doi.org/10.5465/amle.2015.0035

Byrne, Z., Hayes, T., Mort McPhail, S., Hakel, M., Cortina, J., & McHenry, J. (2014). Educating Industrial–Organizational Psychologists for science and practice: Where do we go from here? *Industrial and Organizational Psychology, 7*(1), 2–14. https://doi.org/10.1111/iops.12095

Callaway, E. (2011). Report finds massive fraud at Dutch universities. *Nature, 479*(7371), 15. https://doi.org/10.1038/479015a

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

Castille, C. M., Köhler, T., & O'Boyle, E. H. (2022). A brighter vision of the potential of open science for benefiting practice: A ManyOrgs proposal. *Industrial and Organizational Psychology, 15*(4), 546–550. https://doi.org/10.1017/iop.2022.70

Castille, C. M., Kreamer, L. M., Albritton, B. H., Banks, G. C., & Rogelberg, S. G. (2022). The open science challenge: Adopt one practice that enacts widely shared values. *Journal of Business and Psychology, 37*(3), 459–467. https://doi.org/10.1007/s10869-022-09806-2

Center for Open Science. (2020). *New measure rates quality of research journals' policies to promote transparency and reproducibility*. COS. https://www.cos.io/about/news/new-measure-rates-quality-research-journals-policies-promote-transparency-and-reproducibility.

Center for Open Science (n.d.). Open Science Badges enhance openness, a core value of scientific practice. COS. https://www.cos.io/initiatives/badges.

Chambers, C. (2019). What's next for registered reports? *Nature, 573*(7773), 187–189. https://doi.org/10.1038/d41586-019-02674-6

Chambers, C. D., & Tzavella, L. (2022). The past, present and future of Registered Reports. *Nature Human Behaviour, 6*(1), 29–42. https://doi.org/10.1038/s41562-021-01193-7

Chang, A., .C., & Li, P. (2022). Is economics research replicable? Sixty published articles from thirteen journals say 'usually not'. *Critical Finance Review, 11*, 185–206. https://doi.org/10.1561/104.00000053

Chopik, W. J., Bremner, R. H., Defever, A. M., & Keller, V. N. (2018). How (and whether) to teach undergraduates about the replication crisis in psychological science. *Teaching of Psychology, 45*(2), 158–163. https://doi.org/10.1177/0098628318762900

Claesen, A., Gomes, S., Tuerlinckx, F., & Vanpaemel, W. (2021). Comparing dream to reality: An assessment of adherence of the first generation of preregistered studies. *Royal Society Open Science, 8*(10), Article 211037. https://doi.org/10.1098/rsos.211037

Cobb, C. L., Crumly, B., Montero-Zamora, P., Schwartz, S. J., & Martínez, C. R., Jr. (2023). The problem of miscitation in psychological science: Righting the ship. *American Psychologist. Advance online publication*. https://doi.org/10.1037/amp0001138

Cortina, J. M., Aguinis, H., & DeShon, R. P. (2017). Twilight of dawn or of evening? A century of research methods in the Journal of Applied Psychology. *Journal of Applied Psychology, 102*, 274–290. https://doi.org/10.1037/apl0000163

Cortina, J. M., Green, J. P., Keeler, K. R., & Vandenberg, R. J. (2017). Degrees of freedom in SEM: Are we testing the models that we claim to test? *Organizational Research Methods, 20*(3), 350–378. https://doi.org/10.1177/1094428116676345

Craig, R., Cox, A., Tourish, D., & Thorpe, A. (2020). Using retracted journal articles in psychology to understand research misconduct in the Social Sciences: What is to be done? *Research Policy, 49*(4). https://doi.org/10.1016/j.respol.2020.103930

Credé, M., & Harms, P. D. (2015). 25 years of higher-order confirmatory factor analysis in the organizational sciences: A critical review and development of reporting recommendations. *Journal of Organizational Behavior, 36*(6), 845–872. https://doi.org/10.1002/job.2008

Dane, F. C. (1990). *Research methods*. Thomson Brooks/Cole. https://psycnet.apa.org/record/1990-97456-000.

De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., … Van Der Weyden, M. B. (2004). Clinical trial registration: A statement from the International Committee of Medical Journal Editors. *New England Journal of Medicine, 351*(12), 1250–1251. https://doi.org/10.1056/NEJMe048225

DeSimone, J. A., Brannick, M. T., O'Boyle, E. H., & Ryu, J. W. (2021). Recommendations for reviewing meta-analyses in organizational research. *Organizational Research Methods, 24*(4), 694–717. https://doi.org/10.1177/1094428120967089

Dickersin, K., & Rennie, D. (2012). The evolution of trial registries and their use to assess the clinical trial enterprise. *Journal of the American Medical Association, 307*(17), 1861–1864. https://doi.org/10.1001/jama.2012.4230

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00621

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

Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D. J., Buttrick, N. R., Chartier, C. R., … Szecsi, P. (2020). Many Labs 5: Testing pre-data-collection peer review as an intervention to increase replicability. *Advances in Methods and Practices in Psychological Science, 3*(3), 309–331. https://doi.org/10.1177/251524592095

Efendic, E., & Van Zyl, L. E. (2019). On reproducibility and replicability: Arguing for open science practices and methodological improvements at the South African Journal of Industrial Psychology. *SA Journal of Industrial Psychology, 45*(1), 1–10. https://doi.org/10.4102/sajip.v45i0.1607

Eisend, M., & Tarrahi, F. (2014). Meta-analysis selection bias in marketing research. *International Journal of Research in Marketing, 31*(3), 317–326. https://doi.org/10.1016/j.ijresmar.2014.03.006

Ernst, A. F., & Albers, C. J. (2017). Regression assumptions in clinical psychology research practice—A systematic review of common misconceptions. *PeerJ, 5*, Article e3323. https://doi.org/10.7717/peerj.3323

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS One, 5*, Article e10068. https://doi.org/10.1371/journal.pone.0010068

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90*(3), 891–904. https://doi.org/10.1007/s11192-011-0494-7

Feeney, J. (2018). Robust science: A review of journal practices in industrial-organizational psychology. *Industrial and Organizational Psychology, 11*(1), 48–54. https://doi.org/10.1017/iop.2017.84

Ferguson, C. J. (2015). "Everybody knows psychology is not a real science": Public perceptions of psychology and how we can improve our relationship with policymakers, the scientific community, and the general public. *American Psychologist, 70*(6), 527–542. https://doi.org/10.1037/a0039405

Fisher, C. B., Fried, A. L., & Feldman, L. G. (2009). Graduate socialization in the responsible conduct of research: A national survey on the research ethics training experiences of psychology doctoral students. *Ethics & Behavior, 19*(6), 496–518. https://doi.org/10.1080/10508420903275283

Fisher, G., & Sandell, K. (2015). Sampling in industrial–organizational psychology research: Now what? *Industrial and Organizational Psychology, 8*(2), 232–237. https://doi.org/10.1017/iop.2015.31

Fong, E. A., & Tosi, H. L. (2007). Effort, performance, and conscientiousness: An agency theory perspective. *Journal of Management, 33*(2), 161–179. https://doi.org/10.1177/0149206306298658

Franco, A., Malhotra, N., & Simonovits, G. (2016). Underreporting in psychology experiments: Evidence from a study registry. *Social Psychological and Personality Science, 7*(1), 8–12. https://doi.org/10.1177/1948550615598377

Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: A mixed-methods study. *Journal of Clinical Epidemiology, 150*, 33–41. https://doi.org/10.1016/j.jclinepi.2022.05.019

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science, 1*(2), 198–218. https://doi.org/10.1177/2515245918771329

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

Goldfarb, B., & King, A. A. (2016). Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal, 37*(1), 167–176. https://doi.org/10.1002/smj.2459

Gomez-Mejia, L. R., & Balkin, D. B. (1992). Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal, 35*(5), 921–955. https://doi.org/10.2307/256535

Götz, M., O'Boyle, E. H., Gonzalez-Mulé, E., Banks, G. C., & Bollmann, S. S. (2021). The "Goldilocks Zone": (Too) many confidence intervals in tests of mediation just exclude zero. *Psychological Bulletin, 147*(1), 95–114. https://doi.org/10.1037/bul0000315

Grahe, J. (2021). The necessity of data transparency to publish. *The Journal of Social Psychology, 161*(1), 1–4. https://doi.org/10.1080/00224545.2020.1847950

Grand, J. A., Rogelberg, S. G., Allen, T. D., Landis, R. S., Reynolds, D. H., Scott, J. C., … Truxillo, D. M. (2018). A systems-based approach to fostering Robust Science in Industrial-Organizational Psychology. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 11*(1), 4–42. https://doi.org/10.1017/iop.2017.55

Grand, J. A., Rogelberg, S. G., Banks, G. C., Landis, R. S., & Tonidandel, S. (2018). From outcome to process focus: Fostering a more robust psychological science through registered reports and results-blind reviewing. *Perspectives on Psychological Science, 13*(4), 448–456. https://doi.org/10.1177/1745691618767883

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

Guzzo, R. A., Schneider, B., & Nalbantian, H. R. (2022). Open science, closed doors: The perils and potential of open science for research in practice. *Industrial and Organizational Psychology, 15*(4), 495–515. https://doi.org/10.1017/iop.2022.61

Hardwicke, T. E., Frank, M. C., Vazire, S., & Goodman, S. N. (2019). Should psychology journals adopt specialized statistical review. *Advances in Methods and Practices in Psychological Science, 2*, 240–249. https://doi.org/10.1177/2515245919858428

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., … Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science, 5*(8). https://doi.org/10.1098/rsos.180448

Harms, P. D., Credé, M., & DeSimone, J. A. (2018). The last line of defense: Corrigenda and retractions. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 11*(1), 61–65. https://doi.org/10.1017/iop.2017.86

Hart, W. (2013). RETRACTED: Unlocking past emotion: Verb use affects mood and happiness. *Psychological Science, 24*(1), 19–26. https://doi.org/10.1177/0956797612446351

Hartgerink, C. H., van Aert, R. C., Nuijten, M. B., Wicherts, J. M., & van Assen, M. A. (2016). Distributions of p-values smaller than .05 in psychology: What is going on? *PeerJ, 4*, Article e1935. https://doi.org/10.7717/peerj.1935

Haven, T. L., Errington, T. M., Gleditsch, K. S., van Grootel, L., Jacobs, A. M., Kern, F. G., … Mokkink, L. B. (2020). Preregistering qualitative research: A Delphi study. *International Journal of Qualitative Methods, 19*. https://doi.org/10.1177/1609406920976417

Heilbron, J. L. (2003). *The Oxford companion to the history of modern science*. Oxford University Press. https://doi.org/10.1093/acref/9780195112290.001.0001

Hensel, P. (2021). Reproducibility and replicability crisis: How management compares to psychology and economics – A systematic review of literature. *European Management Journal, 39*(5), 577–594. https://doi.org/10.1016/j.emj.2021.01.002

Hoole, C. (2019). Avoiding the elephant in the room: The real reasons behind our research crisis. *SA Journal of Industrial Psychology, 45*(1), 1–5. https://doi.org/10.4102/sajip.v45i0.1723

Hotez, P. J. (2021). *The antiscience movement is escalating, going global and killing thousands*. Scientific American. https://www.scientificamerican.com/article/the-antiscience-movement-is-escalating-going-global-and-killing-thousands/.

Hüffmeier, J., Torka, A. K., Jäckel, E., & Schäpers, P. (2022). Open science practices in IWO psychology: Urban legends, misconceptions, and a false dichotomy. *Industrial and Organizational Psychology, 15*(4), 520–524. https://doi.org/10.1017/iop.2022.69

Hummer, L., Thorn, F. S., Nosek, B., & Errington, T. (2017). *Evaluating Registered Reports: A naturalistic comparative study of article impact [Working paper]*. https://doi.org/10.31219/osf.io/5y8w7

Ioannidis, J. P. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science, 7*(6), 645–654. https://doi.org/10.1177/1745691612464056

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

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*(2), 149–155. https://doi.org/10.1038/ng.295

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kepes, S., Banks, G. C., & Keener, S. K. (2020). The TOP factor: An indicator of quality to complement journal impact factor. *Industrial and Organizational Psychology, 13*(3), 328–333. https://doi.org/10.1017/iop.2020.56

Kepes, S., Banks, G. C., McDaniel, M. A., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods, 15*(4), 624–662. https://doi.org/10.1177/1094428112452760

Kepes, S., Bennett, A. A., & McDaniel, M. A. (2014). Evidence-based management and the trustworthiness of our cumulative scientific knowledge: Implications for teaching, research, and practice. *Academy of Management Learning & Education, 13*(3), 446–466. https://doi.org/10.5465/amle.2013.0193

Kepes, S., Keener, S. K., McDaniel, M. A., & Hartman, N. S. (2022). Questionable research practices among researchers in the most research-productive management programs. *Journal of Organizational Behavior, 43*(7), 1153–1286. https://doi.org/10.1002/job.2623

Kepes, S., & McDaniel, M. A. (2013). How trustworthy is the scientific literature in industrial and organizational psychology? *Industrial and Organizational Psychology, 6*(3), 252–268. https://doi.org/10.1111/iops.12045

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*(2), 123–143. https://doi.org/10.1007/s10869-013-9300-2

Kepes, S., Wang, W., & Cortina, J. M. (2022). Assessing publication bias: A 7-step user's guide with best-practice recommendations. *Journal of Business and Psychology*. https://doi.org/10.1007/s10869-022-09840-0

Kepes, S., Wang, W., & Cortina, J. M. (2023). Heterogeneity in effect sizes: An assessment of the current state of the literature. *Organizational Research Methods*. https://doi.org/10.1177/10944281231169942

Kern, F. G., & Gleditsch, K. S. (2017). Exploring pre-registration and pre-analysis plans for qualitative inference. *Preprint ahead of publication*, 1–15.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., … Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology, 14*(5), Article e1002456. https://doi.org/10.1371/journal.pbio.1002456

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C., Nosek, B. A., Hilgard, J., … Ratliff, K. A. (2022). Many Labs 4: Failure to replicate mortality salience effect with and without original author involvement. *Collabra: Psychology, 8*(1), Article 35271. https://doi.org/10.1525/collabra.35271

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., … Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology, 45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Köhler, T., & Cortina, J. M. (2019). Play it again, Sam! An analysis of constructive replication in the organizational sciences. *Journal of Management, 47*(2), 488–518. https://doi.org/10.1177/0149206319843985

Köhler, T., González-Morales, M. G., Banks, G. C., O'Boyle, E. H., Allen, J. A., Sinha, R., … Gulick, L. M. V. (2020). Supporting robust, rigorous, and reliable reviewing as the cornerstone of our profession: Introducing a competency framework for peer review. *Industrial and Organizational Psychology, 13*(1), 1–27. https://doi.org/10.1017/iop.2019.121

Kowalczuk, M. K., Dudbridge, F., Nanda, S., Harriman, L., & Moylan, E. C. (2013). *A comparison of the quality of reviewer reports from author-suggested reviewers and editor-suggested reviewers in journals operating on open or closed peer review models*. Chicago, USA: [Poster Presentation]. 7th International Congress on Peer Review and Biomedical Publication. https://f1000research.com/posters/1094564.

Kowaltowski, A. J. (2021). Brazil's scientists face 90% budget cut. *Nature, 598*, Article 566. https://www.nature.com/articles/d41586-021-02882-z.

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*(26), 2734–2736. https://doi.org/10.1056/NEJMe078110

Landis, R. S., James, L. R., Lance, C. E., Pierce, C. A., & Rogelberg, S. G. (2014). When is nothing something? Editorial for the null results special issue of journal of business and psychology. *Journal of Business and Psychology, 29*(2), 163–167. https://doi.org/10.1007/s10869-014-9347-8

Landy, J. F., Jia, M. L., Ding, I. L., Viganola, D., Tierney, W., , … Dreber, A., & Crowdsourcing Hypothesis Tests Collaboration. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin, 146*(5), 451–479. https://doi.org/10.1037/bul0000220

Ledford, H., Reardon, S., Mega, E. R. E. R., Tollefson, J., & Witze, A. (2019). *Trump seeks big cuts to science funding – Again*. Nature. https://www.nature.com/articles/d41586-019-00719-4.

Lee, S. D., Kuncel, N. R., & Gau, J. (2020). Personality, attitude, and demographic correlates of academic dishonesty: A meta-analysis. *Psychological Bulletin, 146*(11), 1042–1058. https://doi.org/10.1037/bul0000300

Lehrer, J. (2010). *The truth wears off*. The New Yorker. https://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off.

Lepine, J. A., Podsakoff, N. P., & Lepine, M. A. (2005). A meta-analytic test of the challenge stressor–hindrance stressor framework: An explanation for inconsistent relationships among stressors and performance. *Academy of Management Journal, 48*(5), 764–775. https://doi.org/10.5465/amj.2005.18803921

Leung, K. (2011). Presenting post hoc hypotheses as a priori: Ethical and theoretical issues. *Management and Organization Review, 7*(3), 471–479. https://doi.org/10.1111/j.1740-8784.2011.00222.x

Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA publications and communications board task force report. *American Psychologist, 73*(1), 26–46. https://doi.org/10.1037/amp0000151

Maassen, E., van Assen, M. A. L. M., Nuijten, M. B., Olsson-Collentine, A., & Wicherts, J. M. (2020). Reproducibility of individual effect sizes in meta-analyses in psychology. *PLoS One, 15*(5), Article e0233107. https://doi.org/10.1371/journal.pone.0233107

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

Martin, G. N., & Clarke, R. M. (2017). Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology, 8*, 523. https://doi.org/10.3389/fpsyg.2017.00523

Matthews, D. (2014). *Why Congressional Republicans want to cut social science research funding*. Vox. https://www.vox.com/2014/11/12/7201487/congress-social-science-nsf-funding.

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods, 9*(2), 147–163. https://doi.org/10.1037/1082-989x.9.2.147

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist, 70*(6), 487–498. https://doi.org/10.1037/a0039400

Mazzola, J. J., & Deuling, J. K. (2013). Forgetting what we learned as graduate students: HARKing and selective outcome reporting in I–O journal articles. *Industrial and Organizational Psychology, 6*(3), 279–284. https://doi.org/10.1111/iops.12049

McCook. (2017). "Hindsight's a bitch:" Colleagues dissect painful retraction. Retraction Watch. https://retractionwatch.com/2017/03/07/hindsights-bitch-colleagues-dissect-painful-retraction/.

Mertens, G., & Krypotos, A. M. (2019). Preregistration of analyses of preexisting data. *Psychologica Belgica, 59*(1), 338. https://psychologicabelgica.com/articles/10.5334/pb.493.

Methner, N., Dahme, B., & Menzel, C. (2022). The "replication crisis" and trust in psychological science: How reforms shape public trust in psychology. *Social Psychological Bulletin*. https://doi.org/10.23668/psycharchives.12192

Miller, C. C. (2006). Peer review in the organizational and management sciences: Prevalence and effects of reviewer hostility, bias, and dissensus. *Academy of Management Journal, 49*(3), 425–431. https://doi.org/10.5465/amj.2006.21794661

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA GROUP. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 6*(7), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Moonesinghe, R., Khoury, M. J., & Janssens, A. C. J. W. (2007). Most published research findings are false—But a little replication goes a long way. *PLoS Medicine, 4, Article e28*, 218–221. https://doi.org/10.1371/journal.pmed.0040028

Moreau, D., & Gamble, B. (2022). Conducting a meta-analysis in the age of open science: Tools, tips, and practical recommendations. *Psychological Methods, 27*(3), 426–432. https://doi.org/10.1037/met0000351

Morgan, J., Lindsay, B., & Moran, C. (2022). Opening a "closed door": A call for nuance in discussions of open science. *Industrial and Organizational Psychology, 15*(4), 537–541. https://doi.org/10.1017/iop.2022.72

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., … Chartier, C. R. (2018). The psychological science accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science, 1*(4), 501–515. https://doi.org/10.1177/2515245918797607

Murtaugh, P. A. (2002). Journal quality, effect size, and publication bias in meta-analysis. *Ecology, 83*(11), 1162–1166. https://doi.org/10.1890/0012-9658(2002)083[1162:JQESAP]2.0.CO;2

Nair, A. (2019). Funding cuts hurt top scientific institutions. The Times of India. https://timesofindia.indiatimes.com/city/pune/funding-cuts-hurt-top-scientific-institutions/articleshow/69208587.cms.

Naudet, F., Sakarovitch, C., Janiaud, P., Cristea, I., Fanelli, D., Moher, D., & Ioannidis, J. P. A. (2018). Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: Survey of studies published in The BMJ and PLOS Medicine. *British Medical Journal, 360*, Article k400. https://doi.org/10.1136/bmj.k400

Neuliep, J. W., & Crandall, R. (1990). Editorial bias against replication research. *Journal of Social Behavior & Personality, 5*(4), 85–90. https://psycnet.apa.org/record/1991-00044-001.

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

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., … Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences, 23*(10), 815–818. https://doi.org/10.1016/j.tics.2019.07.009

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*(6), 615–631. https://doi.org/10.1177/1745691612459058

Nowok, B., Raab, G. M., & Dibben, C. (2016). synthpop: Bespoke Creation of Synthetic Data in R. *Journal of Statistical Software, 74*(11), 1–26. https://doi.org/10.18637/jss.v074.i11

Nuijten, M. B., Hartgerink, C. H., van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods, 48*(4), 1205–1226. https://doi.org/10.3758/s13428-015-0664-2

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of open data and computational reproducibility in registered reports in psychology. *Advances in Methods and Practices in Psychological Science, 3*(2), 229–237. https://doi.org/10.1177/2515245920918872

O'Boyle, E., Banks, G. C., Carter, K., Walter, S., & Yuan, Z. (2019). A 20-year review of outcome reporting bias in moderated multiple regression. *Journal of Business and Psychology, 34*, 19–37. https://doi.org/10.1007/s10869-018-9539-8

O'Boyle, E., Banks, G. C., & Gonzalez-Mulé, E. (2017). The Chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management, 43*(2), 376–399. https://doi.org/10.1177/0149206314527133

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

Ostriker, J. P., Holland, P. W., Kuh, C. V., & Voytuk, J. A. (2009). *A guide to the methodology of the national research council assessment of doctorate programs*. The National Academies Press. https://nap.nationalacademies.org/catalog/12676/a-guide-to-the-methodology-of-the-national-research-council-assessment-of-doctorate-programs.

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528–530. https://doi.org/10.1177/1745691612465253

Paterson, T. A., Harms, P. D., Steel, P., & Credé, M. (2016). An assessment of the magnitude of effect sizes: Evidence from 30 years of meta-analysis in management. *Journal of Leadership & Organizational Studies, 23*(1), 66–81. https://doi.org/10.1177/1548051815614321

Patten, C. (2021). UK funding cuts are a slap in the face for science. Financial Times. https://www.ft.com/content/ac2042fa-41ed-473b-b3d0-0fe27eed38c7.

Philipp-Muller, A., Lee, S. W. S., & Petty, R. E. (2022). Why are people antiscience, and what can we do about it? *Proceedings of the National Academy of Sciences, 119*(30), Article e2120755119. https://doi.org/10.1073/pnas.2120755119

Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., & Bachrach, D. G. (2008). Scholarly influence in the field of management: A bibliometric analysis of the determinants of university and author impact in the management literature in the past quarter century. *Journal of Management, 34*(4), 641–720. https://doi.org/10.1177/0149206308319533

Retraction Watch. (2015). Diederik Stapel now has 58 retractions. https://retractionwatch.com/2015/12/08/diederik-stapel-now-has-58-retractions/.

Retraction Watch. (2018). Management researcher admits to faslification, resigns. https://retractionwatch.com/2018/03/21/marketing-researcher-admits-to-falsification-resigns/.

Retraction Watch. (2023). Wiley and Hindawi to retract 1,200 more papers for compromised peer review. https://retractionwatch.com/2023/04/05/wiley-and-hindawi-to-retract-1200-more-papers-for-compromised-peer-review/.

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*(3), Article e33423. https://doi.org/10.1371/journal.pone.0033423

Rosman, T., Bosnjak, M., Silber, H., Koßmann, J., & Heycke, T. (2022). Open science and public trust in science: Results from two studies. *Public Understanding of Science, 31*(8), 1046–1062. https://doi.org/10.1177/09636625221100686

Ross, E. (2017). *Social scientists tell congress: "Don"t cut our funding*. Nature. https://www.nature.com/articles/nature.2017.21801.

Rupp, D. E. (2011). Research and publishing ethics: Editor and reviewer responsibilities. *Management and Organization Review, 7*(3), 481–493. https://doi.org/10.1111/j.1740-8784.2011.00227.x

Rynes, S. L., Colbert, A. E., & O'Boyle, E. H. (2018). When the "best available evidence" doesn't win: How doubts about science and scientists threaten the future of evidence-based management. *Journal of Management, 44*(8), 2995–3010. https://doi.org/10.1177/0149206318796934

Sacco, D. F., & Brown, M. (2019). Assessing the efficacy of a training intervention to reduce acceptance of questionable research practices in psychology graduate students. *Journal of Empirical Research on Human Research Ethics, 14*(3), 209–218. https://doi.org/10.1177/1556264619840525

Sarafoglou, A., Kovacs, M., Bakos, B., Wagenmakers, E.-J., & Aczel, B. (2022). A survey on how preregistration affects the research workflow: Better science but more work. *Royal Society Open Science, 9*(7), Article 211997. https://doi.org/10.1098/rsos.211997

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. *Advances in Methods and Practices in Psychological Science, 4*(2). https://doi.org/10.1177/25152459211007467

Schmidt, F. L., & Oh, I.-S. (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of Scientific Psychology, 4*(1), 32–37. https://doi.org/10.1037/arc0000029

Schneider, J., Rosman, T., Kelava, A., & Merk, S. (2022). Do open science badges increase trust in scientists among undergraduates, scientists, and the public? *PsychArchives.*. https://doi.org/10.23668/psycharchives.5066

Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C., Van Assen, M. A., … Schulte-Mecklenbeck, M. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes, 165*, 228–249. https://doi.org/10.1016/j.obhdp.2021.02.003

Sides, J. (2015). Why Congress should not cut funding to the social sciences. Washington Post. https://www.washingtonpost.com/news/monkey-cage/wp/2015/06/10/why-congress-should-not-cut-funding-to-the-social-sciences/.

Siegel, M., Eder, J. S. N., Wicherts, J. M., & Pietschnig, J. (2021). Times are changing, bias isn't: A meta-meta-analysis on publication bias detection practices, prevalence rates, and predictors in industrial/organizational psychology. *Journal of Applied Psychology, 107*(11), 2013–2039. https://doi.org/10.1037/apl0000991

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*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science, 9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *P*-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534–547. https://doi.org/10.1037/a0033242

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., … Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour, 5*(8), 990–997. https://doi.org/10.1038/s41562-021-01142-4

Spector, P. (2022). Is open science rewarding A while hoping for B? *Industrial and Organizational Psychology, 15*(4), 516–519. https://doi.org/10.1017/iop.2022.64

Stapel, D. (2011). *Stapel betuigt openlijk 'diepe spijt'* [*Stapel declares publicly 'deep regret'*]. Brabants Dagblad. Retrieved from http://bd.nl/nieuws/tilburg-stad/stapel-betuigt-openlijk-diepe-spijt-1.121338.

Starbuck, W. H. (2005). How much better are the most-prestigious journals? The statistics of academic publication. *Organization Science, 16*(2), 180–200. https://doi.org/10.1287/orsc.1040.0107

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*(285), 30–34. https://doi.org/10.1080/01621459.1959.10501497

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician, 49*(1), 108–112. https://doi.org/10.1080/00031305.1995.10476125

Stricker, J., & Günther, A. (2019). Scientific misconduct in psychology: A systematic review of prevalence estimates and new empirical data. *Zeitschrift für Psychologie, 227*(1), 53–63. https://doi.org/10.1027/2151-2604/a000356

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

Suls, J., & Martin, R. (2009). The air we breathe: A critical look at practices and alternatives in the peer-review process. *Perspectives on Psychological Science, 4*(1), 40–50. https://doi.org/10.1111/j.1745-6924.2009.01105.x

Swift, J. K., Christopherson, C. D., Bird, M. O., Zöld, A., & Goode, J. (2022). Questionable research practices among faculty and students in APA-accredited clinical and counseling psychology doctoral programs. *Training and Education in Professional Psychology, 16*(3), 299–305. https://doi.org/10.1037/tep0000322

Tenney, E. R., Costa, E., Allard, A., & Vazire, S. (2021). Open science and reform practices in organizational behavior research over time (2011 to 2019). *Organizational Behavior and Human Decision Processes, 162*, 218–223. https://doi.org/10.1016/j.obhdp.2020.10.015

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., … Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One, 10*(8), Article e0134826. https://doi.org/10.1371/journal.pone.0134826

Tipu, S. A. A., & Ryan, J. C. (2021). Are business and management journals anti-replication? An analysis of editorial policies. *Management Research Review, 45*(1), 101–117. https://doi.org/10.1108/MRR-01-2021-0050

Tonidandel, S., Bryan, L., & Morgan, W. (2014). Educating Industrial–Organizational Psychologists: Perspectives from SIOP's Education and Training Committee. *Industrial and Organizational Psychology, 7*(1), 58–61. https://doi.org/10.1111/iops.12106

Torka, A.-K., Mazei, J., Bosco, F., Cortina, J., Götz, M., Kepes, S., … Hüffmeier, J. (2023). How well are open science practices implemented in industrial and organizational psychology and management? *European Journal of Work and Organizational Psychology. Advance online publication.*. https://doi.org/10.1080/1359432X.2023.2206571

Toth, A. A., Banks, G. C., Mellor, D., O'Boyle, E. H., Dickson, A., Davis, D. J., … Borns, J. (2021). Study preregistration: An evaluation of a method for transparent reporting. *Journal of Business and Psychology, 36*(4), 553–571. https://doi.org/10.1007/s10869-020-09695-3

Tourish, D., & Craig, R. (2020). Research misconduct in business and management studies: Causes, consequences, and possible remedies. *Journal of Management Inquiry, 29*(2), 174–187. https://doi.org/10.1177/1056492618792621

Uhlmann, E. L., Ebersole, C., Chartier, C., Errington, T., Kidwell, M., Lai, C. K., … Nosek, B. A. (2019). Scientific utopia III: Crowdsourcing science. *Perspectives on Psychological Science, 14*(5), 711–733. https://doi.org/10.1177/1745691619850561

Van den Akker, O., Peters, G. J., Bakker, C., Carlsson, R., Coles, N. A., Corker, K. S., … Yeung, S. K. (2020). *Inclusive systematic review registration form*. https://doi.org/10.31222/osf.io/3nbea

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

Van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. (1999). Effect of open peer review on quality of reviews and on reviewers' recommendations: A randomised trial. *Bmj, 318*(7175), 23–27. https://doi.org/10.1136/bmj.318.7175

Van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology, 67*, 2–12. https://doi.org/10.1016/j.jesp.2016.03.004

Veldkamp, C. L., Nuijten, M. B., Dominguez-Alvarez, L., Van Assen, M. A., & Wicherts, J. M. (2014). Statistical reporting errors and collaboration on statistical analyses in psychological science. *PLoS One, 9*(12), Article e114876. https://doi.org/10.1371/journal.pone.0114876

von Hippel, P. T. (2022). Is psychological science self-correcting? Citations before and after successful and failed replications. *Perspectives on Psychological Science, 17*(6), 1556–1565. https://doi.org/10.1177/17456916211072525

Wagenmakers, E.-J., & Dutilh, G. (2016). Seven selfish reasons for preregistration. *APS Observer, 29*(9). https://www.psychologicalscience.org/observer/seven-selfish-reasons-for-preregistration.

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*(6), 632–638. https://doi.org/10.1177/1745691612463078

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

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*(7), 726–728. https://doi.org/10.1037/0003-066X.61.7.726

Wiley. (2000-2023). *What is peer review?* Wiley Author Services. https://authorservices.wiley.com/Reviewers/journal-reviewers/what-is-peer-review/index.html.

Wingen, T., Berkessel, J. B., & Englich, B. (2020). No replication, no trust? How low replicability influences trust in psychology. *Social Psychological and Personality Science, 11*(4), 454–463. https://doi.org/10.1177/1948550619877412

Wolins, L. (1962). Responsibility for raw data. *American Psychologist, 17*(9), 657–658. https://doi.org/10.1037/h0038819

Woznyj, H. M., Grenier, K., Ross, R., Banks, G. C., & Rogelberg, S. G. (2018). Results-blind review: A masked crusader for science. *European Journal of Work and Organizational Psychology, 27*(5), 561–576. https://doi.org/10.1080/1359432X.2018.1496081

Yong, E. (2017). How the GOP could use science's reform movement against it. The Atlantic. https://www.theatlantic.com/science/archive/2017/04/reproducibility-science-open-judoflip/521952/.

Yong, E. (2018). Psychology's replication crisis is running out of excuses. The Atlantic. https://www.theatlantic.com/science/archive/2018/11/psychologys-replication-crisis-real/576223/.