Technische Universität Dortmund

Fakultät Erziehungswissenschaft, Psychologie und Bildungsforschung

Advancing Theory and Research on Open Science Practices in Psychology: Insights from Five Investigations

Kumulative Dissertation zur Erlangung des akademischen Grades

Doktorin der Philosophie (Dr. phil.)

vorgelegt von Ann-Kathrin Torka

im Juni 2024

Erstgutachter: Prof. Dr. Joachim Hüffmeier

Zweitgutachter: Prof. Dr. Stefan Krumm

Danksagung

Im Rahmen dieser Dissertation eine umfangreiche (Aus-)Bildung zu genießen, sich mit so vielen spannenden Menschen austauschen zu können, tolle Orte zu bereisen und einen Beruf auszuüben, der mich mit Freude erfüllt, sind Privilegien, derer ich mir sehr bewusst bin. Ich möchte mich daher bei den Menschen bedanken, die mich auf meinem Weg begleitet und unterstützt haben. Zunächst möchte ich mich bei meinem Betreuer und Erstgutachter Joachim Hüffmeier bedanken. Ich danke dir, dass du mich mit deiner Leidenschaft für Forschung angesteckt hast, mich jederzeit in meinem Bestreben bestärkt hast, für die unterstützende Betreuung und dafür, dass wir immer offen miteinander sprechen konnten. Außerdem danke ich Stefan Krumm und Stephan Heinzel für ihre Zeit und Unterstützung, die sie mir als Gutachter in meinem Promotionsverfahren entgegenbringen.

Ein besonderer Dank gilt meinen „Geschwistern“: Jens, der mir 5 Jahre lang jeden Tag zur Seite stand. Elena, die erst eine Verbündete und dann eine enge Freundin wurde. Marc, der immer die passenden Worte gefunden hat, sei es witzig oder einfühlsam. Melanie R., die diesen Weg gemeinsam mit mir begonnen und beendet hat. Melanie S., die jede Konferenz bereichert hat. Corinna, die mich beim Endspurt motiviert hat. Ihr wart und seid so viel mehr als bloß Kolleg*innen. Außerdem bedanke ich mich bei allen Co-Autor*innen sowie bei allen Hilfskräften, mit denen ich zusammenarbeiten durfte und die mich in meiner Arbeit unterstützt haben.

Danke an meine großartigen Freunde, allen voran Carina, die jeden wichtigen Meilenstein gemeinsam mit mir feiert. Zu guter Letzt danke ich meiner wundervollen Familie, auf die ich immer zählen kann. Ich danke meinem Ehemann Hendrik, meinem Ruhepol und Gegengewicht zur Forschung. Danke, dass du mich in meinen Entscheidungen immer unterstützt und mir so viel Verständnis entgegenbringst. Danke, dass du mich aber auch immer wieder daran erinnerst, dass das Leben nicht nur aus Forschung besteht und du mich immer suchen kommst, wenn ich mich in der Arbeit verliere. Mein größter Dank gilt meinen Eltern, die mir das Studium ermöglicht haben und damit den Grundstein für diese Dissertation gelegt haben. Ihr habt nicht nur immer in meine Bildung investiert, sondern seid mir auch immer eine Stütze. Ich habe in den letzten Jahren dank euch so viel erleben und lernen dürfen. Ich kann euch gar nicht genug danken für alles, was ihr für mich tut. Diese Arbeit ist für euch.

Table of Contents

Zusammenfassung

Die vorliegende Dissertation zielt darauf ab, Theorie und Forschung zu Open Science Practices (OSPs) in der Psychologie weiterzuentwickeln. Sie umfasst vier Artikel und zusätzliche Analysen. Der erste Artikel gibt einen Überblick über den aktuellen Stand der Reproduzierbarkeit und Replizierbarkeit in der Arbeits- und Organisationspsychologie (AOP) und verwandten Bereichen. Im Rahmen dieses Überblicks diskutiere ich, wie OSPs die Vertrauenswürdigkeit der Forschung verbessern können, indem sie beispielsweise fragwürdige Forschungspraktiken reduzieren. Der zweite Artikel ist ein Kommentar, der gängige Missverständnisse über OSPs anspricht und deren Vorteile hervorhebt. Anschließend vertieft die Dissertation die Implementierung von OSPs mit vier empirischen Studien. Der dritte Artikel umfasst zwei Studien. Die erste Studie untersucht die Implementierung von OSPs in AOP- und Managementzeitschriften anhand der Webseiten dieser Zeitschriften. Wir stellten fest, dass die meisten Zeitschriften OSPs nicht erwähnten. Die zweite Studie untersucht wahrgenommene Barrieren bei der Implementierung von OSPs anhand einer Befragung unter Herausgeber*innen der AOP- und Managementzeitschriften. Die Antworten zeigten die Notwendigkeit, neue OSPs zu entwickeln und bestehende zu verfeinern, sowie die Verfügbarkeit von Informationen über OSPs zu erhöhen. Der vierte Artikel untersucht, ob und wie Replikationen in den Autor*innen-Richtlinien der sozialpsychologischen Zeitschriften in den Jahren 2015 und 2022 erwähnt werden. Obwohl mehr Zeitschriften 2022 Replikationen begrüßen, werden sie auf den Webseiten der meisten Zeitschriften weiterhin nicht erwähnt. Die zusätzlichen Analysen untersuchen, inwieweit AOP-Studien eine Präregistrierung haben und sich an diese halten. Wir fanden heraus, dass nur wenige Studien eine Präregistrierung hatten und dass Abweichungen hiervon häufig waren. Gemeinsam erweitern die Artikel dieser Dissertation das Wissen über OSPs in der Psychologie und können so dazu beitragen, die psychologische Forschung offener und transparenter zu gestalten.

Abstract

The current doctoral dissertation aims to advance theory and research on Open Science Practices (OSPs) in Psychology. It comprises four articles and additional analyses. The first article reviews the current state of reproducibility and replicability in Industrial and Organizational Psychology (IOP) and related fields. As part of this review, we discuss how implementing OSPs can mitigate factors that undermine research trustworthiness, such as Questionable Research Practices (QRPs). The second article is a commentary addressing common misconceptions about OSPs and highlighting OSPs' benefits. The dissertation then delves deeper into the implementation of OSPs with four empirical studies. The third article includes two studies. The first study examines the implementation of OSPs in IOP and Management journals by analyzing whether journal websites refer to OSPs. We found that most journals did not mention OSPs. The second study investigates perceived barriers to implementing OSPs with a survey among editors from the IOP and Management journals featured in the first study. Their responses indicated a need for developing new and refining existing OSPs, as well as increasing the availability of information on OSPs. The fourth article examines whether and how replications are mentioned in the author guidelines on the websites of Social Psychology journals in 2015 and 2022. Although we found an increase in journals welcoming replications in 2022, they were still not mentioned on the websites of most journals. The additional analyses examine the extent to which IOP studies have preregistrations and adhere to their preregistrations. We found that only very few studies had a preregistration and that deviations from preregistrations were common. Together, the articles included in this dissertation expand the knowledge about OSPs in Psychology and may thereby contribute to making psychological research more open and transparent.

Keywords: Open Science Practices, journal policies, replication, preregistration

General Introduction

In the past decade, Psychology has faced challenges with low reproducibility (i.e., researchers obtaining the same results when reanalyzing data of a published study with the same methodology; Artner et al., 2021; Maassen et al., 2020) and low replicability (i.e., a "finding can be obtained with other random samples" Asendorpf et al., 2013, p. 109; Open Science Collaboration, 2015). These issues have led to what is now termed the "replication crisis" (for an overview, see Nosek et al., 2022). To address this crisis, the Transparency and Openness Promotion (TOP) Guidelines (Nosek et al., 2015) were introduced to establish an Open Science culture in Psychology. The related guidelines are commonly known as Open Science Practices (OSPs).

Since the introduction of these guidelines, nearly a decade has passed. It is timely to assess the development of Psychology in the wake of the replication crisis and to ask: *Is Psychology now a more open science?* Unfortunately, we currently lack comprehensive knowledge about the status of OSP implementation in Psychology and whether there has been progress over time. Thus, in this dissertation, I aim to answer the following five key questions to advance theory and research on OSPs in Psychology: I. How can OSPs help Psychology increase trustworthiness? II. Are OSPs broadly implemented by relevant stakeholders? III. Are OSPs broadly implemented by researchers? IV. Has the implementation of OSPs changed over time? V. Are there differences in the implementation of OSPs between psychological fields?

To address these questions, this dissertation comprises four articles and additional analyses, including two theoretical manuscripts and four empirical studies. The first theoretical article reviews the current state of reproducibility and replicability in Industrial and Organizational Psychology (IOP) and related fields. As part of this review, we[1] discuss

[1] I conducted all articles included in this dissertation with the collaboration of co-authors. To acknowledge their contributions, I will use plural pronouns when referring to these studies. My specific contributions are listed in the "Author Contribution" section.

how implementing OSPs can mitigate factors that undermine research trustworthiness, such as Questionable Research Practices (QRPs; Banks et al., 2016). The second theoretical article is a commentary addressing common misconceptions about OSPs and highlighting their benefits. Together, these two theoretical articles contribute to answering the first key question.

The dissertation then delves deeper into the implementation of OSPs. The second key question is addressed through three studies that are described in the third and fourth article, which examine the implementation of OSPs by stakeholders in Psychology. Article 3 includes two studies. The first of these two studies examines the implementation of OSPs in IOP and Management journals by analyzing whether journal websites refer to OSPs. The second of these two studies investigates perceived barriers to implementing OSPs with a survey among editors from the IOP and Management journals featured in the first study. Article 4 examines whether and how replications are mentioned in the author guidelines on the websites of Social Psychology journals. The third key question is examined through additional analyses. These analyses describe a study about the implementation of preregistrations by researchers in Psychology, specifically examining the extent to which IOP studies have a preregistration and adhere to their preregistrations. The fourth key question is addressed by Article 4 and the additional analyses. Article 4 investigates whether the expressed support for replications in the policies of Social Psychology journals increased from 2015 (i.e., the year the replication crisis became widely recognized) to 2022. Similarly, the additional analyses provide data on the development of the proportion of articles with preregistrations in IOP from 2017 to 2023. Finally, the fifth and last key question is addressed by Article 3 and the additional analyses. Article 3 compares the implementation of OSPs between IOP/Management and Social Psychology. The additional analyses compare the number of disclosed and undisclosed deviations in IOP with a recent sample from the journal *Psychological Science*.

Theoretical Background

For decades, there have been warnings about less than optimal research practices characterized for instance by small sample sizes, resulting in inadequate statistical power (often below 0.50; Cohen 1973, 1992; Sedlmeier & Gigerenzer, 1989; Szucs & Ioannidis, 2017), and a focus on statistical significance, leading to publication bias (Fanelli 2012; Greenwald 1975; Rosenthal 1979; Sterling 1959). In Psychology, these warnings culminated in a "replication crisis" (for an overview, see Nosek et al., 2022), as (a) systematic large scale replication projects revealed a high rate of non-replicable findings (e.g., Klein et al., 2014; Open Science Collaboration, 2015), and (b) many findings were found to be non-reproducible (Artner et al., 2021; Maassen et al., 2020).

To understand the emergence of low replicability and reproducibility, it is crucial to examine the incentive structure in Psychology. The scientific community faces immense pressure to publish more and more in prestigious journals. Publication in high-impact, so called A-Journals, remains the gold standard in Psychology (e.g., Aguinis et al., 2020b). Unfortunately, these journals often favor statistically significant results (Fanelli, 2012; Kepes & McDaniel, 2013; Sterling, 1959). Researchers presenting statistically significant results increase their likelihood to be rewarded with scientific publications, which are essential for obtaining valuable resources like funding, grants, tenure, or promotion (Ball, 2005; Gomez-Mejia & Balkin, 1992; Nosek et al., 2012; Podsakoff et al., 2008). Researchers are therefore incentivized to engage in behaviors that increase their chances of publishing in prestigious journals.

Unfortunately, some researcher behaviors, while increasing the likelihood of obtaining significant results and subsequent publication, undermine the trustworthiness of the collective knowledge by making findings appear more impressive than they actually are (e.g., Götz et al., 2021; Kepes et al., 2022; Murphy & Aguinis, 2019; O'Boyle et al., 2017). These Questionable Research Practices (QRPs; Banks et al., 2016) are "ways of producing,

maintaining, sharing, analyzing, or interpreting data that are likely to produce misleading conclusions, typically in the interest of the researcher" (Nagy et al., 2024, p. 9). QRPs include practices such as hypothesizing after the results are known (HARKing; Kerr 1998), selectively reporting hypotheses (Fanelli 2010, 2012; Greenwald, 1975, Sterling et al., 1995), or selectively adding or dropping data (Kepes et al., 2022; O'Boyle et al., 2017). Reporting only analyses and data that "worked", i.e., that produced findings in favor of a hypothesis, is considered unethical behavior (i.e., *p*-hacking; Simonsohn et al., 2014).

Using QRPs due to the pressure to publish (Bedeian et al., 2010; see also Bedeian, 2003; Kepes et al., 2022; O'Boyle et al., 2017) is referred to as a "'bad barrels' problem" (Banks et al., 2016, p. 328). This suggests that it is not the researchers (i.e., the 'apples') who are inherently bad, but rather the environment in which they operate that persuades them to use these practices. Indeed, QRPs are common among psychologists in America (John et al., 2012) and Italy (Agnoli et al., 2017), as well as in related fields like Management (Banks et al., 2016), Education Research (Makel et al., 2021), and in science overall (Xi et al., 2021). QRPs can result in false positive results by manipulating hypothesized effects to appear statistically significant and inflating effect sizes (Rupp, 2011; Simmons et al., 2011). QRPs may therefore play a critical role in the replication crisis (Earp & Trafimow, 2015; Maxwell et al., 2015). Taken together, "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, 1975)" (Kepes et al., 2022, p. 1192; see also Spector, 2022).

OSPs and Trustworthiness

To address these issues, OSPs were introduced. The TOP Guidelines (Nosek et al., 2015) include eight standards that promote transparency and openness through author guidelines. For instance, sharing data, materials, and code with readers and other researchers is one OSP. Conducting replication studies and preregistering research—specifying a research and analysis plan prior to data collection—are other examples of OSPs. Overall, OSPs are

designed to increase reproducibility (Hardwicke et al., 2018; Obels et al., 2020) and replicability (Protzko et al., 2020). A first objective of this dissertation was to review how OSPs can help Psychology in increasing its trustworthiness. This question is critical because the trustworthiness of psychological research is fundamental to its credibility and impact. By understanding how OSPs can enhance trustworthiness, the field of Psychology can address issues related to QRPs and publication bias. This, in turn, can lead to more reliable and valid findings, which are essential for informing future research, practitioners, and policy-making. Thus, the first key question of this dissertation is as follows:

Question I: How can OSPs help Psychology in increasing trustworthiness?

Implementation of OSPs by Stakeholders

The TOP guidelines were established to encourage journals to implement and promote the use of OSPs among their authors. For instance, journals can encourage authors to preregister their study's hypotheses and analysis plan before data collection and to share any associated data and analysis code when submitting manuscripts. However, if stakeholders such as journals and editors do not support and implement these guidelines, this initiative will not be effective. The guidelines were proposed to the journals because they are particularly important for initiating broad cultural changes. Journals decide which studies are published and which are not, thereby acting as gatekeepers for scientific discoveries (Aguinis et al., 2020a; Kepes et al., 2018). The support and implementation of OSPs by relevant stakeholders is therefore particularly important. Facing precarious careers and limited resources (e.g., limited financial resources, limited time to conduct studies), researchers will align their practices with stakeholder expectations. Therefore, a sustainable and broad cultural change might only be achieved if stakeholders change policies and create incentives for engaging in OSPs (Nosek et al., 2022).

Investigating whether stakeholders, such as funding agencies, journal editors, and academic institutions, implement OSPs is crucial for understanding the broader support

structure for Open Science. Stakeholders play a significant role in setting standards and expectations for research conduct. If they endorse and enforce OSPs, it can lead to a systemic shift towards more open and reliable research practices. Examining the implementation of OSPs by stakeholders and the factors influencing their implementation helps to identify potential barriers or facilitators within the ecosystem of psychological research. Thus, the second key question addresses the stakeholders:

Question II: Are OSPs broadly implemented by relevant stakeholders?

Implementation of OSPs by Researchers

Even if stakeholders broadly support and encourage the use of OSPs, their benefits materialize only if researchers actually incorporate them in their daily practices. Unfortunately, there are also some reservations and critical voices about OSPs (Guzzo et al., 2022; Woznyj et al., 2018). Indeed, initial evidence suggests an implementation gap at least regarding some OSPs, indicating that researchers often do not use them (Aguinis et al., 2020a; Ferguson, 2015; Tenney et al., 2021). The effectiveness of OSPs, however, depends heavily on their practical implementation by researchers. Therefore, it is essential to investigate whether researchers are actually adopting these practices. Understanding implementation rates can provide insights into the current state of Open Science in Psychology. It also highlights areas where additional training, resources, or incentives might be needed to promote a more widespread use of OSPs, or areas where currently available OSPs are less suitable. Thus, the third key question deals with the implementation of OSPs by the individual researcher:

Question III: Are OSPs broadly implemented by researchers?

Changes Across Time

As mentioned above, the “big bang” associated with the widespread awareness of the replication crisis took place almost a decade ago (Open Science Collaboration, 2015). In response, at least some parts of the field of Psychology recognized the need to revise and adapt their academic practices (Pashler & Wagenmakers, 2012; Spellman et al., 2017).

Therefore, it is reasonable to assume that Psychology has made use of this time and that there have been changes in the implementation of OSPs over time. Examining the temporal changes can reveal trends and the impact of various interventions or policy changes over the years. This question is crucial for assessing the progress made in implementing OSPs and identifying periods of significant change or stagnation. It also helps to evaluate the long-term effects of efforts to promote OSPs, informing future strategies for enhancing transparency and reproducibility in psychological research. Therefore, the fourth key question oh this dissertation focuses on changes over time:

Question IV: Did the implementation of OSPs change over time?

Changes Across Different Psychological Disciplines

Various fields within Psychology may exhibit different norms, challenges, and levels of adoption of OSPs. Exploring these differences can offer a nuanced understanding of how OSPs are implemented across sub-disciplines. Identifying strong differences between psychological disciplines could serve as a starting point for research on barriers and specific needs within each sub-discipline. It can help pinpoint specific areas where additional support or tailored interventions might be necessary. For instance, implementing OSPs might vary between quantitative and qualitative research, and it could be more challenging to openly share data collected within an organization compared to data collected among students in a laboratory. Understanding these variations can also promote cross-disciplinary learning and collaboration, ultimately enhancing the overall quality and credibility of psychological research. Thus, the fifth and final key question of this dissertation compares different fields of research in Psychology:

Question V: Are there differences in the implementation of OSPs between psychological fields?

Summary

The aim of this dissertation is to advance theory and research on OSPs in Psychology. The included articles and additional analyses complement each other in addressing the five key questions. Figure 1 provides an overview of the included articles and additional analyses, outlining their key objectives and areas of overlap. Article 1 offers a theoretical model encompassing internal and external factors in organizational Psychology and related fields, researcher behaviors, OSPs, and the resulting trustworthiness of the psychological knowledge. It illuminates the relationship between these factors and how OSPs can enhance trustworthiness in Psychology (i.e., *Question I).* Article 2 addresses some critical voices regarding OSPs and confronts common misconceptions. In doing so, it complements Article 1 which focuses on the positive aspects of OSPs.

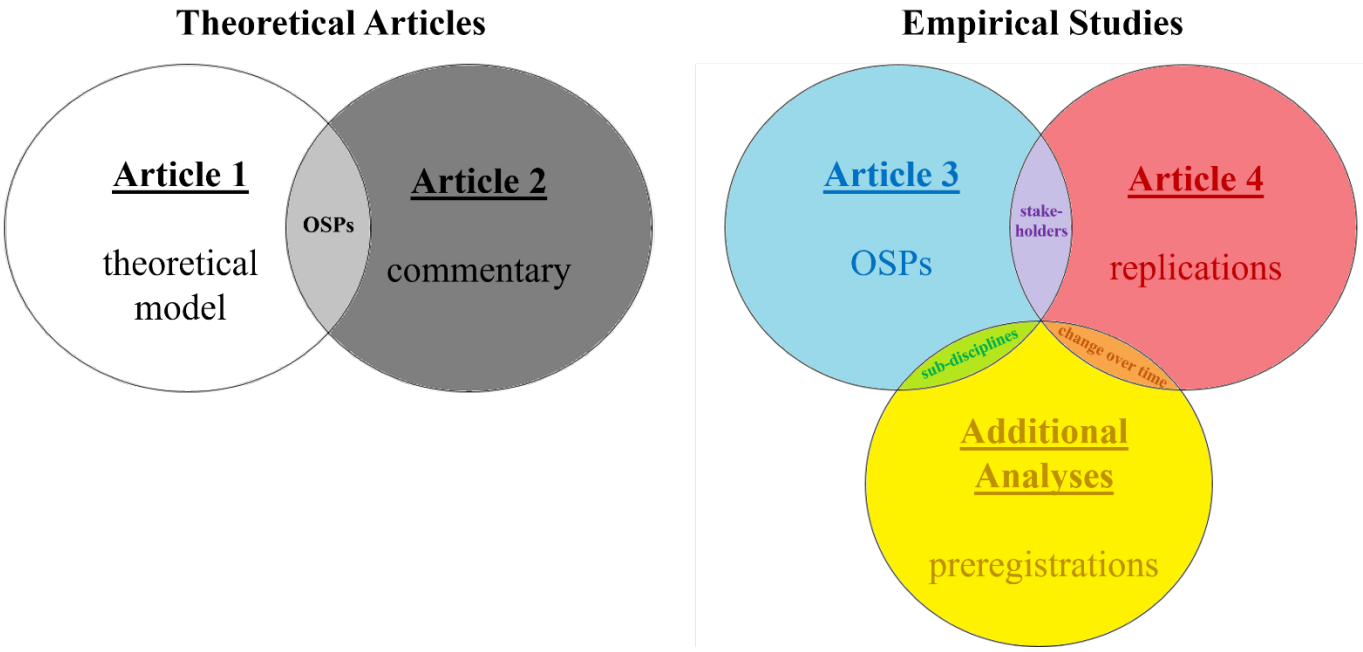
Article 3 examines the support for seven OSPs on psychological journals' websites, as well as editors' reasons for not implementing them. Using a similar methodology, Article 4 investigates psychological journals' support for replication studies and mentions of replications in editorials. However, Article 4 expands the investigation by not only coding *whether* journals say something about OSPs, but rather coding *what* they are actually saying about it (e.g., looking whether they are encouraging or discouraging replications). Both articles shed light on the implementation of different OSPs by (a) journals and (b) editors, which are relevant stakeholders in Psychology (i.e., *Question II*).

The additional analyses take a different approach. By examining all studies with a preregistration in IOP, the study assesses the implementation of preregistrations by individual researchers (i.e., *Question III*). Additionally, by comparing published studies with their preregistrations, we examine the extent to which researchers deviate from their preregistrations, whether the deviations are disclosed or undisclosed, and whether they are likely to be associated with QRPs. The additional analyses have some overlap with Articles 3 and 4. Both Article 4 and the additional analyses focus on changes over time (i.e., *Question*

IV). Article 4 examines changes in OSP implementation by relevant stakeholders, while the additional analyses explore changes over time in individual researchers' adoption of OSPs. Similarly, both the additional analyses and Article 3 examine differences across sub-disciplines within Psychology (*Question V*) by comparing IOP-related research with early adopters of OSPs in Social Psychology (Article 4) and multidisciplinary Psychology (additional analyses).

Figure 1

Articles and Additional Analyses Included in this Dissertation and Their Key Objectives



The four articles included in this dissertation are entitled as follows:

1. 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*, Article 104005. https://doi.org/10.1016/j.actpsy.2023.104005 (* = shared first authorship)

2. 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
3. Torka, A. K., Mazei, J., Bosco, F. A., Cortina, J. M., Götz, M., Kepes, S., O'Boyle, E., & Hüffmeier, J. (2023). How well are Open Science Practices implemented in Industrial and Organizational Psychology and Management? *European Journal of Work and Organizational Psychology*, *32*(4), 461-475. https://doi.org/10.1080/1359432X.2023.2206571
4. Torka, A. K., Mazei, J., & Hüffmeier, J. (2023). Are replications mainstream now? A comparison of support for replications expressed in the policies of Social Psychology journals in 2015 and 2022. *Social Psychological Bulletin*, *18*, 1-22. https://doi.org/10.32872/spb.9695

The additional analyses included in this dissertation are entitled as follows:

- Torka, A. K., & Hüffmeier, J. (2024). *Going off script: Exploring the reporting of preregistration deviations in Industrial and Organizational Psychology and their relationship with QRPs* [Manuscript in preparation]. Department of Psychology, TU Dortmund University.

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. (2020a). 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., Cummings, C., Ramani, R. S., & Cummings, T. G. (2020b). "An a is an a": The new bottom line for valuing academic research. *Academy of Management Perspectives, 34*(1), 135–154. https://doi.org/10.5465/amp.2017.0193

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., Fieder, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., Vanaken, M. A. G., Weber, H., & 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

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

Banks, G. C., O'Boyle Jr, E. H., 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

Bedeian, A. G. (2003). The manuscript review process: The proper roles of authors, referees, and editors. *Journal of Management Inquiry, 12*(4), 331–338. https://doi.org/bw4btf

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://doi.org/10.5465/amle.9.4.zqr715

Cohen, J. (1973). Brief notes: Statistical power analysis and research results. *American Educational Research Journal*, *10*(3), 225-229. https://doi.org/10.3102/00028312010003225

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 1155-1159.

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in Social Psychology. *Frontiers in Psychology*, *6*, Article 138286. https://doi.org/10.3389/fpsyg.2015.00621

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PloS one*, *5*(4), 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

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

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

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., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B., & 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

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., & 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., 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), 1190–1208. https://doi.org/10.1002/job.2623

Kepes, S., List, S. K., & McDaniel, M. A. (2018). Enough talk, it's time to transform: A call for editorial leadership for a robust science. *Industrial and Organizational Psychology*, *11*(1), 43–48. https://doi.org/10.1017/iop.2017.83

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217. https://doi.org/dnqm8w

Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *Academy of Management Journal, 18*(4), 769–783. https://doi.org/10.5465/255378

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., . . . 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

Maassen, E., van Assen, M. A., Nuijten, M. B., Olsson-Collentine, A., Wicherts, J. M., & Gnambs, T. (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., Hodges, J., Cook, B. G., & Plucker, J. A. (2021). Both questionable and open research practices are prevalent in education research. *Educational Researcher*, *50*(8), 493-504. https://doi.org/10.3102/0013189X211001356

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

Murphy, K. R., & Aguinis, H. (2019). HARKing: How badly can cherry-picking and question trolling produce bias in published results*? Journal of Business and Psychology, 34*(1), 1–17. https://doi.org/10.1007/s10869-017-9524-7

Nagy, T., Hergert, J., Elsherif, M. M., Wallrich, L., Schmidt, K., Waltzer, T., Payne, J. W., Gjoneska, B., Seetahul, Y., Wang, Y. A., Scharfenberg, D., Tyson, G., Yang, Y.-F., Skvortsova, A., Alarie, S., Graves, K., Sotola, L. K., Moreau, D., & Rubínová, E. (2024). *Bestiary of Questionable Research Practices in Psychology*. PsyArXiv. https://osf.io/preprints/psyarxiv/fhk98

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/ 10.1126/science.aab2374

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*(1), 719–748. https:// doi.org/10.1146/annurev-psych-020821-114157

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

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

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

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

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

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

Protzko, J., Krosnick, J., Nelson, L. D., Nosek, B. A., Axt, J., Berent, M., Buttrick, N., DeBell, M., Ebersole, C., Lundmark, S., MacInnis, B., O'Donnell, M., Perfecto, H., Pustejovsky, J. E., Roeder, S., Walleczek, J., & Schooler, J. W. (2020). *High replicability of newly-discovered social-behavioral findings is achievable* [Working paper]. Retrieved from https://doi.org/10.31234/osf.io/n2a9x.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

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

Sedlmeier P, & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*(2), 309–316.

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

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

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

Spellman, B., Gilbert, E., & Corker, K. S. (2017). *Open Science: What, why, and how* [Working paper]. https://doi.org/10.31234/osf.io/ak6jr

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

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

Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent Cognitive Neuroscience and Psychology literature. *PLoS Biology*, *15*(3), Article e2000797. https://doi.org/10.1371/journal.pbio.2000797

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/gmt5c5

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

Xie, Y., Wang, K., & Kong, Y. (2021). Prevalence of research misconduct and Questionable Research Practices: A systematic review and meta-analysis. *Science and Engineering Ethics*, *27*, Article 41. https://doi.org/10.1007/s11948-021-00314-9

Article 1

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

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

[a]Old Dominion University

[b]Virginia Commonwealth University

[c]TU Dortmund University

Note. The order of authorship is alphabetical as the authors contributed equally to this work.

This manuscript is published as:
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*, Article 104005. https://doi.org/10.1016/j.actpsy.2023.104005

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.

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

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

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. 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., 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. After reviewing these factors, it is clear that "the only reasonable conclusion left is that I-O psychology is facing the same crisis in confidence/replication as other areas of psychology" (Efendic & Van Zyl, 2019, p. 2). To remedy this, we then provide several suggestions for how the identified problems can be addressed, including highlighting some ways in which we are already making progress.

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 et al., in press-a), which explicitly indicates that research findings, even meta-analytic ones, may not be robust and generalizable.

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.

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 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 1,978 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.

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 (also referred to exact or direct replications; Simons, 2014; Stroebe & Strack, 2014), quasirandom replications (also referred to conceptual extensions or replications; Stroebe & Strack, 2014; Tsang & Kwan, 1999), constructive replications (also referred to as conceptual replications; Stroebe & Strack, 2014), 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).[1]

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,

[1] 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).

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., 2022; Klein et al., 2014; 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 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.

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 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 retractions published in PsycINFO, Craig et al. (2020) found that, across psychology disciplines, data fabrication, falsification, and fraud accounted for 48% of the reasons for 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, the reason for a retraction was coded as data fabrication, falsification and fraud in 33% of cases;

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 .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., .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.

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)

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 .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.[2] 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) 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 et al., 2017; Harms et al., 2018), inconsistencies between reported degrees of freedom and models

[2] 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.

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).

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 60-70% 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).

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 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 1,200 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.

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.

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 et al., 2018a). Education about replicability and reproducibility might (at least temporally) shift undergraduates' attitudes toward 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.

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, 1975)" (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.

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 et al., 2018a). In such a changed reward system, researchers would not need doing everything they can (e.g., misconduct, QRPs) to increase the chances their article will get published in a “prestigious” journal.

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 and colleagues (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.

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 & Rosenbaum, 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 et al., 2018b; 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 et al., 2018b; 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 et al., 2018b).

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 et al., in press-b).

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.

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[3]), the Center for Open Science (COS) developed the TOP factor (COS, 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

[3] 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 et al. (2022a) 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.

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 & 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)

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).

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 collection, like with the results blind review process. The main difference between RRs and RBRs is the timing of data collection (Grand et al., 2018b). 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 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 less 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 and colleagues (2017) found similar to slightly greater numbers of citations for RRs.

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).

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). 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).

Implementation of OSPs. In general, OSPs can increase the perceived trustworthiness of research (Methner et al., in press; Schneider et al., 2022; Rosman 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, 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).

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 (Open Science Collaboration, 2015; Ebersole et al., 2016; Klein et al., 2022). Additionally, only a small number of studies deal with constructively replicating published effects (Makel et al., 2012; Köhler & Cortina, 2019). 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.

Figure 1 summarizes the internal and external factors that threaten our field's trustworthiness. In Box 2, we highlight some internal factors, such as personal characteristics, can raise an individual's probability to engage in destructive behaviors (Lee et al., 2020; Fong & Tosi, 2007). 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 4), 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 5) 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 Figure 1). The first box contains a variety of proposed changes that would combat the current

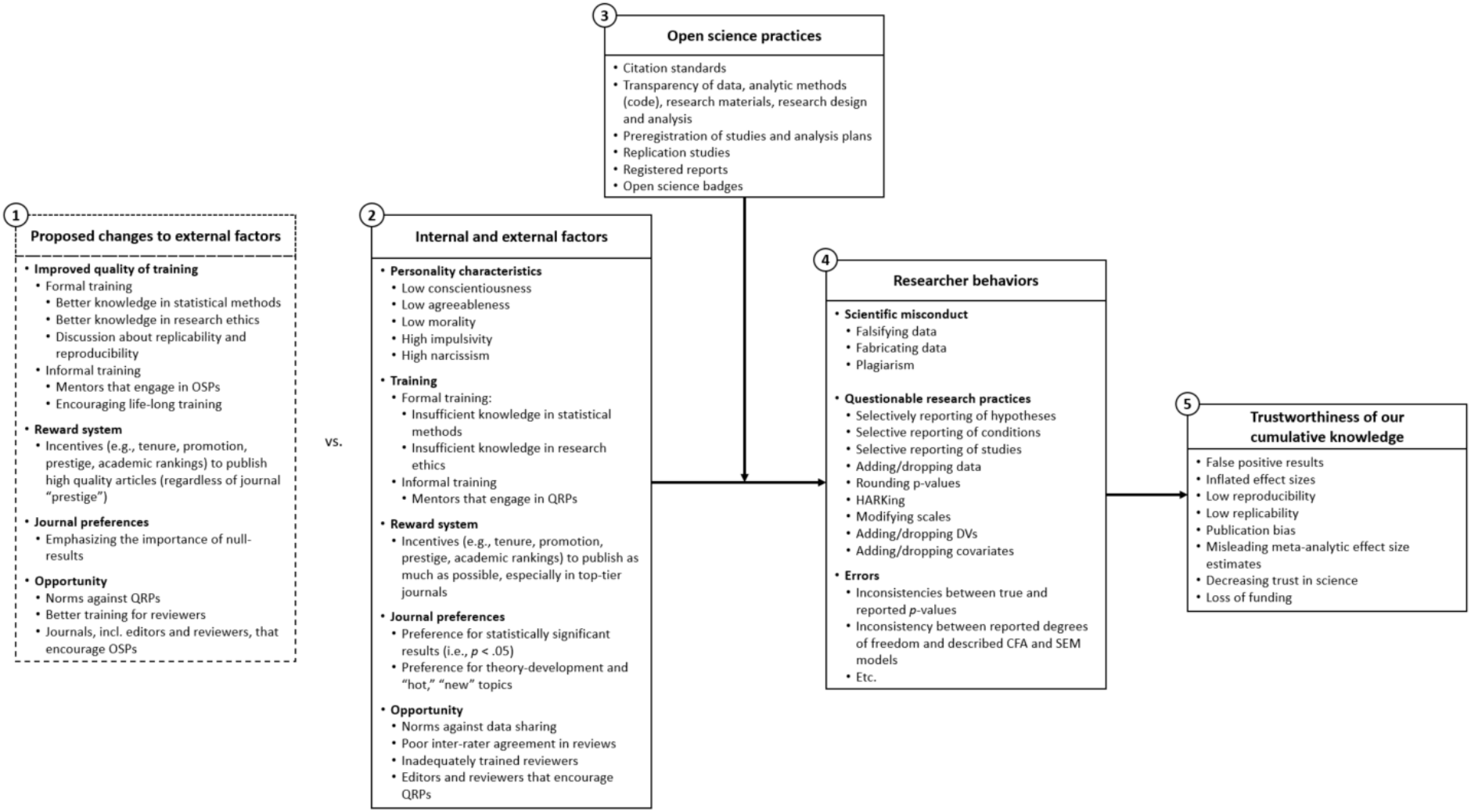
external factors identified in Box 2. One of these recommendations involves enhancing the quality of both formal and informal scientific training, with a particular focus on topics related to replicability and reproducibility. Additionally, it would be beneficial to implement a reward system that promotes 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 3) 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 systematic adoption of suggested changes by academic institutions and journals. Indeed, as noted by Castille et al. (2022b), if each of us take small steps toward 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 and preregistration. For more recommendations for different stakeholders in IO, see Table 2 in Grand et al., (2018a).

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.

Figure 1.
Theoretical model: Reasons for and examples of misconduct, QRPs, and errors



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

Anvari, F., & Lakens, D. (2018). The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology, 3*(3), 266-286. https://doi.org/10.1080/23743603.2019.1684822

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., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., Vanaken, M. A. G., Weber, H., & 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.

Banks, G. C., O'Boyle Jr, E. H., 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.

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.

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.

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.

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., Grange, J. A., Perugini, M., Spies, J. 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), 15–15. https://doi.org/10.1038/479015a

Carey, B. (2011, November 2). *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. (2022b). 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. (2022a). 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. http://dx.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.

De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Maarusic, An., Overbeke, A. J. P. M., Schroeder, T. V., Sox, H. C., & 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., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., . . . 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., Corker, K. S., Corley, M., Hartshorne, J. K., IJzerman, H., Lazarevic´, L. B., Rabagliati, H., Ropovik, I., Aczel, B., Aeschbach, L. F., Andrighetto, L., Arnal, J. D., Arrow, H., Babincak, P., & 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. doi: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), 1037-1037. 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., Tonidandel, S., & Truxillo, D. M. (2018a). 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. (2018b). 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., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B., & 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*, 1609406920976417.

Heilbron, J. L. (2003). *The Oxford companion to the history of modern science*. Oxford University Press.

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://dx.doi.org/10.4102/sajip.v45i0.1723

Hotez, P. J. (2021, March 29). *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. 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. (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., Allison, D. B., Ball, C. A., Coulibaly, I., Cui, X., Culhane, A. C., Falchi, M., Furlanello, C., Game, L., Jurman, G., Mangion, J., Mehta, T., Nitzberg, M., Page, G. P., Petretto, E., 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., & 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. (2015). The validity of conscientiousness is overestimated in the prediction of job performance. *PloS One*, *10*(10), e0141468. https://doi.org/10.1371/journal.pone.0141468

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., 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., 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., 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. (in press-b). 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. (in press-a). Heterogeneity in effect sizes: An assessment of the current state of the literature. *Organizational Research Methods*.

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/dnqm8w

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik,A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, 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., Ahn, P. H., Brady, A. J., Chartier, C. R., Christopherson, C. D., Clay, S., Collisson, B., Crawford, J. T., Cromar, R., Gardiner, G., Gosnell, C. L., Grahe, J., Hall, C., Howard, I., . . . 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., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., . . . 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., Woo, S. E., & 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, September 8-10). 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. [Poster Presentation]. 7th International Congress on Peer Review and Biomedical Publication, Chicago, USA. https://f1000research.com/posters/1094564

Kowaltowski, A. J. (2021, October 25). *Brazil's scientists face 90% budget cut*. Nature. https://www.nature.com/articles/d41586-021-02882-z

Laine, C., Horton, R., DeAngelis, C. D., Drazen, J. M., Frizelle, F. A., Godlee, F., Chir, B., Haug, C., Hébert, P. C., Kotzin, S., Marusic, A., Sahni, P., Schroeder, T. V., Sox, H. C., Van Der Weyden, M. B., & 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. http://dx.doi.org/10.1037/bul0000220

Ledford, H., Reardon, S., Mega, E. R., Tollefson, J., & Witze, A. (2019, March 11). 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, December 13). *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, November 12). *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, March 7). *“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.

Methner, N., Dahme, B., & Menzel, C. (in press). The "replication crisis" and trust in psychological science: How reforms shape public trust in psychology. *Social Psychological Bulletin*.

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., & and 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

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

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

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J.,Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J.,Keene, J. R., Protzko, J., Aczel, B.,:::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

Nagy, T., Hergert, J., Elsherif, M. M., Wallrich, L., Schmidt, K., Waltzer, T., Payne, J. W., Gjoneska, B., Seetahul, Y., Wang, Y. A., Scharfenberg, D., Tyson, G., Yang, Y.-F., Skvortsova, A., Alarie, S., Graves, K., Sotola, L. K., Moreau, D., & Rubínová, E. (2024). *Bestiary of Questionable Research Practices in Psychology*. PsyArXiv. https://osf.io/preprints/psyarxiv/fhk98

Nair, A. (2019, May 7). *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.

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., Buch, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … & 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., van 't Veer, A. E., & 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

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

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

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

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.

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, April 17). *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, December 8). *Diederik Stapel now has 58 retractions.* https://retractionwatch.com/2015/12/08/diederik-stapel-now-has-58-retractions/

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

Retraction Watch (2023, April 5). *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, April 14). *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? *Psychological Science, 33*(9), 1588-1604. https://doi.org/10.1177/09567976221097499

Schwartz, S. J., Lilienfeld, S. O., Meca, A., & Sauvigné, K. C. (2016). The role of neuroscience within psychology: A call for inclusiveness over exclusiveness. *American Psychologist*, *71*(1), 52–70. https://doi.org/10.1037/a0039678

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, June 10). *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., Esterling, K. M., & 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

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

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

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., Pollock, D., & Dorsett, K. (2015) Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE 10*(8): 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., O'Boyle, E., & 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., DeHaven, A., Bochantin, 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

Tsang, E. W. K., & Kwan, K.-M. (1999). Replication and theory development in organizational science: A critical realist perspective. *Academy of Management Review, 24,* 759-780. https://doi.org/10.2307/259353

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

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

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.318.7175.23Vazire, S., & Holcombe, A. O. (2022). Where are the self-correcting mechanisms in science? *Review of General Psychology, 26*(2), 212–223. https://doi.org/10.1177/10892680211033912

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), 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, October 31). Seven selfish reasons for preregistration. *APS Observer, 29*(9).

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, April 5). *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, November 19). *Psychology's replication crisis is running out of excuses.* The Atlantic. https://www.theatlantic.com/science/archive/2018/11/psychologys-replication-crisis-real/576223/

Article 2

Open Science Practices in IWO Psychology: Urban Legends, Misconceptions, and a False Dichotomy

Joachim Hüffmeier[a], Ann-Kathrin Torka[a], Elisabeth Jäckel[b], and Philipp Schäpers[c]

[a]TU Dortmund University

[b]University of Amsterdam

[c]University of Münster

This manuscript is published as:
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

Open Science Practices in IWO Psychology: Urban Legends, Misconceptions, and a False Dichotomy

Although we appreciate Guzzo et al. (2022) addressing the issue of Open Science Practices (OSPs) and pointing out potential risks, we believe that their focal article is neither an accurate reflection of OSPs nor of the related perils. In our commentary, we highlight and elaborate the following four (partly) interrelated and problematic issues that run the risk of misrepresenting the usefulness of OSPs: (i) There are very good reasons why OSPs are currently discussed, although they are hardly mentioned by Guzzo and colleagues. The perils that the authors perceive are either (ii) exaggerated and dramatized or (iii) simply due to misconceptions related to OSPs on their part. (iv) Guzzo et al. see a dichotomy between different types of science within Industrial, Work, and Organizational (IWO) Psychology and suppose that the usefulness of OSPs is limited to only one of them (i.e., the hypothetico-deductive approach).

(i) OSPs as a Bad "New" Key Norm

In view of an article that intends to advocate for the value of open science, it was astonishing to see how little attention the authors devoted to illustrating how OSPs can, in fact, address current problems that plague many different literatures. These problems comprise (among others): (i) a widespread unwillingness by authors to share their data, (ii) big difficulties to reproduce published findings even if the data are available, (iii) an alarmingly high rate of confirmed hypotheses in the published literature that is most likely due to (iv) the journals' unwillingness to publish null findings and (v) resulting file-drawing of statistically non-significant findings, and (vi) an underrepresentation of replication studies (Ryan & Tipu, 2021). Notably, the way in which the authors describe these problems is rather misleading. For instance, there is a broad consensus in the literature that Questionable Research Practices (QRPs; e.g., only reporting supported hypotheses) are not the same thing as plain fraud (e.g., Banks et al., 2016). However, the authors equate these two and incorrectly

describe that OSPs were developed to prevent fraud, although OSPs were in fact developed to reduce researchers' use of QRPs.

However, more importantly for this commentary and deviating from Guzzo et al. (2022), we are not convinced that IWO Psychology is currently governed by norms that define "good" research. By contrast, OSPs have been repeatedly shown to have the potential to improve the currently problematic state. For instance, studies with high statistical power, methodological transparency, and preregistration were found to be highly replicable—a very desirable, but often not realized feature in psychology research (Protzko et al., 2020). Moreover, preregistered studies were found to be more transparent in their findings than non-preregistered studies and exhibited a lower rate of confirmed hypotheses (Toth et al., 2020). As a last example, Open Science Badges were not only found to increase materials and data sharing. With OS Badges, the shared data were also more likely available and correct, as well as usable and complete (as compared to when authors only indicated data availability; Kidwell et al., 2016). Altogether, some (flawless) research practices would change when applying OSPs to IOW Psychology research.

(ii) There is a Movement in the Direction of Transparency and Full Data Sharing

Guzzo et al. (2022) perceived a peril in terms of disclosure and sharing. First, we want to acknowledge that data sharing can sometimes be problematic or even impossible, especially in view of company data that would give competitors valuable insights or that allows the identification of single employees. Nonetheless, this problem has already been recognized for decades and there is no lack of constructive ideas on how to address this challenge (Banks et al., 2016). In fact, Guzzo et al. themselves name a few of these measures (e.g., sharing only a relevant portion of a full data set).

It is, however, more important in the context of the current commentary, that there is clearly *no visible movement* in IWO Psychology to the direction of complete transparency and full data sharing as an absolute requirement. There are two crucial pieces of evidence that

Guzzo et al. (2022) neglect, but that clearly confirm our perspective. First, the Transparency and Openness Promotion (TOP) guidelines (Nosek et al., 2015), which are used to design and evaluate journal policies regarding Open Science, very clearly articulate that even the two strictest levels of data transparency (Level 3 and Level 2) do *not* prescribe *public* data sharing. Importantly, even if data sharing would be required (which is currently not the case, for an overview see Table 1), data would not have to be shared *publicly.* Moreover, even if the second strictest transparency level were applied by a journal (which is currently the case for only 29 of 240 Psychology and six of 40 Business & Management journals, see Table 1), researchers could always argue why they cannot or do not want to share (or upload) their data.

Second, data on how journals currently adopt these standards clearly reflect that most journals have not implemented any data sharing requirements at all so far. For instance, an analysis of the current implementation of data transparency requirements for "Psychology" and "Business and Management" journals using the TOP factor website (https://topfactor.org/) reveals that more than half of the journals in Psychology and about two thirds of the journals in Business and Management do not have any data sharing-related requirements at all (see Table 1). Moreover, there is only one single journal—the journal *Meta-Psychology*—in Psychology (one out of 240) and no journal in Business and Management where data sharing is an absolute requirement. In our own current research, we found similar results specifically for IWO journals (Torka et al., 2022). Thus, the authors' claims regarding data sharing appear to be exaggerated and their recommendation on how to solve this problem is already fully implemented in the TOP guidelines (Nosek et al., 2015). To sum it up, it is *neither* the idea underlying of OSPs to force researchers to engage in unintended research practices, *nor* is there any movement in such a direction. Consequently, there is no competition between journals to achieve the strictest levels of transparency and openness.

(iii) Replications as a Driver of Research with Inadequate Statistical Power

When describing the paradox of replication, it appears as if Guzzo et al. (2022) have (fundamental) misconceptions about OSPs. First, given the current underrepresentation of replication studies (Ryan & Tipu, 2022) that is accompanied by typically low statistical power in the original studies (Paterson et al., 2016), we do not currently know which findings from IWO Psychology can be replicated at all and which findings cannot. There is little reason to assume that IWO Psychology has a higher replication rate than the rather low rates of its neighbouring fields like Management, Marketing, or Economics. This is because the prevailing incentives and publishing practices are highly similar across these fields.

Second, even if all IWO Psychology findings were replicable, it would be necessary to conduct further independent and exact replication studies. These studies would allow to control for QRPs (e.g., reporting only supported hypotheses or arbitrarily excluding outliers to obtain significant results) and test whether findings are generalizable across different populations, countries, and cultures. This is why replication studies typically find smaller effect sizes and provide more accurate effect size estimates than the original studies (e.g., Ebersole et al., 2020; Stewart & Shapiro, 2000). Thus, in contrast to the assertions of Guzzo et al. (2022), IWO Psychology does not only need conceptual, but also exact replications.

Third and perhaps most importantly, the authors assert that replications will encourage projects with small sample sizes (Guzzo et al., 2022). However, since the beginning of replication efforts, it is broad consensus and widely established that replication studies have to be adequately powered (e.g., Simonsohn, 2015). In line with this consensus and in contrast to the assertions by Guzzo et al., replication studies typically have much more statistical power than the original studies (e.g., Ebersole et al., 2020; Stewart & Shapiro, 2000). Hence, given the higher power of replication studies, the assertion that replications run the risk of overestimating the magnitude of “true” relationships is simply wrong. In fact, the opposite is true: Well-powered replication studies often help to adequately assess the magnitude of “true”

relationships and they typically find (much) smaller effect sizes than the original studies (e.g., Ebersole et al., 2020; Stewart & Shapiro, 2000).

(iv) Preregistrations Explicitly Rely on an Hypothetico-deductive Approach

Guzzo et al. (2022) postulate a pronounced dichotomy between hypothetico-deductive research and applied research (i.e., field research in organizations) regarding their respective adequacy in view of OSPs (e.g., regarding preregistrations). We believe that this dichotomous perspective is false for several reasons. First, preregistering a study does *not* limit the possibilities to engage in any type of research. In fact, it does *not* inhibit creating new explanations after data collection. It rather prevents presenting a post hoc finding as an a priori hypothesis (i.e., Hypothesizing After the Results are Known [HARKing]). Moreover, not OSP, but research practices like HARKing and further QSPs devalue exploratory studies because much too often it is not clear how researchers end up with the findings they report in their manuscripts. By contrast, preregistering a study simply helps (a) to distinguish between studies that develop versus test hypotheses or (b) to illustrate at which points of time (or in which analytical steps) hypotheses are developed and then tested when using the same data set for both activities. In fact, preregistrations were likened to a systematic research log that clarifies which findings are consistent with and which are diverged from expectations (Haven et al., 2020). Such a research log can build the needed trust in reported (exploratory) findings because the research process becomes more transparent.

Moreover, preregistrations can also accommodate innovative approaches, for instance analysing big data from organizations, which Guzzo et al. (2022) describe as particularly problematic in their article. The different preregistration templates offered in the Open Science Framework (OSF)[1], the most prominent tool for preregistrations in psychology, and

[1] As a side note and as a reply to the authors' observations (see Guzzo et al., 2022, p. 44), it is possible to preregister a study on the OSF with email addresses that end in .com, also when using the aspreditced.org template (and further templates). It is furthermore possible to preregister studies when the data are already collected at the time of preregistration.

especially the template for analysing pre-existing data by Mertens and Krypotos (2019) are known for their flexibility. We suspect that the authors' wrong impression was guided by their apparently exclusive focus on the preregistration platform aspredicted.org and the related template, which is primarily tailored to the needs of experimental research.

Second, preregistrations are neither rigid nor "one-size-fit-all" devices that cannot be accommodated to own projects. Thus, there are various templates for all kinds of research approaches and methods that clearly show that preregistration is not limited to the hypothetico-deductive research approach. For instance, there are templates available for qualitative research (Haven et al., 2020), the analysis of pre-existing data (Mertens & Krypotos, 2019), or meta-analyses. Hence, we do not share the authors presented dichotomy of research approaches because preregistration comes in many different forms and shapes and can be handled very flexibly. In summary, we do not believe that OSPs would prevent any research in IWO Psychology because no one will be forced to apply OSPs that would harm a research project, is inconsistent with ethical guidelines or would get a person/company in trouble. By contrast, we firmly believe that more transparency and openness have the potential to improve and strengthen current and future IWO Psychology research.

References

Banks, G. C., O'Boyle Jr, E. H., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., Abston, K. A., Bennett, A. A., & 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/gf2k58

Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D. J., Buttrick, N. R., Chartier, C. R., Corker, K. S., Corley, M., Hartshorne, J. K., IJzerman, H., Lazarević, L. B., Rabagliati, H., Ropovik, I., Aczel, B., Aeschbach, L. F., Andrighetto, L., Arnal, J. D., Arrow, H., Babincak, P., ... & 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/gkqrqx

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.*

Haven, T. L., Errington, T. M., Gleditsch, K. S., van Grootel, L., Jacobs, A. M., Kern, F. G., Piñeiro, R., Rosenblatt, F., & Mokkink, L. B. (2020). Preregistering qualitative research: A Delphi study. *International Journal of Qualitative Methods, 19*, 1–13. https://doi.org/gjzbw6

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, 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/f8pkck

Mertens, G., & Krypotos, A.-M. (2019). Preregistration of analyses of preexisting data. *Psychologica Belgica, 59*(1), 338-352. https://doi.org/ggdc45

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... & Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*(6242), 1422-1425. https://doi.org/gcpzwn

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/bjz9

Protzko, J., Krosnick, J., Nelson, L. D., Nosek, B. A., Axt, J., Berent, M., Buttrick, N., DeBell, M., Ebersole, C. R., Lundmark, S., MacInnis, B., O'Donnell, M., Perfecto, H., Pustejovsky, J. E., Roeder, S., Walleczek, J., & Schooler, J. W. (2020). *High replicability of newly-discovered social-behavioral findings is achievable*. PsyArXiv. https://doi.org/h5rt

Ryan, J. C., & Tipu, S. A. (2022). Business and management research: Low instances of replication studies and a lack of author independence in replications. *Research Policy, 51*(1), Article 104408. https://doi.org/gnk94x

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*(5), 559-569. https://doi.org/gfpn9t

Stewart, M. M., & Shapiro, D. L. (2000). Selection based on merit versus demography: Implications across race and gender lines. *Journal of Applied Psychology, 85*(2), 219–231. https://doi.org/bwv89b

Torka, A.-K., Mazei, J., Bosco, F., Cortina, J., Götz, M., Kepes, S., O'Boyle, E., & Hüffmeier, J. (2022). *Open Science Practices in management and industrial and organizational psychology: Not much talk, not much walk.* Manuscript submitted for publication.

Toth, A. A., Banks, G. C., Mellor, D., O'Boyle, E. H., Dickson, A., Davis, D. J., DeHaven, A., Bochantin, J., & Borns, J. (2020). Study preregistration: An evaluation of a method for transparent reporting. *Journal of Business and Psychology, 36*, 553-571. https://doi.org/ghcrqm

Table 1

Overview of the currently realized TOP standard "data transparency" for the journal categories "Business & Management" and "Psychology" (from TOP factor website [https://topfactor.org/])

TOP factor level	Business & Management		Psychology	
	N	%	N	%
Level 3: Data must be posted to a trusted repository and reported analyses will be reproduced independently prior to publication.	0	0 %	1	0.42 %
Level 2: Data must be posted to a trusted repository. Exceptions must be identified at article submission.	6	15 %	29	12.08 %
Level 1: Article states whether data are available, and, if so, where to access them.	8	20 %	82	34.17 %
Not Implemented: Journal encourages data sharing or does not provide any information about data transparency.	26	65 %	128	53.33 %
Overall	40	100 %	240	100 %

Note. Some journals are included in "Business & Management" and "Psychology."

Article 3

How Well Are Open Science Practices Implemented in Industrial and Organizational Psychology and Management?

Ann-Kathrin Torka[a], Jens Mazei[a], Frank Bosco[b], Jose Cortina[b],

Martin Götz[c], Sven Kepes[b], Ernest O'Boyle[d], and Joachim Hüffmeier[a]

[a]TU Dortmund University

[b]Virginia Commonwealth University

[c]University of Zurich

[d]Indiana University

This manuscript is published as:
Torka, A. K., Mazei, J., Bosco, F. A., Cortina, J. M., Götz, M., Kepes, S., O'Boyle, E., & Hüffmeier, J. (2023). How well are Open Science Practices implemented in Industrial and Organizational Psychology and Management? *European Journal of Work and Organizational Psychology*, *32*(4), 461-475. https://doi.org/10.1080/1359432X.2023.2206571

Article 4

Are Replications Mainstream now?

A Comparison of Support for Replications Expressed in the Policies of Social Psychology Journals in 2015 and 2022

Ann-Kathrin Torka, Jens Mazei, and Joachim Hüffmeier

TU Dortmund University

This manuscript is published as:
Torka, A. K., Mazei, J., & Hüffmeier, J. (2023). Are replications mainstream now? A comparison of support for replications expressed in the policies of Social Psychology journals in 2015 and 2022. *Social Psychological Bulletin*, *18*, 1-22. https://doi.org/10.32872/spb.9695

Highlights

- The expressed support for replication studies in policies of social psychology journals increased between 2015 and 2022, while the majority of journals still do not mention replications on their websites.
- Exploratory analyses suggested that journals that expressed support for replications on their websites are actually more likely to publish articles about replication.
- Further exploratory analyses of the journals' TOP factors indicated similar rates of support for replications as for other rigor and transparency promoting policies.
- In sum, our findings suggest that appreciation for replication has increased, but is not yet part of mainstream culture in social psychology.

Abstract

A decade ago, replications were typically not conducted and appreciated in social psychology, although replications play a central role in ensuring trust in scientific fields. Without systematic replication efforts, it is not clear whether findings are trustworthy. As journals can function as gatekeepers for publications, they can influence whether researchers conduct (and publish) replications. Yet, the scholarly culture in social psychology might have changed over the last decade because numerous highly visible studies did not replicate past findings. In light of these insights and the resulting learning opportunities for the field, we predicted an increase in the expressed support for replications in the policies of social psychology journals from 2015 (i.e., the year the replication problem became widely known) to 2022. We coded whether and how replications were mentioned in the author guidelines on the websites of social psychology journals ($N = 51$). As expected, replications were welcomed more often in 2022 (25%) than they were in 2015 (12%), but they were not mentioned on the websites of most journals (71% in 2022 vs. 82% in 2015). An exploratory analysis suggested that journals that expressed support for replications on their websites were also more likely to publish articles about replication. Further, exploratory analyses of the journals' TOP factors indicated similar rates of support for replications as for other rigor and transparency promoting policies. In sum, our findings suggest that appreciation for replication has increased, but is not yet part of mainstream culture in social psychology.

Keywords: replication, journal policies, open science, replication crisis

Are Replications Mainstream now?

A Comparison of Support for Replications Expressed in the Policies of Social Psychology Journals in 2015 and 2022

More than a decade ago, social psychology was confronted with a "crisis of confidence" (Pashler & Wagenmakers, 2012, p. 528). Systematic replication efforts suggested a relatively high non-replication rate (e.g., Klein et al., 2014; Open Science Collaboration, 2015), and the use of questionable research practices (QRPs; Banks et al., 2016) appeared to be common (Götz et al., 2021; John et al., 2012). This "manifest" crisis was preceded by another, more "latent" crisis that was perhaps even more problematic—a "period, when we were [even] unaware of the problem and thus did nothing about it" (Nelson et al., 2018, p. 512). Back then, it seems, the field did not value replications and, thus, was not even aware of the possible low replicability of its research findings. Yet, replications play a central role in ensuring trust in scientific fields, such as social psychology (Simons, 2014). As they clarify research findings' reliability, they can be seen as the "Supreme Court of the scientific system" (Collins, 1985, p. 19). Conversely, *without* replications, it is more difficult to assess the reliability of our findings.

Since the beginning of the manifest replication crisis, methods have been developed to improve the field's research and publication practices. For example, open science practices (OSPs) aim to increase transparency, openness, and reproducibility. In 2015, OSPs, including support for publishing replications, were described in a policy framework called the Transparency and Openness Promotion (TOP) Guidelines for journals (Nosek et al., 2015). Initiatives like the TOP Guidelines may not be sufficient to initiate broad cultural change without support from relevant stakeholders. Indeed, we may *still* work in a "dysfunctional social culture" (Nosek et al., 2022, p. 733) that discourages replications. Journals are particularly important for advancing culture change as they are gatekeepers for scientific discoveries (Aguinis et al., 2020; Kepes et al., 2018). Thus, we investigated whether social

psychology journals have increased their willingness to publish replications between 2015 and 2022 as indicated by the public policies on their websites. Moreover, we explored whether social psychology journals have increased the rate of publishing replications since 2015 and depending on whether their websites welcomed replications.

Literature Review

Replications play a central role in ensuring trust in entire scientific fields, such as social psychology, because they examine their reliability (Simons, 2014). They offer the possibility to evaluate scientific discoveries and, thus, should be "a fundamental feature of the scientific process" (Zwaan et al., 2018, p. 3). Replications, if conducted rigorously, inform theory and matter independently of their results. They allow examination of sampling errors, artefacts, and may even unveil fraud (Schmidt, 2009). In the long term, they ensure the stability of our knowledge (Hüffmeier et al., 2016; Radder, 1996). However, historical evidence suggests that social psychology has not (a) valued (Giner-Sorolla, 2012; Hüffmeier et al., 2016), (b) incentivized (Koole & Lakens, 2012), or (c) conducted replications (Makel et al., 2012; Schmidt, 2009). It follows that, *in the absence* of a research culture that facilitates conducting and appreciating replications, we know too little about whether individual findings, or the field as a whole, are trustworthy.

Relevance of Journals' Expressed Support for Replications

At first glance, replications may appear to garner greater appreciation in social psychology today. For example, different systematic replication efforts (e.g., the "Many Labs" projects; e.g., Ebersole et al., 2016; Klein et al., 2014) were published in prestigious journals, like *Advances in Methods and Practices in Psychological Science*, *Journal of Experimental Social Psychology*, and *Social Psychology*. In addition, various researchers advocate for making replications mainstream (Hüffmeier et al., 2016; Zwaan et al., 2018).

However, even if individual researchers or entire labs begin to appreciate and conduct replications, these efforts may not be enough to initiate broad cultural change (Nosek et al.,

2022). Researchers' career advancement is often influenced by the quantity of publications, especially those in highly prestigious journals (Aguinis et al., 2020). Publications in such outlets can be even more important than research quality, especially when it comes to promotion, tenure, and reward decisions (Gervais et al., 2015; Gomez-Mejia & Balkin, 1992). In turn, journals decide what gets published: They determine which type of studies have a high chance of getting published and, thus, are particularly appealing for researchers to conduct. As such, journals function as gatekeepers for what type of research is valued and rewarded, including replications (Kepes et al., 2018).

For example, the *Journal of Personality and Social Psychology* declined to send a replication (Ritchie et al., 2012) of Bem's (2011) precognition research out to peer-review, offering the following explanation: "This journal does not publish replication studies, whether successful or unsuccessful" (Aldhous, 2011). Hence, journals' expressed support for replications, or lack thereof, probably has a strong impact on whether researchers conduct replications. As Giner-Sorolla (2012, p. 566) put it: "The unglamorous nature of replication work, confronted with the narrow publishing bottleneck, makes much of it unpublishable, and therefore not worth starting, in a world of precarious careers and limited resources." Therefore, a sustainable and broad cultural change might only be achieved if stakeholders change policies and create incentives for conducting replications (Nosek et al., 2022).

Are Replications Mainstream now?

Social psychology's "manifest" replication crisis came with a "big bang" and was covered by neighboring fields as well as the media (for an overview, see Pashler & Wagenmakers, 2012). However, errors enable learning processes (Zhao, 2011). Negative emotions related to one's errors trigger a motivation to learn from them (Zhao, 2011), and negative media coverage and critical scrutiny of neighboring disciplines most likely led to negative emotions among social psychologists. Further, errors with major (vs. minor) negative consequences lead to even more learning because they attract more attention (i.e., "negative

outcome bias", Zakay et al., 2004, p. 151). As public trust in social psychology findings and in its scientific community could strongly diminish in view of low replication rates, the field's errors clearly had the potential to trigger severe negative consequences.

This collective awakening was followed by intense discussions about methods to improve the field's publication practices—for example, the introduction of the TOP guidelines, including replications (Nosek et al., 2015). Taken together, the field of social psychology had (a) the motivation (i.e., impending negative consequences), (b) the tools (i.e., preregistrations or replications), and (c) the time (i.e., one decade) to learn from its errors (i.e., the low appreciation of replications; Zakay et al., 2004; Zhao, 2011;) and to work towards the needed change in its scholarly culture (Pashler & Wagenmakers, 2012; Spellman et al., 2017). Therefore, over time, the field of social psychology may have changed its replication policies to address its replication problems. Thus, we hypothesized the following:

> *Hypothesis:* Replications are welcomed[1] more often in social psychology journals in 2022 than they were in 2015.

Overview of the Investigation

We first examined whether expressed support for replication studies in the policies of social psychology journals has changed from 2015 to 2022. We coded journal policies in 2015 and used this as a benchmark date because the publication of a large-scale replication study by the Open Science Collaboration (2015) that year made replication widely visible. We recoded those same journals' policies in 2022 to assess whether they were responsive to the advocacy to increase cultural acceptance of replication research. If explicit support for replications is still missing, it could hint that social psychology upholds its dysfunctional culture (Nosek et al., 2022). Second, we estimated in an exploratory fashion how many articles were published in social psychology journals that dealt with replications, relative to all published articles

[1] Due to the suggestion of a reviewer, we revised the wording of this hypothesis. However, this revision neither changed the content of the prediction itself, nor did it influence any corresponding analyses or results.

comparing articles published before and after 2015 (cf. Makel et al., 2012). We also compared expressed support for replications on the journals' websites and the number of articles published in the examined journals that dealt with replications. This can provide insight into whether any observed change in policy is associated with actual change in publishing replication-related research.

Third, we examine the general adoption of TOP-compliant policies (Nosek et al., 2015) in social psychology journals in 2022. This examination complements our data with an independent coding of present journal policies, and it provides an opportunity to characterize the policy landscape for replication in comparison with other transparency- and rigor-enhancing policies. Thus, the corresponding analyses show whether the development of expressed support for replication studies differs from the development of other open science policies.

Method

We preregistered this project on the OSF: https://osf.io/urf2c/.[2] Supplemental materials, including the full dataset, are publicly available at https://osf.io/n5mbq/.

Journal Selection

Our study compared the expressed support for replications in the policies of social psychology journals in 2015 and 2022. We analyzed the journals' stated policies on their websites. For the data from 2022, we based our content analysis on all journals that were listed in the 2021 *Journal Citation Reports* ($N = 65$), for the journal category of "Psychology, Social." For the data from 2015, we included all journals that were listed in the 2013 *Journal Citation Reports* in the same journal category. A summary of all included journals is provided in the Appendix. Journals were excluded from our analyses (a) if they exclusively published theoretical papers, meta-analyses, narrative reviews, or conference contributions ($n = 4$), (b) if

[2] We preregistered the data collection for the year of 2022, but not for the year of 2015, as we were not fully aware of the extant preregistration opportunities and the related benefits in 2015.

they did not publish in English or German (n = 1), or (c) if they were not included in our dataset for 2015 (i.e., if they were not included in the 2013 *Journal Citation Reports*, n = 9). The final sample included N = 51 journals.

Coding Procedure

For our coding of expressed support for replications, we examined the following sections on the journals' websites (i.e., an exhaustive coding): (a) aims and scopes, (b) author guidelines, (c) the general journal description, and (d) manuscript submission. We coded if and how journals mentioned replications using one variable with four levels: discouraged versus not mentioned versus considered secondary versus welcomed (see Table 1).

Table 1

Coding of Replications

Coding category
• Replications are mentioned, considered as equally important as original research and welcomed
• Replications are mentioned, but considered secondary
• Replications are not mentioned
• Replications are mentioned, but are not welcomed or are even discouraged

Note. "Considered secondary" means that journals mention replication studies, but consider them as secondary contributions and, for instance, publish them only online.

The data from 2015 were coded in January 2015 by two of the authors (κ = .873; two discrepancies were resolved through discussion). The data from 2022 were coded between February 7, 2022 and February 17, 2022. The first author and a second coder began by coding an initial set of five randomly selected journals. All discrepancies were discussed until consensus was reached. Afterwards, both coders coded all remaining journals independently. The resulting average interrater reliability was high (κ = .956; Landis & Koch, 1977). The one discrepancy that occurred was resolved through discussion. For exploratory purposes, we also collected (a) the journals' impact factor, (b) the journals' publisher, and (c) whether the

journals were signatories of the Committee on Publication Ethics (COPE) or mentioned that their guidelines are based on the COPE standards, which include "principles of transparency and best practice in scholarly publishing" (COPE et al., 2018).

Results

Confirmatory Analyses

We examined the expressed support for replications in 2022 and provide a comparison between 2015 and 2022. Table 2 shows whether and how social psychology journals mentioned replications in 2015 and 2022. In 2022, most social psychology journals (36 of 51; 71%) did not mention replications at all. There was one journal (1 of 51; 2%; *Organizational Behavior and Human Decision Processes*) that discouraged the submission of replication studies for publication in its section on "submission guidelines" as follows: "Significant contributions are less likely from research that merely replicates previous findings, revisits established findings using different samples or measures, or offers an incremental advancement to an existing body of knowledge" (information obtained in February 2022). Another journal (1 of 51; 2%; *Journal of Personality and Social Psychology*) mentioned replications in the coded sections, but considered them as secondary contributions (by indicating that it only publishes them online, see the section on "guide for authors"): "Replication manuscripts, if accepted, will be published online only and will be listed in the Table of Contents in the print journal" (information obtained in February 2022). The remaining journals that mentioned replication (13 of 51; 25%) expressed their support for replications. However, even these journals did not communicate a respective policy in their aims and scopes sections, arguably the most prominent part of their websites: In fact, only three of those 13 journals mentioned that replication studies were welcomed in their aims and scopes sections. One of those three journals specified that it focuses on original contributions, but is also open to replications (therefore, this instance could also have been reasonably coded as "replications being considered secondary"; see Table 1).

Table 2

Frequencies: Treatment of the Topic of Replications in Social Psychology Journals

	2015		2022	
	N	%	*N*	%
Discouraged	1	2	1	2
Not Mentioned	42	82	36	71
Considered Secondary	2	4	1	2
Welcomed	6	12	13	25
Total	51	100	51	100

In comparison to 2015, the number of journals that discouraged replications on their websites did not change. There were six fewer journals in 2022 that did not mention replications on their websites (36 in 2022 vs. 42 in 2015; see Table 2) and one journal less in 2022 that considered replications secondary (one in 2022 vs. two in 2015). Further, there were seven more journals in 2022 than in 2015 that welcomed replications on their websites (13 in 2022 vs. 6 in 2015). Thus, the number of journals that mentioned openness to replications (i.e., welcomed or considered secondary) increased from 8 to 14, indicating that replications were supported more often in 2022 than in 2015, which is consistent with our hypothesis.[3]

[3] We deviate from our pre-registration because a reviewer made us aware that there is no need for any inferential statistics for our analysis. Please note that, instead of a random sampling, we assessed the entire population of Social Psychology journals, and thus, sampling error does not seem to represent a problem in the current analyses. If we were looking at this as a sample of journals from a broader population (i.e., when also thinking about social psychology journals that are not part of the Journal Citation Reports), the amount of change would not be considered different than what might be expected due to sampling error alone.

Exploratory Analyses

As preregistered, we explored possible relationships between journals' expressed support for replications in 2022 and (a) the journals' impact factor, (b) whether the journals were a signatory of COPE, and (c) the journals' publisher (e.g., Wiley or Taylor & Francis). We found a small and positive correlation between journals' expressed support for replications (coded as follows: 0 = discouraged, 1 = no mention, 2 = secondary, 3 = supported) and the journals' impact factor, Kendall-Tau-b r = .14. Journals with a higher impact factor were slightly more likely to mention support for replications on their websites. Exploratory analyses with COPE and publisher variables are in the Appendix.

We next explored whether journals' support for replications (vs. not) was associated with an increased rate of publishing replication-related research, and whether this rate was different before versus after 2015.[4] In September 2022, using Web of Science, we searched the entire publication history of each of the 51 coded social psychology journals, which are part of the Journal Citation Reports, to identify (a) the total number of articles published and (b) the number of articles that contained the search term "replicat*" in the search category "topic" (i.e., any articles containing words with the stem of "replicat" in their title, abstract, author keywords, or keywords plus; see Makel et al., 2012, for a first example of this approach). The replication rate of each journal, estimating the percentage of articles that discussed replications, was calculated as follows: number of articles containing "replicat*" as divided by the total number of published articles. This analysis was conducted separately for two time periods (i.e., 1945-2015 and 2016-2022) to examine whether the replication rate changed over time (Table 3). As we did not examine whether the obtained 3,582 hits that were retrieved with the term "replicat*" were in fact replications (cf. Makel et al., 2012), our estimates of the replication rates might have been overestimated (i.e., including false

[4] Our reviewers kindly suggested conducting these additional searches. We thank the reviewers for this helpful suggestion!

positives). Note that this search strategy is inclusive of papers that discuss replication prominently but do not, themselves, report replication studies.

Overall, from 1945 to 2022, the term “replicat*” was used in 4.5% (3,582 of 79,655) of articles, with specific journals ranging from 0% to 22.3% (see the online supplement for individual journal data). However, in the period from 2016 to 2022, the term “replicat*” was used more (i.e., 5.6%, 1,574 of 28,195 articles) than in the period from 1945 to 2015 (i.e., 3.9%, 2,008 of 51,460 articles), a 43% increase. Usage of the term “replicat*” was more common in the years after 2015 than before across these journals, however, the term was used in a relatively small portion of all articles in both time periods.

To examine potential differences between journals expressing support versus no support for replications on their websites, we further estimated replication rates separately for the four groups of journals: Journals that (a) discouraged/did not mention replications in 2015 and 2022, (b) considered replications secondary in 2015 and did not mention replications in 2022, (c) became supporters of replications in 2022, and (d) were supporters of replications in 2015 and 2022. Interestingly, we observed that journals consistently expressing support for replications on their websites (group d), in fact, appeared to publish more articles dealing with replications than journals consistently not expressing support (group a). We found this pattern for both time periods, 1945-2015 (i.e., 5.9% vs. 2.8% for journals that did vs. did not express support for replications on their websites both in 2015 and 2022, respectively) and 2016-2022 (i.e., 11.8% vs. 3.7% for journals that did vs. did not express support for replications on their websites both in 2015 and 2022, respectively). Thus, journals that expressed support for replications on their websites appeared more likely to actually publish replications.

Further, the overall increase in the ratio of published replications was present in all four groups of journals. However, journals that were supporters of replications already in 2015 (group d), showed a steeper increase over time than journals that consistently discouraged/did not mention replications (group a) or than journals that became new

supporters of replications (group c; i.e., an increase of 100% vs. 32% vs. 30% for journals that were consistent supporters vs. consistently discouraged/did not mention replications vs. became supporters, respectively).

Table 3

Articles That Mentioned "Replicat" (N = 3,582)*

		Time periods						
		1945-2015			2016-2022			
	N journals	*N* articles	*N* "replicat*"		*N* articles	*N* "replicat*"		increase
Journals that …	… discouraged/did not mention replications in 2015 and 2022 (*N* = 36)	*N* = 25,319	*N* = 707	2.8%	*N* = 15,734	*N* = 586	3.7%	32 %
	… considered replications secondary in 2015 and did not mention replications in 2022 (*N* = 1)	*N* = 1,085	*N* = 50	4.6%	*N* = 570	*N* = 52	9.1%	97 %
	… became supporters of replications (*N* = 7)	*N* = 13,658	*N* = 578	4.2%	*N* = 7,402	*N* = 407	5.5%	30 %
	… were supporters of replications in 2015 and 2022 (*N* = 7)	*N* = 11,398	*N* = 673	5.9%	*N* = 4,489	*N* = 529	11.8%	100 %
Total		*N* = 51,460	*N* = 2,008	3.9%	*N* = 28,195	*N* = 1,574	5.6%	43 %

Note. "N journals" indicates the number of journals within each group. "*N* articles" indicates the total number of published articles. "*N* 'replicat*'" indicates the number of articles that used the term "replicat*" in their title, abstract, or keywords. Percentages are estimated replication rates (i.e., number of articles containing "replicat*" as divided by the total number of articles), and the increase in the replication rate from 2015 to 2022.

As an additional exploratory analysis, we examined the adoption of TOP-compliant policies for replication and other open science practices (Nosek et al., 2015) in social psychology journals in 2022 by examining the journals' TOP Factor. The director of policy at the Center for Open Science (COS), David Mellor, described the TOP factor as "a modular set of indicators of journal policies to facilitate the visibility of good research practices" (COS, 2020). The TOP Factor is based primarily on the eight standards[5] from the TOP guidelines (Nosek et al., 2015) and, thus, evaluates the degree to which journals support transparency and reproducibility. It assesses the degree to which a journal adopts each of the eight standards, and two other initiatives, Registered Reports (Chambers & Tzavella, 2022) and open science badges (Kidwell et al., 2016), on four increasing levels (i.e., Level 0 to Level III). The maximum score that can be achieved when a journal adopts all standards at the highest level is a TOP factor of 30. We collected the journals' TOP Factor from the open database of TOP Factor codings (https://topfactor.org/). There was a TOP Factor rating available for $N = 40$ of our included journals. The TOP Factor ratings for the missing 11 journals were coded by the first author for the ensuing analysis.

On average, social psychology journals had a TOP factor of $M = 5.92$, $SD = 6.56$. The lowest observed TOP factor was $Min = 0$ and the highest was $Max = 23$. Overall, there were 13 journals (25%) with a TOP Factor of 0 and 11 journals (22%) with a TOP Factor of 1. Three journals (6%) received a TOP Factor of or above 20. Table 4 shows the journals' TOP Factor rating concerning replications. Figure 1 offers a visualization of replication policy adoption in comparison with the other guidelines. Of the 51 included journals, 38 (75%) discouraged submission of replication studies or said nothing about it (i.e., TOP Factor level "Not Implemented"). Four journals (8%) encouraged submissions of replication studies (i.e., TOP Factor level I). One journal (2%) encouraged submissions of replication studies and

[5] The eight standards of the TOP guidelines comprise: citation standards, data transparency, analytic methods (code) transparency, research materials transparency, design and analysis transparency, preregistration of studies, preregistration of analysis plans, and replication.

conducted results-blind reviews (i.e., TOP Factor Level II). Finally, eight journals (16%) used Registered Reports as a submission option for replication studies with peer-review prior to observing the study outcomes (i.e., TOP Factor Level III). Taken together, the TOP Factor rating for replications (i.e., 75% not implemented) reveals a similar picture as our coding of the journals' support for replications (i.e., 71% not mentioned).

Table 4

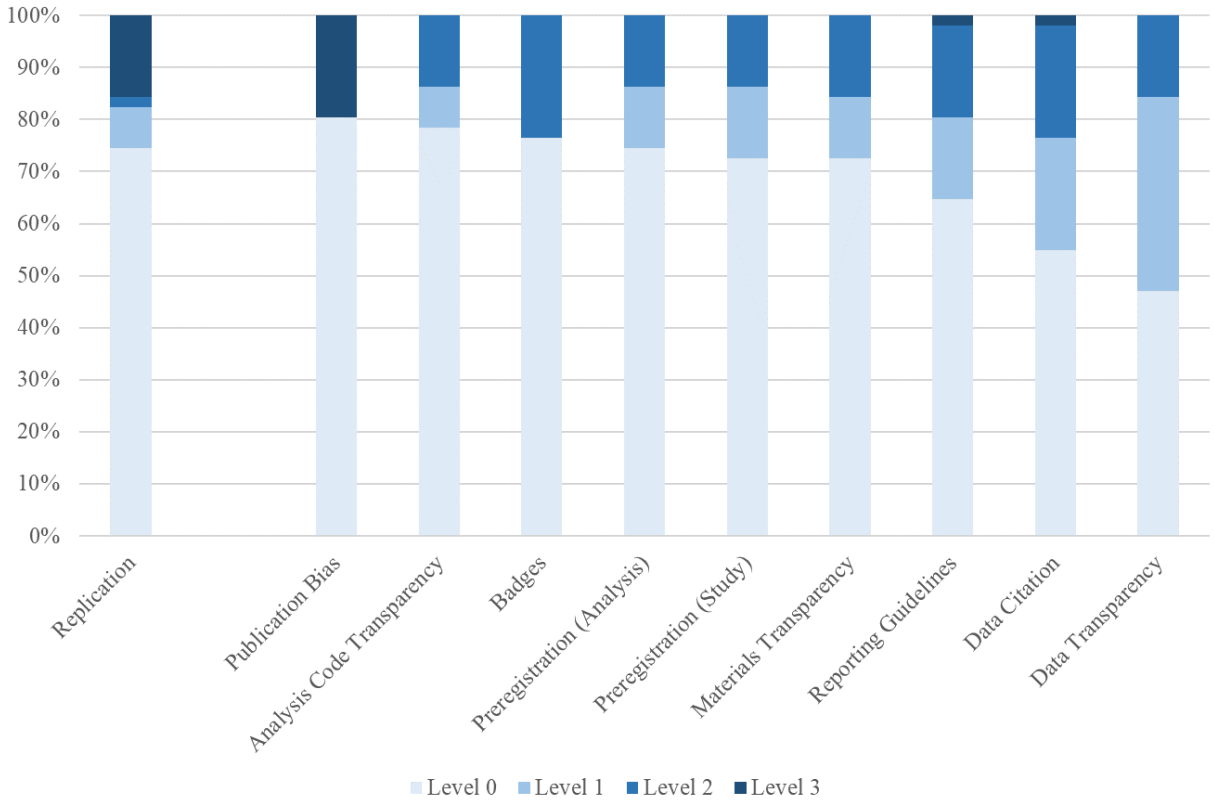
TOP Factor Scores of Social Psychology Journals for Replication

	TOP Factor	
	N	%
Not Implemented: Journal discourages submission of replication studies, or says nothing about it	38	75
Level I: Journal encourages submission of replication studies.	4	8
Level II: Journal encourages submission of replication studies and conducts results blind review.	1	2
Level III: Journal uses Registered Reports as a submission option for replication studies with peer review prior to observing the study outcomes.	8	16
Total	51	100

Note. Percentages may not total 100 due to rounding.

Figure 1

TOP Factor Scores of Social Psychology Journals



Note. Policies are contrasted with "Replication" as our reference point (depicted as the left-most bar). The remaining policies are ordered by the proportion of journals adopting the policy at any level. Visualization adapted from Nosek et al. (2022).

Finally, to explore whether newer journals potentially express more support for replications and, thus, represent a separate and more appreciative "market" for replication studies, we coded the policies of some prominent psychological open access journals that were created near or after 2015 in response to the call to increase rigor and transparency of research (i.e., *Collabra: Psychology*, *Comprehensive Results in Social Psychology*, *Meta-Psychology*, *Personality Science*, and *Social Psychological Bulletin*). We observed that *all* of these journals welcomed replications on their websites, and their average TOP Factor score of 22.4 (range from 18 to 27) was well-above the average of the social psychology journals included in our study. New journals, created, in part, to improve credibility of research, have a strongly positive stance toward open science and replications specifically. As a group of

journals, they may represent an emerging alternative market for replications because most of the more “traditional” journals still did not explicitly welcome replications.

Discussion

Theoretical Implications and Future Theoretical Implications and Future Research

Our study examined whether the expressed support for replication studies in policies of social psychology journals changed between 2015 and 2022. Our hypothesis was supported: Replications were welcomed more often in social psychology journals in 2022 than in 2015. However, expressed support for replications is still represented by a significant minority of journals. Most journals did not update their policies after 2015 and still did not mention replications at all on their websites. Further, even those journals that generally expressed support for replications mostly did not mention them as a possible publication option in their most visible sections (i.e., the aims and scopes).

Our exploratory search on the amount of published replication studies suggests that journals that mention and support replications on their websites are in fact more likely to publish replications. Further, journals that expressed support for replications on their websites in 2015 and 2022 showed a larger increase in their replications rate during that time period than journals that started supporting replications sometime between 2015 and 2022, or did not explicitly support replications in 2015 or 2022. Those journals that had a positive public stance toward replications in 2015 may have been particularly attractive to researchers interested in replication, and changing policies may take some time to bring about change in scholarly practice. Further, the analyzed group of mainstream social psychology journals are comparatively slow to adapt their guidelines compared with journals such as *Collabra* and *Social Psychological Bulletin* that emerged during the period with the aim, in part, to improve research credibility. The introduction of these journals is possibly a reaction to the slow changes of more traditional journals and appears to fulfill the extant need to have journals that support the publication of replications. Altogether, our findings suggest that the field of social

psychology may *still* be characterized by a "dysfunctional social culture" (Nosek et al., 2022, p. 733), in which replications are not considered ordinary scientific practice. This applies to most traditional journals, and is qualified by some change among journals, appearance of more discussion of replication across all journals regardless of policy, and emergence of more progressive journals promoting strong open practices including replication.

Our exploratory analyses on the TOP Factor suggests that journal adoption of language supporting replication occurs at similar rates as other rigor and transparency promoting policies such as data sharing or preregistration. Many journals have adopted some TOP-compliant policies, but a majority still have no or minimal such policies. As such, the adoption of TOP-compliant policies is still coming along slowly. To better understand and change this state of affairs, future research could investigate the journal publishers', editors', and boards' reasons to not actively support replications among other open science practices in their policies.

Practical Implications

If journals, as important gatekeepers, do not support and incentivize replications, then researchers may simply not conduct them. As a result, the actual "latent" crisis underlying the "manifest" crisis—a field that does not value replications, resulting in a lack of replication studies—cannot be addressed. In turn, the insufficient appreciation of replications threatens to leave unclear which of our scientific findings are reliable and trustworthy. Thus, the field (i.e., researchers, reviewers, editors, journals, publishers) may need to think again or more strongly about the role played by replications to support its renaissance as a stronger field (Nelson et al., 2018).

Limitations

Our article contains a number of limitations. For instance, we interpreted missing statements about replications as journals not supporting and valuing replications, which may overstate the extent of the problem. However, if a journal does not mention replications on its

website at all, researchers are at least not encouraged to conduct and submit a replication to this journal, as it remains unclear whether a replication has a good chance of being published. Lacking communication about replications is noteworthy because journals often mention welcomed article types (e.g., reviews, meta-analyses, or quantitative and qualitative studies). Further, our additional search for articles containing the term replicat* in the title, abstract, or keywords (e.g., Makel et al., 2012) suggests that there is a relationship between expressed support for replications on social psychology journals' websites and how often replications were mentioned in their published articles. Thus, missing statements about replications seem to be associated with fewer published replications. Still, there is some ambiguity in the interpretation of missing statements about replications.

Using the term replicat* does not mean that replication studies were included in the paper, just that replicat* appeared in key locations in the paper – the title, abstract, or keywords. It is plausible that use of the term is associated with the likelihood that replication studies are part of the paper, but future research could parse between talking about versus doing replication research.

The TOP Factor ratings we used for our exploratory analysis are provided by individual coding of community members. Different researchers can complete a journal evaluation form on the TOP Factor website, which is then reviewed and added to the database. Therefore, the TOP Factor scores for different journals are recorded at different times (ranging from December 2019 to September 2022 for our sample). Thus, journals may have updated their standards in the meantime. However, our own coding of the journals' replication policies in 2022 is not affected by this limitation and paints a very similar picture of the journals' policies.

Conclusion

Our findings underscore the slow change toward a replication culture in a substantial portion of mainstream social psychology journals. We observed a modest change over time in

their expressed support for replications and amount of prominent discussion of replication in published articles. These findings overall reveal an enduring lack of encouragement as communicated by current journal policies. Thus, the field's norms and practices regarding replications, and other open science practices, do not seem to have substantially changed in the last decade—replications still appear to be far from becoming mainstream.

References

Aguinis, H., Cummings, C., Ramani, R. S., & Cummings, T. G. (2020). "An a is an a": The new bottom line for valuing academic research. *Academy of Management Perspectives, 34*(1), 135–154. https://doi.org/10.5465/amp.2017.0193

Aldhous, P. (2011, May 5). *Journal rejects studies contradicting precognition.* Retrieved from https://www.newscientist.com/article/dn20447-journal-rejects-studies-contradicting-precognition/

Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business and Psychology, 31*(3), 323–338. https://doi.org/10.1007/s10869-016-9456-7

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality & Social Psychology, 100*(3), 407–425. https://doi.org/10.1037/a0021524

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

Collins, H. (1985). *Changing order: Replication and induction in scientific practice.* University of Chicago Press.

COPE, OASPA, DOAJ, & WAME. (2018, January 15). *Principles of transparency and best practice in scholarly publishing*. Retrieved from https://doi.org/10.24318/cope.2019.1.12

COS (2020, February 10). *New measure rates quality of research journals' policies to promote transparency and reproducibility*. Retrieved from https://www.cos.io/about/news/new-measure-rates-quality-research-journals-policies-promote-transparency-and-reproducibility

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., … 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

Gervais, W. M., Jewell, J. A., Najle, M. B., & Ng, B. K. (2015). A powerful nudge? Presenting calculable consequences of underpowered research shifts incentives toward adequately powered designs. *Social Psychological and Personality Science, 6*(7), 847–854. https://doi.org/10.1177/1948550615584199

Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science, 7*(6), 562–571. https://doi.org/10.1177/1745691612457576

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/10.5465/256535

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

Hüffmeier, J., Mazei, J., & Schultze, T. (2016). Reconceptualizing replication as a sequence of different studies: A replication typology. *Journal of Experimental Social Psychology, 66*, 81–92. https://doi.org/10.1016/j.jesp.2015.09.009

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., List, S. K., & McDaniel, M. A. (2018). Enough talk, it's time to transform: A call for editorial leadership for a robust science. *Industrial and Organizational Psychology*, *11*(1), 43–48. https://doi.org/10.1017/iop.2017.83

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., & 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., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … 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

Koole, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science*, *7*(6), 608–614. https://doi.org/10.1177/1745691612462586

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

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

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, *69*, 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buch, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … & Yarkoni, T. (2015). Promoting an open research culture. *Science, (348)*6242, 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D. , Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology, 73*, 719-748. https://doi.org/10.1146/annurev-psych-020821-114157

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

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

Radder, H. (1996). *In and about the world: Philosophical studies of science and technology*. SUNY Pr.

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). https://doi.org/10.1371/journal.pone.0033423

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*(2), 90–100. https://doi.org/10.1037/a0015108

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

Spellman, B., Gilbert, E., & Corker, K. S. (2017). *Open science: What, why, and how.* Working paper. Retrieved from https://doi.org/10.31234/osf.io/ak6jr

Zakay, D., Ellis, S., & Shevalsky, M. (2004). Outcome value and early warning indications as determinants of willingness to learn from experience. *Experimental Psychology, 51*(2), 150–157. https://doi.org/10.1027/1618-3169.51.2.150

Zhao, B. (2011). Learning from errors: The role of context, emotion, and personality. *Journal of Organizational Behavior*, *32*(3), 435–463. https://doi.org/10.1002/job.696

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*. Article E120. https://doi.org/10.1017/S0140525X17001972

Appendix

Table A1

Overview of Included Journals with Coded Variables

No	Journal	Impact Factor	COPE	Replications	
				2015	2022
	Journals that changed their policy				
1	Analyses of Social Issues and Public Policy	1.16	1	not mentioned	welcomed
2	Journal of Personality	5.12	1	not mentioned	welcomed
3	Journal of Social and Personal Relationships	3.04	1	not mentioned	welcomed
4	Law and Human Behavior	3.80	1	not mentioned	welcomed
5	Personality and Individual Differences	3.00	1	not mentioned	welcomed
6	Social Cognition	1.82	0	not mentioned	welcomed
7	Social Psychology Quarterly	1.76	1	not mentioned	welcomed
8	Journal of Personality Assessment	3.78	1	considered secondary	not mentioned

	Journals that did not change their policy				
9	Cultural Diversity & Ethnic Minority Psychology	3.23	1	welcomed	welcomed
10	European Journal of Personality	5.84	1	welcomed	welcomed
11	Journal of Experimental Social Psychology	3.60	1	welcomed	welcomed
12	Journal of Research in Personality	3.07	1	welcomed	welcomed
13	Personality and Social Psychology Bulletin	4.38	1	welcomed	welcomed
14	Social Psychology	2.47	1	welcomed	welcomed
15	Journal of Personality and Social Psychology	7.67	1	considered secondary	considered secondary
16	Organizational Behavior and Human Decision Processes	4.94	1	discouraged	discouraged
17	Asian Journal of Social Psychology	1.42	1	not mentioned	not mentioned
18	Basic and Applied Social Psychology	1.53	1	not mentioned	not mentioned
19	British Journal of Social Psychology	4.69	1	not mentioned	not mentioned

20	Child Abuse & Neglect	3.93	1	not mentioned	not mentioned
21	Cyberpsychology Behavior and Social Networking	4.16	1	not mentioned	not mentioned
22	Deviant Behavior	1.98	1	not mentioned	not mentioned
23	European Journal of Social Psychology	3.38	1	not mentioned	not mentioned
24	Group Dynamics: Theory, Research, and Practice	0.76	1	not mentioned	not mentioned
25	Group Processes & Intergroup Relations	3.13	1	not mentioned	not mentioned
26	International Journal of Intercultural Relations	2.67	1	not mentioned	not mentioned
27	Journal of Applied Social Psychology	2.12	1	not mentioned	not mentioned
28	Journal of Community & Applied Social Psychology	2.37	1	not mentioned	not mentioned
29	Journal of Cross-Cultural Psychology	2.62	1	not mentioned	not mentioned
30	Journal of Diversity in Higher Education	2.70	1	not mentioned	not mentioned
31	Journal of Health and Social Behavior	4.46	1	not mentioned	not mentioned

32	Journal of Individual Differences	2.08	0	not mentioned	not mentioned
33	Journal of Language and Social Psychology	2.25	1	not mentioned	not mentioned
34	Journal of Loss & Trauma	1.06	1	not mentioned	not mentioned
35	Journal of Nonverbal Behavior	2.94	1	not mentioned	not mentioned
36	Journal of Social and Clinical Psychology	1.95	0	not mentioned	not mentioned
37	Journal of Social Psychology	2.71	1	not mentioned	not mentioned
38	Kölner Zeitschrift für Soziologie und Sozialpsychologie	0.82	1	not mentioned	not mentioned
39	Motivation and Emotion	2.34	1	not mentioned	not mentioned
40	Personal Relationships	1.90	1	not mentioned	not mentioned
41	Personality and Mental Health	3.82	1	not mentioned	not mentioned
42	Political Psychology	4.33	1	not mentioned	not mentioned
43	Psychology of Men & Masculinities	2.95	1	not mentioned	not mentioned

44	Research on Language and Social Interaction	3.08	1	not mentioned	not mentioned
45	Revista de Psicologia Social	0.62	1	not mentioned	not mentioned
46	Self and Identity	3.30	1	not mentioned	not mentioned
47	Sex Roles	4.15	1	not mentioned	not mentioned
48	Small Group Research	2.66	1	not mentioned	not mentioned
49	Social Behavior and Personality	0.98	1	not mentioned	not mentioned
50	Social Influence	0.93	1	not mentioned	not mentioned
51	Social Justice Research	1.44	1	not mentioned	not mentioned

Note. Codings are as follows.
Impact Factor: 2020 journal impact factor as listed in the 2021 Journal Citation Reports;
COPE: 0 = journal (or its publisher) is neither a signatory of COPE nor does is mention COPE, 1 = the journal (or its publisher) is a signatory of COPE or mentions that its guidelines are based on the COPE standards;
Replications: discouraged = replications are mentioned in the respective sections, but are not welcomed or are even discouraged; not mentioned = replications are not mentioned in the respective sections; considered secondary = replications are mentioned in the respective sections, but considered secondary; welcomed = replications are mentioned in the respective sections, considered as equally important as original research and welcomed.

Table A2

Expressed Support for Replications in 2022, Divided by the Journals' Publisher and COPE (N = 51)

	COPE		Publisher									
Replication	No Signatory	Signatory	Wiley	Taylor & Francis	SAGE	APA	Elsevier	Springer	Guilford Press	Hogrefe	Mary Ann Libert, Inc.	Scientific Journal Publishers Ltd
Discouraged	0	1	0	0	0	0	1	0	0	0	0	0
Not Mentioned	7	29	8	9	5	3	2	5	1	1	1	1
Considered Secondary	1	0	0	0	0	1	0	0	0	0	0	0
Welcomed	4	9	3	0	3	2	3	0	1	1	0	0
Total	12	39	11	9	8	6	6	5	2	2	1	1

Note. Discouraged = replications are mentioned in the respective sections, but are not welcomed or are even discouraged; Not mentioned = replications are not mentioned in the respective sections; Considered Secondary = replications are mentioned in the respective sections, but considered secondary; Welcomed = replications are mentioned in the respective sections, considered as equally important as original research and welcomed.

Additional Analyses

Going off Script: Exploring the Reporting of Preregistration Deviations in Industrial and Organizational Psychology and Their Relationship with QRPs

Ann-Kathrin Torka and Joachim Hüffmeier

TU Dortmund University

Abstract

The current study aims to take stock of, encourage, and contribute to establishing preregistrations as a standard to increase transparency in Industrial and Organizational Psychology (IOP). To do so, we examined to what extent IOP research is preregistered and adheres to preregistrations. We assumed that only little IOP research is preregistered and that many of the related studies may deviate from their preregistration without transparently reporting it. In 27 IOP journals, we found that only $N = 91$ articles had a preregistration (i.e., 0.70% of all studies published between 2017 and 2023). Drawing a random sample of 25 of these preregistered studies, we examined deviations from preregistrations and their transparent disclosure. The studies in our IOP sample deviated on average $M = 1.60$ ($SD = 1.61$) times from their preregistrations. Put differently, sixty-four percent of the studies deviated in at least one aspect, and 60% had at least one undisclosed deviation, indicating that deviations from preregistrations are common. Importantly, the vast majority of these deviations was not linked to Questionable Research Practices. We further compared the IO sample with a recent sample from *Psychological Science* because the journal is an early adopter of preregistration. We found no difference between the two samples in their number of (undisclosed) deviations, which suggests that preregistering can be done effectively even without being a standard in a field (yet). We provide an overview of valuable resources for preregistrations and provide best practice recommendations to transparently report any deviations.

Keywords: preregistration, transparency, Open Science, Questionable Research Practices, Industrial and Organizational Psychology

Going off Script: Exploring the Reporting of Preregistration Deviations in Industrial and Organizational Psychology and Their Relationship with QRPs

The trustworthiness of the cumulative knowledge in Industrial and Organizational (IO) Psychology is currently at risk, just as in related fields (for an overview, see Keener et al., 2023). One factor contributing to this problem is the use of Questionable Research Practices (QRPs; O'Boyle & Götz, 2022). QRPs are researcher behaviors that can distort findings in favor of significant results. When using QRPs, researcher exploit their degrees of freedom to present biased evidence (Banks et al., 2016). Common QRPs include selectively reporting hypotheses, hypothesizing after the results are known (HARKing; Kerr, 1998), *p*-hacking (i.e., "report only [. . .] analyses [. . .] that 'work'" Simonsohn et al., 2014, p. 534), and cherry picking results (John et al., 2012). There is some evidence that the field of organizational research in general, and IO Psychology in particular, is affected by QRPs. For instance, a survey by Banks et al. (2016) found that many Management researchers admitted engaging in QRPs (e.g., 49.6% in HARKing and 33.3% in *p*-hacking). Similar practices have been reported among psychologists in general (John et al., 2012; Agnoli et al., 2017). Consistent with HARKing and selective outcome reporting, journal articles in IO Psychology and Management often report more supported and fewer unsupported hypotheses than the underlying dissertations (Mazzola & Deuling, 2013; Kepes et al., 2022).

A suitable Open Science practice to reduce the use of QRPs and to increase transparency consists in preregistering study hypotheses and analysis plans (Banks et al., 2019; Kepes & McDaniel, 2013). By doing so, researchers make their research plans transparent and publicly accessible, allowing others to scrutinize and verify the research process. This helps to enhance the reproducibility (i.e., two researchers achieve identical results by analyzing the same data using the same methodology, Asendorpf et al., 2013) and

trustworthiness of scientific findings (Nosek et al., 2018). Research on preregistration has shown that it can help increase transparency and rigor in scientific reporting (Toth et al., 2021; see also Scheel et al., 2023), and that researchers perceive it to improve the quality of their projects and research processes (Sarafoglou et al., 2022).

However, it is currently unclear to which extent studies in IO Psychology are actually preregistered and how strongly authors deviate from their preregistrations. A recent study found that only one of 244 coded studies from four leading journals in organizational sciences (i.e., *Journal of Applied Psychology, Journal of Management, Journal of Organizational Behavior,* and *Personnel Psychology*) had preregistered its hypotheses (Crede & Sotola, 2024). Thus, preregistration rates in IO Psychology might be rather low. Moreover, there is an ongoing debate about when and how to deviate from a preregistration (Lakens, 2024). Deviations from a preregistered plan can be either unintentional or intentional, and they can either increase or decrease the quality of a study (Lakens, 2024). Previous research on the adherence to preregistrations in Psychology found that researchers generally deviated substantially from their preregistrations (Van den Akker et al., 2023).

Our study aims to advance the current research practice towards a broader implementation of preregistrations in IO Psychology. As our first contribution, we assess the current prevalence of preregistrations in IO Psychology, and the extent of disclosed versus undisclosed deviations in preregistered studies. Identifying a potentially low rate of articles with preregistration practices in IO Psychology could highlight the need for increased awareness of, education about, and interventions on preregistration. Moreover, understanding disclosure practices can inform strategies for improving transparency. As our second contribution, we illuminate and discuss whether undisclosed deviations in our sample and in IO Psychology generally are associated with QRPs. If we find associations of

deviations and QRPs, our study could provide a basis of efforts to improve research integrity and methodological rigor in IO Psychology. As our third contribution, we offer a comparison with a recent sample of articles from the journal *Psychological Science*, which is an early adopter of preregistrations. This comparison provides valuable insights in potential differences among fields of research. If IO Psychology strongly differs from other fields in terms of preregistration practices, these differences could be a starting point for initiatives to enhance preregistration adherence in IO Psychology. Conversely, if there is no significant difference between IO Psychology and this early adopter of preregistrations, the finding could encourage researchers who have not yet engaged with preregistrations by reducing the perceived barriers to preregistering. Finally, as fourth contribution, we provide an overview of relevant sources that provide hands-on advice on how to conduct them (e.g., preregistration forms) and on best practices for handling deviations from preregistrations (e.g., templates for deviations). This is crucial because providing clear, accessible tools and guidelines can lower the barrier to entry for IOP researchers, encouraging them to attempt a first preregistration. Facilitating the management of deviations (according with best practices for their reporting) can also increase transparency and credibility of research.

Theoretical Background

Preregistrations

In the last decade, Psychology was confronted with low reproducibility (Artner et al., 2021; Maassen et al., 2020) and low replicability (i.e., obtaining an equivalent finding with another random sample, Asendorpf et al., 2013; Open Science Collaboration, 2015; Pashler & Wagenmakers, 2012), resulting in the so-called "replication crisis" (for an overview, see Nosek et al., 2022). In response to this crisis, Nosek et al. (2015) introduced preregistrations as part of their Transparency and Openness Promotion (TOP) Guidelines. By preregistering their studies, researchers outline a detailed plan for their research projects before collecting

and analyzing data, thereby committing themselves to their hypotheses, methods, and analysis plans. The preregistration is then deposited in one of various online repositories, such as the Open Science Framework (OSF) or aspredicted.com.

By limiting researchers' degrees of freedom, preregistrations can reduce the use of QRPs. Overall, they increase transparency and aim to improve the credibility and reproducibility of scientific findings (Nosek et al., 2018). Indeed, Registered Reports (RRs), which are preregistered studies that undergo peer review before data collection, have been found to exhibit higher methodological and analytical rigor, as well as to possess higher overall quality than comparison papers that were no RRs (Soderberg et al., 2021). Further, as there are fewer supported hypotheses in preregistered than non-preregistered studies, they decrease the amount of false positive findings (Toth et al., 2021; see also Scheel et al., 2021). Taken together, preregistrations are perceived by many as "a game changer" (Simmons et al., 2021, p. 177). However, despite its potential to enhance research credibility, Hardwicke et al. (2022) for instance found that, in a sample of 250 randomly selected published psychological articles from 2014 to 2017, only 3% had a preregistration (for comparable results in top-tier organizational behavior journals from 2011 to 2019, see Tenney et al., 2021). Thus, as a first focus of our research, we are interested in the general amount of articles[1] with preregistration in IO Psychology. We accordingly pose our first Research Question (*RQ1)*: How many published articles in IO Psychology have a preregistration?[2]

Deviations from Preregistrations

Even researchers who preregister their studies might not always adhere to their preregistered plans. For instance, Van den Akker et al. (2023) assessed the extent of selective

[1] In the analyses of our manuscript, we differentiate between articles and studies: Articles are research manuscripts published in a scientific journal. With studies, we refer to individual studies published as part of a research manuscript, i.e., there are multi-study articles that include more than one individual study.

[2] As a first deviation from our preregistration (https://osf.io/g69v4/?view_only=5082510df627479ea803ffb089318488), we changed the wording of *RQ1* from „How many published articles in IO Psychology are preregistered?" to „How many published articles in IO Psychology have a preregistration?" to reflect that studies, but not articles as a whole are preregistered.

hypothesis reporting in Psychology by comparing preregistrations with the corresponding articles. In their sample, 52% and 57% of studies contained omitted and added hypotheses, respectively, and 18% included hypotheses that had undergone a change in direction. Furthermore, in their research on gambling studies with $N = 20$ investigated preregistrations, Heirene et al. (2021) found that 65% of the included studies exhibited undisclosed deviations. In a sample of $N = 27$ preregistered studies from the journal *Psychological Science,* even 89% included undisclosed deviations (Claesen et al., 2021).

Conversely, Toth et al. (2021) compared published preregistered studies against published non-preregistered studies in Applied Psychology and Management. They found that only 16% of all preregistered studies explicitly discussed some kind of deviations from their preregistration. However, as outlined above, research in other fields suggests that the number of deviations from preregistrations is usually much higher. Taken together, it appears possible that a considerable number of studies in IO Psychology deviate from their preregistrations without transparently reporting these deviations. Nonetheless, we currently cannot say anything about the true extent to which preregistered studies in fact deviate from their preregistrations and do or do not disclose it. Therefore, we are interested in the extent to which preregistered studies in IO Psychology (a) deviate from their preregistrations and (b) disclose these deviations. We therefore pose the following two research questions.

RQ2: How many preregistered studies in IO Psychology deviate from their preregistration in at least one aspect?

RQ3: How many preregistered studies in IO Psychology do not report at least one extant deviation from their preregistration in the published article?

Differences in Preregistration Practices between Different Fields of Research

Exploring how disclosed and undisclosed deviations from preregistrations might differ between different fields of research can help to gain deeper insights into current preregistration practices and the current preregistration quality in IO Psychology.

Understanding these differences can potentially also provide valuable insights into specific challenges within IO Psychology. Claesen et al. (2021) investigated the adherence to preregistrations and the disclosure of preregistration deviations in *Psychological Science*, which is an early adopter of Open Science practices generally and preregistrations in particular and only very rarely publishes articles from IO Psychology. The 27 studies they examined were conducted during a period when preregistration was a rather new Open Science practice in Psychology, spanning from February 2015 to November 2017. Thus, their sample represents what could be termed the "first generation" (Claesen et al., 2021, Title) of preregistrations. It is reasonable to assume that *Psychological Science* authors gained greater familiarity with preregistering studies over time, leading to improvements in the specificity of (i.e., "all phases of research process are described", Heirene et al., 2021, p. 4) and adherence to their preregistrations (a trend also noted in research on gambling, as observed by Heirene et al., 2021).

To investigate the deviations from preregistrations across different research fields, we will generally build on the approach of Claesen et al. (2021). However, we will include an updated, more recent sample of preregistered studies published in *Psychological Science*, as a benchmark for comparison, reflecting the current "state of the art" standard. For a fair comparison, we will match the publication year of the studies in the samples from IO Psychology and from *Psychological Science*. This allows us to contrast recent studies from IO Psychology with those from *Psychological Science*. We accordingly ask the following:

> *RQ4*: Are there more undisclosed deviations from preregistrations in published IO Psychology studies than in a random sample of currently published studies from *Psychological Science*?

Deviations from Preregistrations and Their Relationship with QRPs

While the full benefits and the rationale behind preregistering a study are realized only when researchers generally adhere to their predetermined plans, there are of course (and

often) circumstances that lead them to deviate from their preregistration. Many of these circumstances are not problematic at all (e.g., after preregistering their study, researchers learn about a methodologically more sound way to analyze their data). However, one problematic reason to deviate from a preregistration is that the study does not produce the desired empirical outcome. For example, a researcher might consciously decide to selectively omit preregistered dependent variables from the manuscript that do not support the hypotheses. In these cases, deviations from preregistrations are associated with QRPs.

However, as mentioned above, QRPs are not the only reason why researchers deviate from preregistered plans. Toth et al. (2021) explored various explanations for why studies deviate from their preregistrations, encompassing both minor and significant changes. They differentiate between changes that occurred (i) due to (unintentional) errors, (ii) due to adding to (e.g., adding variables) or slightly changing the procedure, (iii) due to updating sample size, (iv) due to changes made to the analysis plan, and (v) due to preregistration-introduced issues related to QRPs. As illustrated by this typology, deviations from a preregistration are often not a sign of problematic behavior, but researchers may learn new things they did not know at the time of the preregistration or there are deviations that are beyond their control. For instance, a smaller sample than preregistered might result if a cooperating organization ends data collection ahead of time. Other deviations can even improve the quality of a study compared to the preregistered plan, such as incorporating helpful comments from the review process or correcting an error in the preregistration (Willroth & Atherton, 2024). We agree with Toth et al. (2021) that different types of deviations from a preregistration can be associated with QRPs to different degrees. We will therefore also examine different types of deviations from preregistrations in IO Psychology with regards to their association with QRPs. We therefore pose our fifth research question as follows:

RQ5: How many undisclosed deviations from preregistrations in published IO Psychology studies might be (versus are likely not) associated with QRPs?

Method

Transparency and Openness

Below, we report how we determined our sample size, all data exclusions, and all measures (Simmons et al., 2012). The current study is preregistered. The preregistration of this study can be accessed via the following link: https://osf.io/g69v4/?view_only=5082510df627479ea803ffb089318488. Supplemental materials, including the full dataset, will be publicly available after submission of the article under https://osf.io/7v2ya/?view_only=38bd429f83314f52903d8c8fb3456e6b. All deviations from the preregistration are disclosed in the manuscript.

Sample

We base our analysis on the IO Psychology journals in the ABS Academic Journal Guide 2021 (https://charteredabs.org/academic-journal-guide-2021/), which generally includes 66 journals. However, we excluded journals that 1) are not published in English or German (e.g., *Gedrag en Organisatie*), 2) focus on theoretical contributions and commentaries rather than empirical research (e.g., *Organizational Psychology Review*), 3) have a topical rather than disciplinary focus (e.g., *Accident Analysis & Prevention*), or 4) focus on a discipline other than IO Psychology (e.g., Military: *Military Psychology*, or School Psychology: *Journal of School Psychology*). This resulted in a final sample of 27 journals (see Table 2).

In these journals, we first searched for all published studies performing an electronic keyword search within *Web of Science* (search string: IS=*insert journal ISSNs*). Afterwards, we searched for studies that were preregistered performing a second electronic keyword search within *Web of Science* (search string: IS=*insert journal ISSNs* AND ALL=prereg* OR OSF OR aspredicted OR registered OR pre-reg*). We deviated from our preregistration with this search string. As preregistered, we first used the following search string: IS=*insert journal ISSNs* AND ALL=prereg*. However, this search yielded only $N = 46$ hits (i.e.,

preregistered studies), which appeared (too) low to us. In order to not miss any preregistered studies, we expanded our search string as described above.

We then drew a random sample of 25 studies, while limiting each journal's prevalence in our sample to no more than five studies (the prevalence of preregistered studies was very unequal across journals, see below). To limit each article to no more than one included study, we drew one random study from multi-study papers (where applicable). To do so, for each multi-study paper, we used a random number generator to generate one random number in the range of the total number of studies. The study that matched the generated number was selected for inclusion. To summarize, it was our aim to include $N = 25$ preregistered studies from IO Psychology in our analyses.

Further, we drew a sample of 25 additional studies from *Psychological Science* as a "state of the art baseline". To rule out the possibility that potential differences between this sample and the sample from IO Psychology are due to the year in which they were published (i.e., authors may have gained more knowledge about preregistering studies over time and improved the specificity of and adherence to their preregistrations), we matched the studies from *Psychological Science* to the included studies from IO Psychology regarding year of publication (i.e., we matched both sub-samples according to the year of publication). More specifically, for each IO Psychology study included in our sample, we searched for articles with preregistered studies in *Psychological Science* that were published in the same year (search string: IS= 0956-7976 AND ALL=prereg* AND PY=*insert year*)[3] and then drew a random article (that was not already included in the sample). Again, to limit each article to no more than one included study, we drew one random study from multi-study papers using a random number generator. To summarize, it was our aim to include $N = 25$ preregistered studies from *Psychological Science* in our analyses that match the sample from IO

[3] Contrary to the search for articles with preregistration from IO Psychology, these searches yielded a sufficient number of suitable articles for each year.

Psychology regarding the publication year. To ensure objective coding, we excluded our own preregistered studies from the samples.

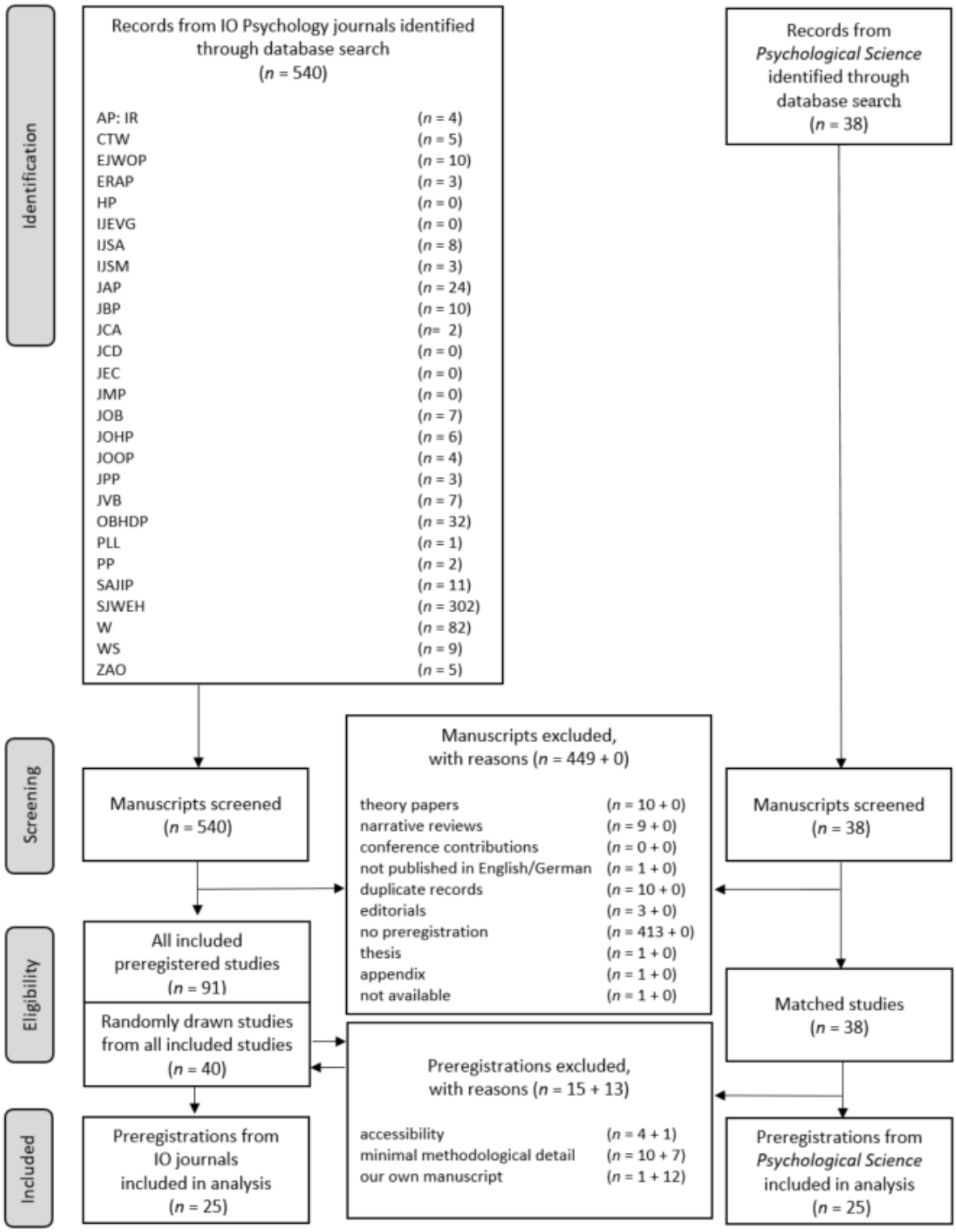
Exclusion Criteria

Articles were excluded from our analyses if they were theory papers, narrative reviews, or conference contributions, or if they were not published in English or German. Our broad search string yielded many false positive hits (e.g., studies about *registered* nurses). Thus, as a second deviation from our preregistration, we also excluded all articles that did not include at least one preregistered study. Preregistered studies were excluded from our analyses following the exclusion criteria formulated by Claesen et al. (2021, p. 3):

> "Both exclusion criteria for the adherence assessment, accessibility and minimal methodological detail, were assessed using six items each that were scored with either 0 (unfulfilled) or 1 (fulfilled). The items for accessibility were the following: a preregistration plan should be a (i) permanent, (ii) read-only, (iii) time-stamped, (iv) public, (v) non-ambiguously available, and (vi) on a third-party repository stored document. If not all items were scored with a 1, the plan was excluded. Minimal detail was assessed for the remaining preregistration plans by evaluating whether they contained some information on the following methodological aspects: (i) a research question and/or hypothesis, (ii) a list of variables, (iii) operationalization of the variables, (iv) a sample size, (v) a procedure, and (vi) a statistical model. Again, if not all items were present, the plan was excluded."

However, we noticed that it is not possible for some studies to specify a sample size in advance (e.g., when working with archival data). In these situations, it was sufficient for our minimal detail assessment to have preregistered inclusion criteria or a stopping rule. For a detailed overview of included and excluded manuscripts, see the flowchart in Figure 1.

Figure 1
Flowchart of Identification and Selection of Preregistrations



Coding Procedure

Our first criterion variable (see *RQ1*) is the number of published articles with a preregistration in IO Psychology. We provide an estimate of published articles with a preregistration in IO Psychology as well as an estimate of the proportion of articles with a preregistration in IO Psychology (i.e., number of published articles with a preregistration from IO Psychology divided by the number of all published articles from IO Psychology). Our second (see *RQ2*) and third (see *RQ3*) criterion variables are (a) the proportion of studies that deviate from their preregistration in at least one aspect (i.e., number of studies from IO Psychology with at least one deviation from their preregistration divided by the number of all studies from IO Psychology included in our sample), and (b) the proportion of studies that deviate from their preregistration but do not report it in the published article (i.e., number of studies from IO Psychology with at least one *undisclosed deviation* from their preregistration divided by the number of all studies from IO Psychology included in our sample).

To examine whether there are deviations from the preregistrations in a study and how these deviations are reported versus not reported, we followed the approach by Claesen et al. (2021) and compared the following aspects of each preregistration with the final (published) study: (i) hypotheses/research questions, (ii) variables/covariates, (iii) sample size, (iv) exclusion criteria, (v) procedure, and (vi) analysis (exploratory analyses will not be included in the comparison). For each aspect, we coded whether there were no deviations, only disclosed deviations, or at least one undisclosed deviation.[4] We then used these codes to form three count variables ranging from zero to six (because we looked at six aspects of each preregistration [i-vi], see above): (a) number of aspects with no deviations, (b) number of aspects with only disclosed deviations, and (c) number of aspects with undisclosed deviations.

[4] We did not preregister our approach for handling meta-analyses and reviews. We included them, but coded sample size criteria as not applicable.

This number of aspects with undisclosed deviations from preregistrations in the IO studies serves as another criterion variable (see *RQ3*).

We describe how adherence to the preregistration was coded for each of these aspects (adapted from Claesen et al., 2021) in Table 1. In the case of reasonable doubt about a deviation from the preregistration plan, this was coded as no deviation. In line with Claesen et al. (2021), accessibility was rated for all preregistrations, minimal methodological detail was rated for all preregistrations that passed the accessibility criteria, and adherence was assessed for all preregistrations that passed the exclusion criteria.

Table 1

Coding of Deviations from the Preregistrations

Aspects	No deviations from the preregistration	Deviation(s) from the preregistration
Hypotheses/ Research Questions	The hypotheses/research questions (RQs) are either exactly the same as preregistered, or they do at least not contradict the preregistered hypotheses/RQs and analyses. For example, they contain more detail that does not contradict the preregistered hypotheses/RQs, or the wording is changed to be more precise or had to be changed due to an error in the preregistration, etc.	The hypotheses/RQs are not the same as preregistered, insofar that they are not in line with the preregistered hypotheses/RQs. For example, hypotheses/RQs were deleted or added, the direction of a predicted effect shifted, etc.
Variables/ Covariates	The variables/covariates are exactly the same as preregistered. If the wording is changed to be more precise or due to an error in the preregistration (e.g., the name of a variable was changed while the operationalization stayed the same), this does not constitute a deviation.	The variables/covariates are not the same as preregistered. For example, dependent variables were deleted (i.e., not reported although they were preregistered) or added (i.e., reported although they were not preregistered).
Sample Size	The sample size and/or stopping rule is the same as preregistered. When it is not clear in the preregistration whether the planned sample size is before or after exclusions, it must match with either the sample size before or after exclusions in the article. Deviating from Claesen et al. (2021) a sample size slightly lower (i.e., -5%) or slightly higher (+5%) than preregistered will not be considered as a deviation if it is reasonable that (termination of) data collection was not based on significance tests (e.g., if the study was run by an external provider [prolific or amazon turk], because those external providers are known to run studies with some, usually small, deviations from the originally determined sample size). However, if only a minimum is preregistered, but the obtained sample size is more than 5% higher than this preregistered minimum, this will be considered a deviation.	The reported sample size and/or stopping rules is not the same as was preregistered.

Exclusion Criteria	The applied exclusion criteria are the same as preregistered. We do not consider implicit exclusion criteria as deviations, for example, if authors write "the sample consisted of US adults" while non-US residents and minors were the preregistered exclusion criteria. If no exclusion criteria were preregistered, there are also no exclusions in the published article.	The exclusion criteria are not the same as preregistered. If no exclusion criteria were preregistered, but there were exclusions in the published article, this is a deviation. Because of terminology, exclusion criteria may appear to vary, while they are in fact the same (e.g., living in the US is technically not the same as being in the US, but this is not a deviation). Non-preregistered criteria are deviations, but this does not hold for implicit exclusion criteria, for example "the sample consisted of psychology students."
Procedure	The procedure is the same as the preregistered procedure, regardless of whether the authors provide more or less detail in the article.	Large changes in the procedure (including the design, e.g., changing from a within- to a between-subjects design).
Analyses	The model is either exactly the same, or contains less or more detail that does not contradict the original analyses. Terminology may be inconsistent.	Changes to the model, such as for instance extra covariates in a regression, using other variables as dependent/predictor variables than preregistered, etc. Deviating from Claesen et al. (2021), extra analyses due to extra hypotheses are also coded as deviations.

For each aspect that deviates from the preregistration, we also coded whether the deviations were disclosed or not disclosed. We coded an aspect as "all deviations disclosed" if the authors clearly acknowledge all deviations in the main text, the notes, the supplemental material, and/or the open practices disclosure (e.g., by writing "we deviate from our preregistration by doing XY"). By contrast, at least one undisclosed deviation in an aspect was coded as present if the authors do not clearly acknowledge all deviations in the main text, the notes, the supplemental material, and/or the open practices disclosure. If a deviation is only visible through a shared statistical analysis code (i.e., there is no explicit mention of this deviation), this will also be coded as undisclosed deviation.

Finally, as additional analysis, we aim to obtain first evidence on the relationship between disclosed and undisclosed deviations in IO Psychology and QRPs. Therefore, for each aspect with disclosed and undisclosed deviations from the preregistration that we found, we also coded whether the deviations may be (versus are likely not) associated with QRPs (see also Heirene et al., 2021; Toth et al., 2021). Deviations that are likely not associated with QRPs are, for instance, changes due to errors in the preregistration, or having to terminate data collection due to extenuating circumstances. By contrast, deviations that may be associated with QRPs are, among others, failures to report preregistered hypotheses, reporting of non-preregistered hypotheses, altering of preregistered hypotheses [i.e., HARKing], adding or dropping data/conditions/covariates/dependent variables to avoid null results, rounding p-values, or modifying measures after data collection.

Results

Preregistrations in IO Psychology

In the following, to answer how many published articles in IO Psychology have a preregistration (*RQ1*), we provide descriptive statistics for an estimate of the number of published articles with preregistration in IO Psychology as well as of the proportion of published articles with preregistration (i.e., number of published articles with preregistration

in IO Psychology divided by the number of all published articles in IO Psychology). We found that only $N = 91$ articles in our IO Psychology sample had a preregistration at all. These articles were published starting from 2017 with three exceptions (one article was published in 2012 and two were published in 2013). Figure 2 gives an overview of the number of articles with preregistration in IO Psychology per year.

Interestingly, the 91 articles with preregistration from IO Psychology were published in only 13 different journals (i.e., roughly half of our sample of 27 included journals, see Table 2). Specifically, more than half of the articles (i.e., $n = 49$ articles, 53.85%) were published in only two different journals (i.e., *Organizational Behavior and Human Decision Processes*, and *Journal of Applied Psychology*). As shown in Figure 2, there is an increase of articles with preregistration in IO Psychology in the last years, with a peak of 31 articles with preregistration published in 2023.

Table 2

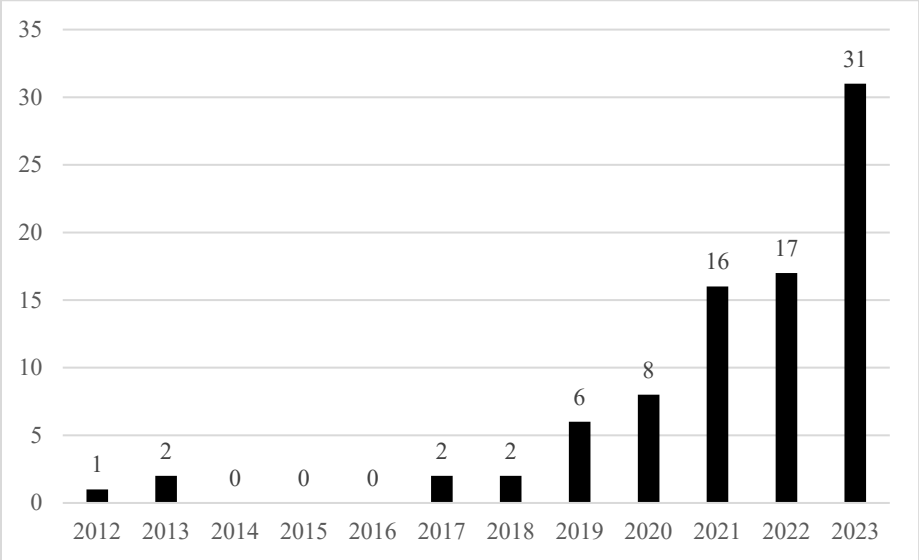
List of Included IO Psychology Journals and the Number of Articles with Preregistration

Title	Number of articles with preregistration
Applied Psychology: An International Review	3
Cognition, Technology and Work	1
European Journal of Work and Organizational Psychology	5
Human Performance	0
International Journal for Educational and Vocational Guidance	8
International Journal of Selection and Assessment	0
International Journal of Stress Management	0
Journal of Applied Psychology	18
Journal of Business and Psychology	6
Journal of Career Assessment	0
Journal of Career Development	0
Journal of Employment Counseling	0
Journal of Managerial Psychology	0
Journal of Occupational and Organizational Psychology	3
Journal of Occupational Health Psychology	1
Journal of Organizational Behavior	0
Journal of Personnel Psychology	1
Journal of Vocational Behavior	0
Organizational Behavior and Human Decision Processes	31
Personnel Psychology	2
Psychologist-Manager Journal[a]	0
Revue Europeenne de Psychologie Appliquee[b]	0
Scandinavian Journal of Work, Environment and Health	10
South African Journal of Industrial Psychology	0
Work	2
Work and Stress	0
Zeitschrift für Arbeits- und Organisationspsychologie	0

Note. [a] Name changed to Psychology of Leaders and Leadership; [b]French articles excluded

Figure 2

Number of Articles with Preregistration in IO Psychology per Year

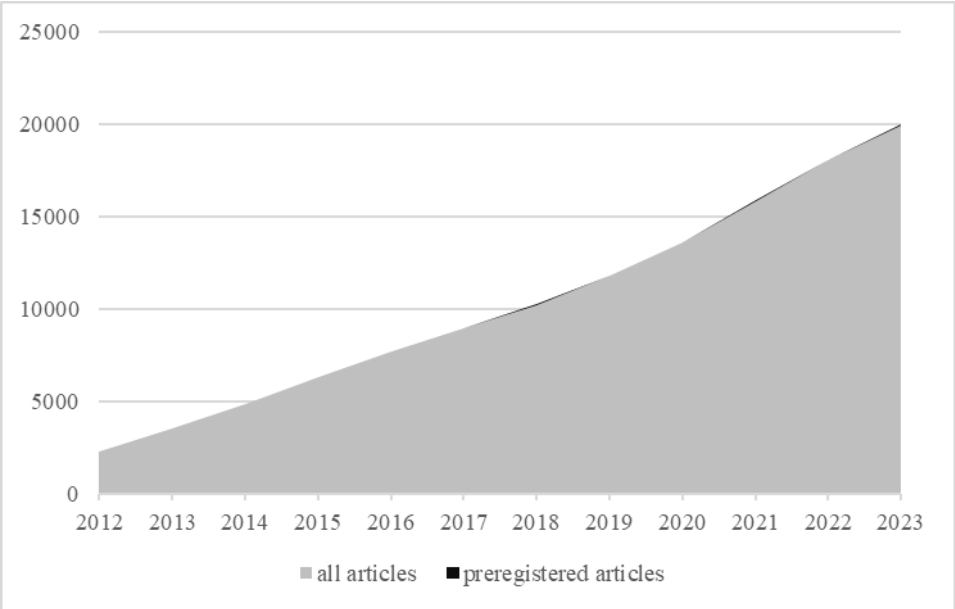


Note. Data for 2024 are not depicted as the year has not yet concluded.

However, the proportion of articles with preregistration among all published articles is still negligible. Only 0.70% of all studies published between 2017 (i.e., the year of the first articles with preregistration after the replication crisis became apparent in 2015) and 2023 have a preregistration. Figure 3 and Table 3 show the development of the proportion of articles with preregistration in IO Psychology.

Figure 3

Stacked Area Chart of Articles in IO Psychology, Divided into Articles without and with Preregistration



Note. Data for 2024 are not depicted as the year has not yet concluded.

Table 3

Proportion of Articles in IO Psychology with Preregistration

year	published articles cumulated *N*	articles with preregistration cumulated *N*	articles with preregistration cumulated %
2017	1268	5	0.39
2018	2535	7	0.28
2019	4072	13	0.32
2020	5901	21	0.36
2021	8153	37	0.45
2022	10325	54	0.52
2023	12227	85	0.70

Note. The data for the *N* of published articles and the *N* of articles with preregistration are aggregated as cumulative totals. For instance, there were a total of *N = 1268* articles published in 2017, and a combined total of *N* = 2535 articles published in both, 2017 and 2018. Further, data for 2024 are not depicted as the year has not yet concluded.

Deviations from Preregistrations in IO Psychology

In view of *RQ2*, we found that 16 of the 25 included studies from the IO Psychology sample (i.e., 64%) deviated from their preregistration in at least one aspect. We found eight

deviations in the aspect hypotheses/RQs and eight in analyses, seven in variables/covariates and seven in sample size, five in procedure, and five in the exclusion criteria aspect. Figures 4 and 5 offer more detailed insights in the distribution of deviations across different methodological aspects (Figure 4) and across the included studies (Figure 5).

Figure 4

Overview of Adherence per Methodological Aspect (IO Psychology)

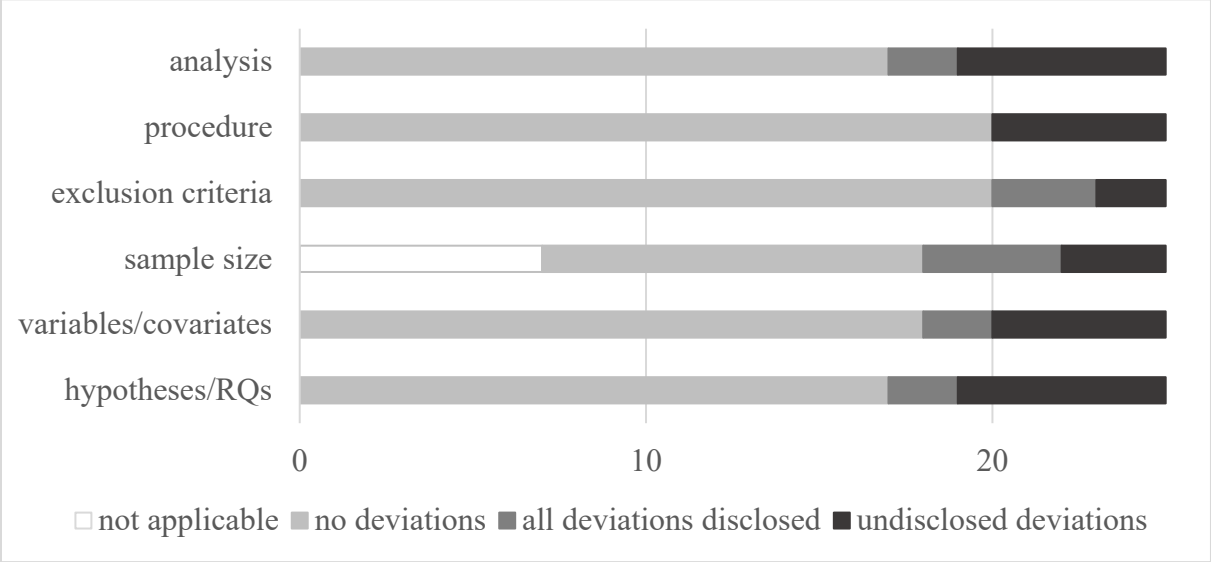
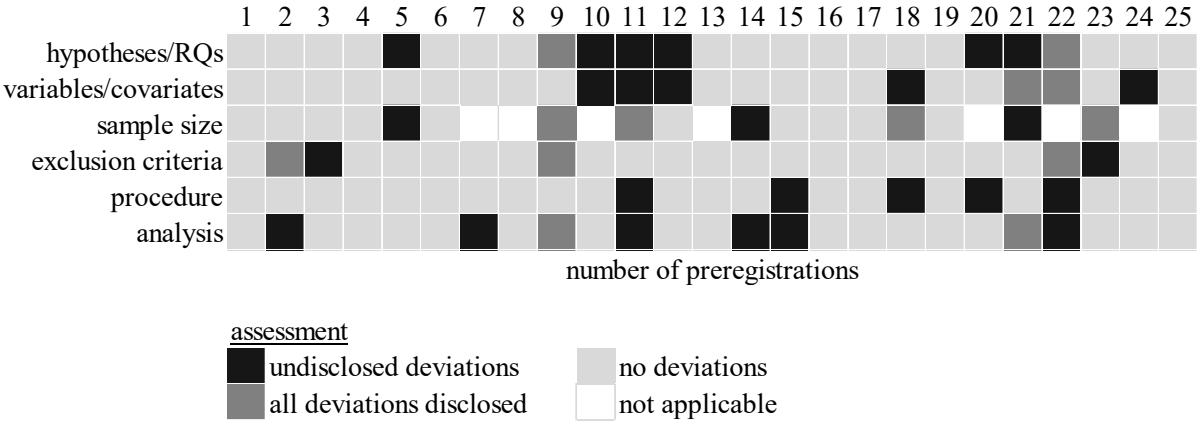


Figure 5

Tile Plot of the Assessment of Each Methodological Aspect per Study (IO Psychology)



To answer our third and main research question *RQ3* (i.e., *How many preregistered studies in IO Psychology do not report at least one extant deviation from their preregistration in the published article?*) we first provide descriptive statistics for the proportion of studies that deviate from their preregistration but do not report it in the published article (i.e., number of studies from the IO Psychology sample with at least one undisclosed deviation from their respective preregistrations divided by the number of all studies included in our IO Psychology sample). Second, we will also provide descriptive statistics of the specific areas where the studies deviate from their preregistrations without reporting it (i.e., hypotheses/RQs, variables/covariates, sample size, exclusion criteria, procedure, or analysis). Overall, there were 15 studies from our IO Psychology sample (i.e., 60%) with at least one undisclosed deviation from their preregistration. We found six undisclosed deviations in the aspect hypotheses/RQs and six in analyses, five in variables/covariates and five in procedure, three in sample size, and two in the exclusion criteria aspect (cf. Figure 5). Taken together, the studies from our IO sample had on average $M = 1.60$ ($SD = 1.61$) aspects with deviations, divided into $M = 0.52$ ($SD = 1.05$) aspects with disclosed deviations, and $M = 1.08$ ($SD = 1.08$) aspects with undisclosed deviations.

Comparing the IO Psychology and the *Psychological Science* samples

In the following, we compare our findings from IO Psychology with our recent study sample from *Psychological Science* to answer *RQ4*: *Are there more undisclosed deviations from preregistrations in published IO Psychology studies than in a random sample of currently published studies from Psychological Science?* In our sample of preregistered studies from *Psychological Science*, 21 of the 25 reviewed studies (i.e., 84%) deviated from their preregistration in at least one aspect. Further, there were 16 studies (i.e., 64%) with at least one undisclosed deviation from their preregistration. Overall, the studies had on average $M = 1.40$ ($SD = 1.00$) deviations, divided into $M = 0.36$ ($SD = 0.57$) disclosed deviations, and $M = 1.04$ ($SD = 0.98$) undisclosed deviations. As preregistered, we conducted a chi-square

test[5] to compare the number of undisclosed deviations in preregistered, published IO Psychology studies with the number of undisclosed deviations in current, published, and preregistered *Psychological Science* studies. We found no difference between the two samples in their number of undisclosed deviations, Fisher's exact test $p = .408$. As additional, non-preregistered and exploratory analyses, we also found no difference between the two samples concerning the number of studies that deviated from their preregistration in at least one aspect, $\chi^2(1) = 2.60$, $p = .107$, nor concerning the number of disclosed deviations, Fisher's exact test $p = .742$. We offer more detailed insights into the distribution of deviations across different methodological aspects and across the included *Psychological Science* studies in the online supplement (SF1 and SF2).

Undisclosed Deviations' Relationship with QRPs

Finally, we were interested in how many disclosed and undisclosed deviations from preregistrations in published IO Psychology studies might be (versus are likely not) associated with QRPs. As described in the methods section, we coded six aspects for each of our 25 included studies from IO Psychology for deviations versus no deviations (see Table 1). Thus, in total, we examined 150 aspects and found a total of 40 deviations, divided into 27 undisclosed and 13 disclosed deviations. Importantly, the vast majority of these deviations is clearly not linked to QRPs. Most deviations were due to unforeseen circumstances (i.e., not being able to recruit the preregistered sample size), learning (i.e., choosing a more suitable statistical analysis), suggested changes during the review process (i.e., moving some hypotheses into the online supplement), or small errors in the preregistration (i.e., wording errors) or during data collection (i.e., participants completed the study more than once).

For the disclosed deviations, we found only one instance that could be interpreted as associated with QRPs. Specifically, an open ended research question from the preregistration

[5] As the expected frequencies in our sample were smaller than five (Field, 2013), we used Fisher's exact test (Fisher, 1922). However, results did not change when conducting a Pearson-chi-square test.

was transformed into a directional hypothesis in the manuscript. However, this change was clearly described in the manuscript as a result of a learning process after gaining a better understanding of the literature following the preregistrations. Regarding undisclosed deviations, we found three instances that could (or perhaps even should be) perceived as related to QRPs. In two cases, a preregistered predictor variable was omitted from the manuscript (and also not reported in an online supplement). In the third case, participants were excluded based on a manipulation check, leading to an exclusion of over a quarter of the data. Although such an exclusion criterion may be reasonable, it was not preregistered, and the deviation is not mentioned in the manuscript. Because the authors shared their data, we could rerun their analyses while including the previously excluded participants and found that the authors' results were no longer statistically significant. Consequently, this change in statistical significance, combined with the undisclosed deviation from the preregistration, could be interpreted as being associated with QRPs.

Discussion

We were interested in the general extent to which research in IO Psychology is preregistered, how it handles deviations, in differences between different fields of research, and the relationship between undisclosed deviations and QRPs. Despite a small increase of articles with preregistration on a low level in the last years, we found that research in IO Psychology is, with few exceptions, currently not preregistered. Although preregistration has been introduced to Psychology for almost a decade (Nosek et al., 2015), IO Psychology as a research field obviously has yet to adopt this Open Science practice. Future research should examine reasons and potential barriers that hinder researchers in IO Psychology from preregistering their studies (for a first survey in Psychology, see Spitzer & Müller, 2023).

It is, however, encouraging that IO Psychologists who *do* preregister their studies, do this just as well as Psychologists in other fields. Furthermore, our findings suggest that there are comparable rates of undisclosed deviations across different fields of research. We found

that 65% of our included IO Psychology studies had undisclosed deviations, the same rate as Heirene et al. (2021) found in their sample of gambling studies. Similarly, in our sample of studies published in *Psychological Science*, 64% included undisclosed deviations. In studies with "first generation preregistrations" in *Psychological Science,* Claesen et al. (2021) found that 89% included undisclosed deviations. Their sample included studies published from 2015 to 2017, while our sample included studies published from 2017 to 2024. This may indicate that researchers are more consequent in disclosing preregistration deviations (or/and in adhering to their preregistrations). However, we were also not as strict in coding some aspects as Claesen et al. (2021; for instance, we did not flag small deviations from the preregistered sample size as deviations, see our methods section). Another encouraging finding of our study is that, in general, the undisclosed deviations in IO Psychology do not seem to be associated with QRPs. At the contrary, most deviations were due to unforeseen circumstances or minor, harmless errors.

As mentioned before, most of the preregistered studies were published in only two journals: *Organizational Behavior and Human Decision Processes* and the *Journal of Applied Psychology*. Interestingly, those journals mention preregistrations prominently on their webpages. *Organizational Behavior and Human Decision Processes* refers to preregistrations in its guide for authors. Specifically, authors aspiring to publish in the journal need to attach a completed 5-Item Submission Checklist that includes the statement that all deviations from any preregistered plans are reported. However, even for the articles published here, we found some undisclosed deviations. Similarly, since November 2021, all articles submitted to the *Journal of Applied Psychology* must state whether the study was preregistered. The journals' support of and request for preregistrations therefore seem to increase the likelihood that authors will actually preregister their studies (for comparable results regarding replication studies, see Torka et al., 2023).

Practical Implications

Notably, the studies that adhered to their preregistration in every aspect were *not* only those studies that preregistered the bare minimum of details. We observed a wide range of specificity in the coded preregistrations (cf. Allen & Mehler, 2019; Bakker et al., 2020). Despite preregistering a study, researchers still have many degrees of freedom, such as preregistering only the dependent and independent variables without control variables, or preregistering variables without specifying their operationalization (cf. Heirene et al., 2021). Even vague preregistrations are, however, helpful as they reduce flexibility in making decisions *after* collecting and seeing the data, and increase the transparency of decisions (Simmons et al., 2021). Thus, we encourage IO psychologists to preregister at least the essential methodological detail, including hypotheses, dependent and independent variables (predictor and criterion variable), planned sample size, and planned analyses to test the hypotheses.

Furthermore, matching studies to their preregistration was sometimes challenging in our analysis, particularly for multi-study papers, for OSF projects with multiple preregistrations, or when the study name in the preregistration did not match with the name in the manuscript (e.g., Study 2 in the preregistration was labeled as Study 1 in the manuscript). Therefore, echoing others who conducted similar studies on preregistrations (Claesen et al., 2021; Heirene et al., 2021), we recommend to include a direct link to the preregistration (rather than the overall project page) in the manuscript.

As previously mentioned, authors in IO Psychology who preregister their studies appear to perform this task just as well as researchers in other fields. Thus, preregistering studies is probably as suitable and as easy to learn for researchers in IO Psychology as it is for researchers from other psychological disciplines even though the preregistration of studies is not yet a standard in IO Psychology. It may even be a motivating experience. Indeed, a survey of psychologists revealed that motivation to preregister again had increased, instead of

decreased, for most participants who had already preregistered before (Spitzer & Müller, 2023). To contribute to and support the establishment of preregistrations as a standard Open Science practice in IO Psychology, we compiled resources for conducting a preregistration in Table 4.

Table 4

Resources for Conducting a Preregistration

Content	Resources
Basic and helpful information about preregistration	• Nosek et al. (2018) • Nosek et al. (2019) • Simmons et al. (2021) • https://www.cos.io/initiatives/prereg
Survey studies about preregistrations	• Sarafoglou et al. (2022) • Spitzer & Müller (2023)
Preregistration forms and templates, e.g., • for quantitative and qualitative research • for systematic reviews and meta-analyses • for projects that use pre-existing datasets	• https://osf.io/zab38/ • https://aspredicted.org/
Recommendations for preregistering studies	• Claesen et al. (2021) • Heirene et al. (2021)
Information about withdrawing and updating preregistrations	• https://help.osf.io/article/152-withdraw-a-registration • https://www.cos.io/blog/preregistration-plan-not-prison • https://help.osf.io/article/110-introduction-to-updating • https://help.osf.io/article/112-starting-an-update
Recommendations on whether and when to deviate from a preregistration	• Lakens (2024) • Willroth & Atherton (2024)
Templates for deviations	• Willroth & Atherton (2024): https://docs.google.com/document/d/1m7k53z38w18AJe56ucftunnHuFM7wDlMFjpoGenwN6k/edit?usp=fatal-error-notifier-usp • Mellor et al. (2023): https://osf.io/yrvcg

As a key finding of our research, we found that almost two thirds of studies do deviate from their preregistration. Thus, it is common and normal to deviate, which may be due to

various reasons such as unforeseen circumstances, gaining new knowledge, or suggestions during the review process. This finding has practical implications for two audiences: authors on the one hand, and reviewers and editors on the other. Regarding authors, we hope that our findings contribute to alleviating their fear of (failures in) preregistrations and show that "pre-registrations are a plan, not a prison" (DeHaven, 2017, Title). Doing a preregistration does not bind researchers; they can still adapt their research for valid reasons. Indeed, there are three different stages of research when preregistration plans ca be modified.

First, prior to data collection, it is possible to withdraw a preregistration. On the OSF, for instance, withdrawing a preregistration is easily made via clicking on the "Withdrawing Button" and providing a justification for withdrawing. This will remove all content (e.g., the preregistration and attached files) from the OSF. However, the "Registration Overview" page will still be visible (containing title, contributors, DOI, etc.) with the additional information that this preregistration was withdrawn, and the author's justification for its withdrawal (Center for Open Science, 2023a, see also Table 4). Second, if it is *necessary*, a preregistration can be updated during data collection. This may be the case, for example, if a recruiting approach needs to be changed (e.g., "Incentives changed from $20 gift cards to $50 gift cards as the team was unable to obtain the pre-specified sample size"; Center for Open Science, 2023b). To accommodate these situations, the OSF, for instance, has updated its repository to allow authors to revise an existing preregistration. To do so, admin collaborators of a preregistration can visit their "My Registrations Page" and click the new "Update" button at the bottom of their registration card (Center for Open Science, 2023c, see also Table 4). The updated version will then be available alongside the initial preregistration.

Third, after data collection, any deviations from a preregistration can (and should) be transparently reported (DeHaven, 2017). As our analyses showed, deviations that are *not* clearly documented risk being interpreted as QRPs. Therefore, as best practice recommendation, reporting any deviations from a preregistration can increase transparency

and trust in the reported findings. Fortunately, there are already suggestions on how to do this (Lakens, 2024) and different templates for reporting deviations are available (e.g., Willroth & Atherton, 2024, see also Table 4). Taken together, preregistration is not meant to restrict researchers rigidly. Instead, it is designed to promote transparency and flexibility, allowing researchers to update their plans or acknowledge deviations as long as these changes are clearly documented.

The finding that deviating from a preregistration is common, has also practical implications for reviewers and editors. They should perceive deviations from preregistrations as a normal occurrence and thereby normalize "failure", rather than criticizing authors for them. In fact, their normalizing of these deviations can help create a more open research environment. Moreover, by communicating their understanding that deviations are part of the research process, reviewers and editors can encourage more researchers to adopt preregistrations.

Limitations

Although we preregistered a detailed coding scheme, the coding of adherence to a preregistration involved some subjectivity in judgments. In certain cases, we made a conscious choice not to code somewhat ambiguous decisions by authors as non-adherence. For instance, some exclusion criteria were not explicitly preregistered, such as excluding participants who (a) were unable to fully participate in the research study (e.g., they cancelled the study before completion or they were not able to participate due to developmental delays), (b) did not give informed consent, (c) did not correctly follow the study (due to their own fault, experimenter error, or technical error), or (d) completed the study more than once. In these cases, where situations occurred that were obviously not anticipated by the authors (and at least partly could not have been anticipated), it seemed too strict to code a deviation.

However, there were other instances where we made decisions to code somewhat ambiguous decisions by authors as non-adherence. For example, we coded a deviation when

the preregistered hypotheses were not reported in the manuscript, even if the discussion stated that preregistered hypotheses were tested. Given the subjectivity in the coding schema, having a second coder to estimate the inter-rater reliability would be beneficial.

Furthermore, there were instances where manuscripts deviated from their preregistration, but reported the planned aspects in a supplement. For instance, some hypotheses and analyses were only reported in the online supplement. If authors disclosed this, we coded these cases as "no deviation" because the preregistration was followed, albeit reported differently (i.e., not in the main manuscript). To be transparent concerning these decisions, we provide an overview of the related instances in our supplement (see SF3a and SF3b).

Conclusion

We hope that the results of our study encourage researchers in IO Psychology to try their first preregistration—after all, preregistering studies can be learned and done quite effectively even if a scientific field has not implemented preregistration as established practice. Additionally, we recommend that authors, reviewers, and editors perceive deviations from their preregistrations as normal and that authors are accordingly transparent about any deviations from their preregistrations. This transparency protects against any accusations of engaging in QRPs. Furthermore, doing a preregistration can also prevent a study from becoming a victim of CARKing (i.e., "critiquing after the results are known"; Nosek & Lakens, 2014, p. 138) and, on a broader scale, help our field build more robust and rigorous cumulative knowledge.

References

Allen, C., & Mehler, D. M. (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

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., Fieder, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., Vanaken, M. A. G., Weber, H., & 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

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

Bakker, M., Veldkamp, C. L., van Assen, M. A., Crompvoets, E. A., Ong, H. H., Nosek, B. A., Soderberg, C. K., Mellor, D., & Wicherts, J. M. (2020). Ensuring the quality and specificity of preregistrations. *PLoS biology*, *18*(12), Article e3000937. https://doi.org/10.1371/journal.pbio.3000937

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

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

Center for Open Science (2023b, May 5). *Withdraw a Registration*. OSF Support. https://help.osf.io/article/152-withdraw-a-registration

Center for Open Science (2023b, May 5). *Introduction to Updating (Registrations)*. OSF Support. https://help.osf.io/article/110-introduction-to-updating

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

Crede, M., & Sotola, L. K. (2024). All is well that replicates well: The replicability of reported moderation and interaction effects in leading organizational sciences journals. *Journal of Applied Psychology*. Advance online publication. https://doi.org/10.1037/apl0001197

DeHaven, A. (2017, May 23). *Preregistration: A Plan, Not a Prison*. COS. https://www.cos.io/blog/preregistration-plan-not-prison

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th Edition ed.). SAGE Publications Ltd.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, *222*, 309–368. https://doi.org/10.1098/rsta.1922.0009

Hardwicke, T. E., Thibault, R. T., Kosie, J. E., Wallach, J. D., Kidwell, M. C., & Ioannidis, J. P. A. (2022). Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014-2017). *Perspectives on Psychological Science, 17*(1). 239-251. https://doi.org/10.1177/1745691620979806

Heirene, R., LaPlante, D., Louderback, E. R., Keen, B., Bakker, M., Serafimovska, A., & Gainsbury, S. M. (2021). *Preregistration specificity & adherence: A review of preregistered gambling studies & cross-disciplinary comparison.* PsyArXiv. https://doi.org/10.31234/osf.io/nj4es

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

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*, Article 104005. https://doi.org/10.1016/j.actpsy.2023.104005

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), 1190–1208. 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

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

Lakens, D. (2024). When and How to Deviate from a Preregistration. *Collabra: Psychology*, *10*(1), Article 117094. https://doi.org/10.1525/collabra.117094

Maassen, E., van Assen, M. A., Nuijten, M. B., Olsson-Collentine, A., Wicherts, J. M., & Gnambs, T. (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

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

Mellor, D. T., Esposito, J., Hardwicke, T. E., Nosek, B. A., Cohoon, J., Soderberg, C. K., Kidwell, M. C., Clyburne-Sherin, A., Buck, S., DeHaven, A. C., & Speidel, R. (2023). *Transparent Changes Template Document*. Retrieved May, 21, 2024, from https://osf.io/yrvcg

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., . . . 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., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600-2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*(1), 719–748. https:// doi.org/10.1146/annurev-psych-020821-114157

Nosek, B. A., & Lakens, D. (2014). Registered reports. *Social Psychology*, *45*(3), https://doi.org/10.1027/1864-9335/a000192

O'Boyle, E. H., & Götz, M. (2022). Questionable research practices. In L. J. Jussim, J. A. Krosnick, & S. T. Stevens (Eds.), *Research integrity: Best practices for the social and behavioral sciences* (pp. 260–294). Oxford University Press. https://doi.org/10.1093/oso/9780190938550.003.0010

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

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

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), 211997. https:// doi.org/10.1098/rsos.211997

Scheel, A. M., Schijen, M. R., & 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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology, 26*, 4–7. http://dx.doi.org/10.2139/ssrn.2160588

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2021). Pre-registration is a game changer. But, like random assignment, it is neither necessary nor sufficient for credible science. *Journal of Consumer Psychology*, *31*(1), 177-180. https://doi.org/10.1002/jcpy.1207

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

Spitzer, L., & Mueller, S. (2023). Registered report: Survey on attitudes and experiences regarding preregistration in psychological research. *Plos one*, *18*(3), Article e0281086. https://doi.org/10.1371/journal.pone.0281086

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

Torka, A. K., Mazei, J., & Hüffmeier, J. (2023). Are replications mainstream now? A comparison of support for replications expressed in the policies of social psychology journals in 2015 and 2022. *Social Psychological Bulletin*, *18*, Article e9695. https://doi.org/10.32872/spb.9695

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

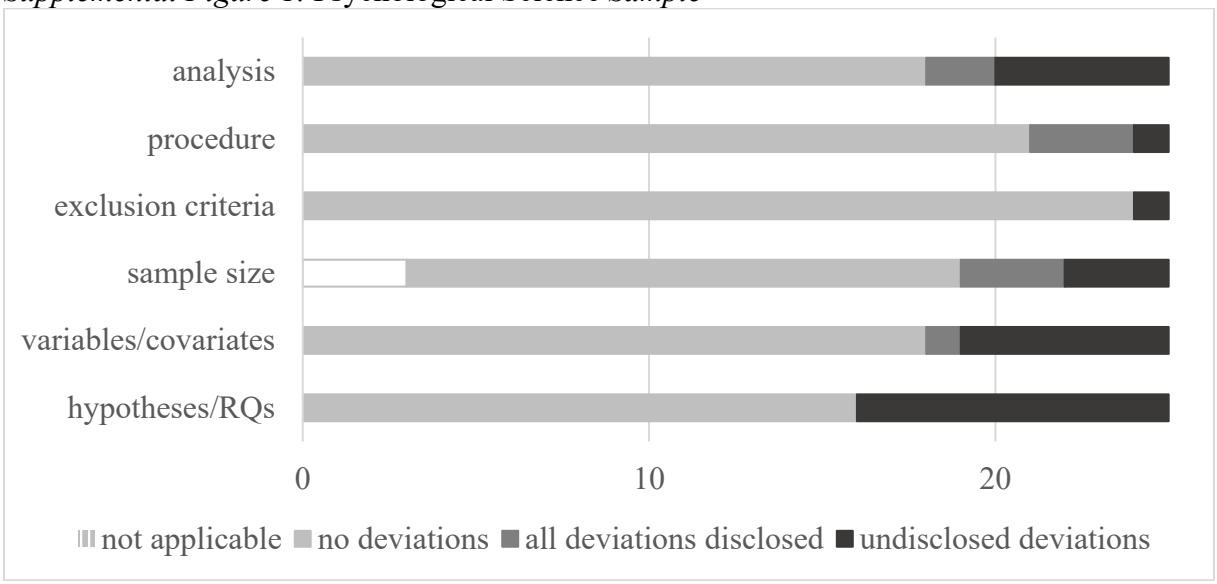
Van den Akker, O. R., van Assen, M. A., Enting, M., de Jonge, M., Ong, H. H., Rüffer, F., ... & Bakker, M. (2023). Selective hypothesis reporting in psychology: Comparing preregistrations and corresponding publications. *Advances in Methods and Practices in Psychological Science*, *6*(3). https://doi.org/10.1177/25152459231187988

Willroth, E. C., & Atherton, O. E. (2024). Best laid plans: A guide to reporting preregistration deviations. *Advances in Methods and Practices in Psychological Science*, *7*(1). https://doi.org/10.1177/25152459231213802

SUPPLEMENT

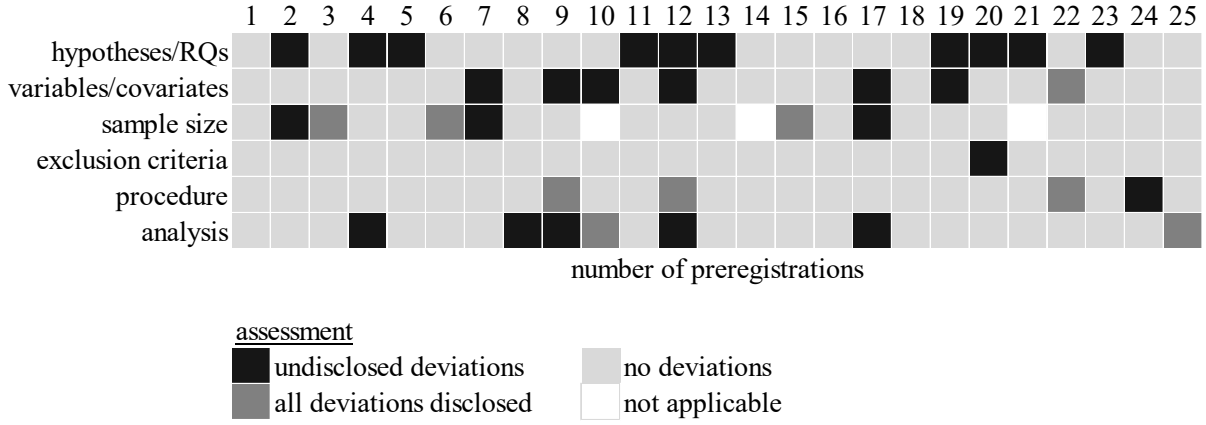
SF1

Supplemental Figure 1. Psychological Science *Sample*



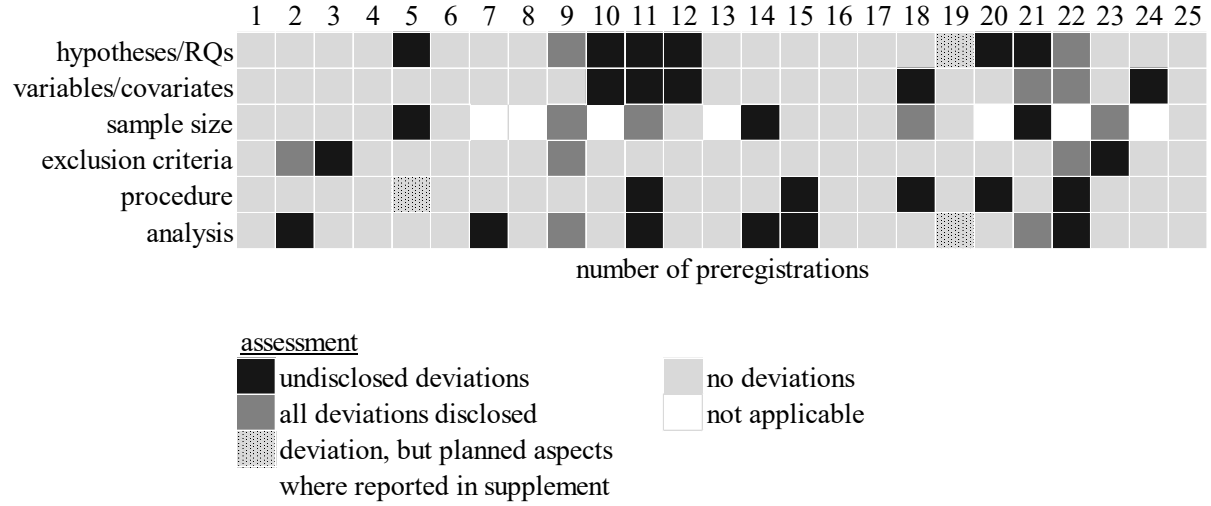
SF2

Supplemental Figure 2. Psychological Science *Sample*



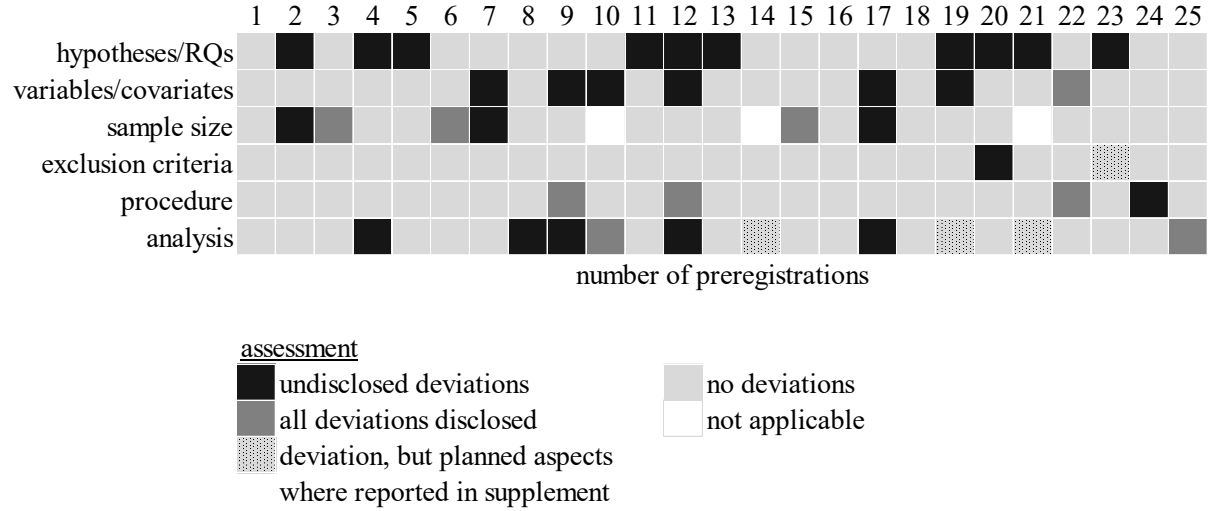
SF3a

Supplemental Figure 3a. IOP Sample with more Nuanced Coding



SF3b

Supplemental Figure 3b. Psychological Science *Sample with more Nuanced Coding*



General Discussion

In this doctoral dissertation, I aimed to advance theory and research on OSPs in Psychology by providing answers to five key questions. Specifically, I aimed to examine (I) how OSPs can help Psychology increase trustworthiness, whether OSPs are broadly implemented by (II) stakeholders and (III) researchers, (IV) whether the implementation of OSPs has changed over time, and (V) whether there are differences in the implementation of OSPs between psychological fields. These questions were addresses with four articles and additional analyses, including two theoretical manuscripts and four empirical studies. In the following, I will summarize the main findings of this dissertation, discuss their theoretical and practical implications, as well as limitations of this dissertation, and present ideas for future research.

Theoretical Implications

In the following, I will summarize the main findings of my dissertation and discuss their theoretical implications in relation to the five key questions raised in the general introduction.

Question I: How can OSPs help Psychology in Increasing Trustworthiness?

The first key Question (i.e., *Question I*) was addressed with two articles. Article 1 reviewed the current state of affairs in IOP and related fields. The article presented a theoretical model that summarizes internal and external factors that lead to various researcher behaviors, such as scientific misconduct, errors, and the use of QRPs. These behaviors adversely affect the trustworthiness of our cumulative scientific knowledge, resulting in issues like false positive results, low replicability, and publication bias (Rupp, 2011; Simmons et al., 2011). Furthermore, the model proposes a variety of changes to the harmful external factors and concludes that OSPs offer the potential to moderate the effects of internal and external factors on researcher behaviors. Complementing Article 1, Article 2 reviewed misconceptions

about OSPs, clarifying what OSPs *are not*, and what Open Science does *not* intend to do. It also highlighted the benefits that OSPs offer to the field of IWO Psychology, arguing that the application of OSPs could lead to research practices with more statistical power (e.g., Brandt et al., 2014; Simonsohn, 2015), increased transparency and openness, fewer QRPs, and more replications. These changes have the potential to improve and strengthen current and future research in Psychology.

Together, the two theoretical articles call attention to system-wide factors such as insufficient training (e.g., Tonidandel et al., 2014; Swift et al., 2022), a reward system that incentivizes the wrong behaviors (e.g., a higher likelihood of publication when using QRPs), harmful journal preferences (e.g., Kepes & McDaniel, 2013), and a suboptimal review process (e.g., Hardwicke et al., 2019; Miller, 2006). Following this argumentation, it is especially crucial for the trustworthiness of Psychology that OSPs are supported and implemented by gatekeepers in the field. However, the articles also argue that the implementation of OSPs by individual researchers offers the possibility to reduce the use of QRPs and the engagement in scientific misconduct. Thus, grass-route initiatives and individual researchers implementing OSPs can also gradually change the field. In these ways, the implementation of OSPs by stakeholders and researchers can help Psychology increase its trustworthiness. This is achieved through changes in external factors, such as mentors engaging in OSPs, and journals, including reviewers and editors, encouraging OSPs. Additionally, the diminishing effect that OSPs, such as preregistration and replication studies, might have on the effect of threatening external factors on researchers' behaviors, also plays a crucial role. In summary, and to answer *Question I*, OSPs can help Psychology in increasing trustworthiness by addressing factors that lead to scientific misconduct and QRPs (e.g., norms against data sharing), by increasing transparency and openness, and by encouraging systemic changes within the field (e.g., journals emphasizing Registered Reports, regardless of statistically significant results).

Question II: Are OSPs Broadly Implemented by Relevant Stakeholders?

We addressed *Question II* with two articles, both examining journals as particularly relevant stakeholders. Article 3 revealed an implementation gap concerning OSPs in IOP/Management journals, as noted by Aguinis et al. (2020a). We found that most OSPs were not implemented by the majority of these journals, with a third of the reviewed journals not implementing any OSPs at all. Article 4 complements this research, mirroring the findings from Article 3 by investigating the implementation of replication studies in Social Psychology journals. Only a minority of these journals implemented replication studies, while most did not mention replications at all on their websites. Furthermore, even those journals that generally implemented replications as a publication option often did not highlight this option in their most visible sections, such as the aims and scopes. Therefore, the answer to *Question II* is negative: OSPs do not appear to be broadly implemented by relevant stakeholders.

Question III: Are OSPs Broadly Implemented by Researchers?

We examined the implementation of OSPs by researchers (i.e., *Question III*) by focusing on preregistrations, as it is one of the most promising OSP (Nosek et al., 2018) and can be implemented by researchers independently of external factors, such as organizational restrictions on data sharing or the availability of outlets willing to publish replication studies. To answer this key question, the study reported in the additional analyses assessed the extent to which research in IOP is preregistered. We found that, with (very) few exceptions, research in IOP is currently not preregistered. Despite preregistration being introduced to Psychology nearly a decade ago (Nosek et al., 2015), IOP has yet to widely adopt this OSP. Thus, similar to *Question II*, OSPs like preregistration also do not appear to be broadly implemented by researchers.

Question IV: Did the Implementation of OSPs Change Over Time?

To address *Question IV*, we examined changes in the implementation of OSPs over time through comparisons in Article 4 and the additional analyses. Article 4 found that Social

Psychology journals welcomed replications more frequently in 2022 than in 2015. However, most journals did not update their policies in response to the replication crisis, and the majority still did not implement replication studies as a publication option. Furthermore, an exploratory analysis suggested that the adoption of language supporting replication in journals, occurred at similar rates as other rigor and transparency-promoting policies, such as data sharing and preregistration. While many journals have implemented at least some OSPs, the majority still have minimal or no such policies. The additional analyses examined changes over time in the preregistration of research by IOP researchers. We found that only a few articles in our IOP sample included a preregistration, with almost all of these being published after the replication crisis in Psychology became widely recognized. The proportion of articles with preregistration increased from 2017 to 2023, but still, less than 1% of all published articles in IOP have a preregistration. Thus, while there has been a slight increase in preregistration over recent years, it remains at a very low level. In summary, and to answer *Question IV*, the implementation of OSPs has progressed only slowly over time, both among stakeholders like journals (as shown in Article 4) and among individual researchers (as demonstrated in the additional analyses).

Question V: Are There Differences in the Implementation of OSPs Between Psychological Fields?

The final key question, *Question V*, addressed differences in the implementation of OSPs between psychological fields. Article 3 compared the implementation of OSPs between IOP/Management journals and Social Psychology journals. The study found that IOP/Management journals implemented some OSPs to a lesser extent than Social Psychology journals, specifically in areas such as material or stimuli transparency, preregistrations, Registered Reports, replications, and Open Science Badges. Despite these differences, the overall implementation of OSPs was still relatively low in Social Psychology journals. Our additional analyses did not compare the implementation of OSPs, but rather the adherence to

preregistrations between IOP and a multidisciplinary psychological journal (*Psychological Science*). We found no difference between the two samples in their number of undisclosed or disclosed deviations, nor in the number of studies that deviated from their preregistration in at least one aspect. Furthermore, similar to Heirene et al. (2021) in gambling studies, we found that 65% of our included IOP studies had undisclosed deviations. This suggests that there are comparable rates of undisclosed deviations across different fields of research.

In summary, there are some differences in the implementation of OSPs between different psychological fields. For instance, Social Psychology appears to be further along in the implementation of OSPs compared to IOP/Management. However, the overall implementation of OSPs remains relatively low even in Social Psychology journals, which should have been particularly motivated to initiate change towards a more trustworthy science, given their prominent role in the replication crisis (for an overview, see Pashler & Wagenmakers, 2012). This finding indicates that there is room for improvement across all psychological fields. The rate of undisclosed deviations, however, is comparable across different fields of research. In conclusion, while there are notable differences in the explicit implementation of certain OSPs between IOP/Management journals and Social Psychology journals, the adherence to preregistrations appears to be consistent across different fields, with a significant portion of studies having undisclosed deviations. This indicates that while some fields may be more proactive in implementing OSPs, challenges in broadly adopting and adhering to these practices persist across various areas of Psychology.

Practical Implications

In the following, I will discuss how the findings of this dissertation can be applied in practical settings within Psychology and how they can inform practices within Psychology. I will provide specific recommendations for relevant stakeholders and researchers based on my findings.

Practical Implications for Relevant Stakeholders

Article 1 proposed changes to external factors that could support and accelerate the implementation of OSPs in Psychology. Stakeholders such as funding agencies, journal editors, and academic institutions could change the academic reward system. For instance, academic institutions could revise promotion and tenure processes to emphasize high-quality publications over the quantity of research in prestigious journals (Grand et al., 2018a; Spector, 2022). Journals, as another relevant stakeholder, could emphasize the importance of null-results (Landis et al., 2014), which would also contribute to the realignment of the academic reward system and reduce the motivation to engage in QRPs to find statistically significant results. Additionally, editors and reviewers could encourage the use of OSPs during the review process.

A simple way for journals to support the implementation of OSPs is to communicate this encouragement on their websites. Thus, Articles 3 and 4 suggest that journals should more explicitly discuss OSPs on their websites. By doing so, journals could increase their TOP factor (Nosek et al., 2015; for more recommendations for relevant stakeholders, see Table 2 in Grand et al., 2018a). However, as Article 2 showed, there is currently no competition between journals to achieve the strictest levels of transparency and openness. This low number of journals mentioning and supporting OSPs may explain why researchers often do not adopt these practices (see Aguinis et al., 2020a). Authors typically read the submission guidelines before submitting a manuscript, and if they see that a journal encourages the use of OSPs, they might be more inclined to comply.

I do not suggest that journals engage in a race to achieve the strictest level of transparency, as forcing all researchers to implement every OSP is not reasonable. There are instances where OSPs are not easily implemented, such as when there is no permission to share all data openly. Despite this, my dissertation found that most research outlets do not require most or even *any* transparency and openness. This is unfortunate, given the immense

potential for stakeholders to promote the widespread and long-term implementation of OSPs in Psychology. Furthermore, to encourage more researchers to adopt preregistrations and ultimately create a more open research environment, our additional analyses argued that editors and reviewers should normalize deviations from preregistrations. This flexible approach could also apply to other OSPs. For instance, reviewers could encourage authors to share their datasets, including dependent and independent variables, even if it is not possible to share all raw data due to external constrains. In this way, a flexible rather than strict implementation of OSPs could increase trustworthiness in Psychology while allowing for the inherent flexibility of the research process.

Practical Implications for Researchers

The findings of this dissertation also offer practical implications for researchers. Article 1 offered recommendations for researchers, such as being as transparent as possible (e.g., sharing data and code) and taking advantage of alternative publishing avenues (e.g., results blind review and Registered Reports; Grand et al., 2018b). Similarly to the practical implications for stakeholders, Article 2 emphasized that it is neither necessary nor desirable to force researchers to adopt practices they are uncomfortable with. Fortunately, the article found no evidence of such coercion occurring.

Article 3 builds on the recommendations from Article 1 by suggesting various ways researchers can support the use of OSPs. Beyond implementing OSPs themselves and leading by example, researchers can educate others (e.g., student assistants, PhD students), share their knowledge at conferences, assist colleagues new to OSPs, and raise awareness among co-authors (see also Aguinis et al., 2020a). Our additional analyses highlighted that even “first-generation” implementers of OSPs can make valuable contributions. For instance, researchers from fields less familiar with preregistration performed as well as those from more mature fields. Even vague preregistrations help by reducing flexibility in post-data collection decisions and increasing the transparency of decisions (Simmons et al., 2021). Therefore, our

additional analyses suggest that using OSPs is suitable and relatively easy to learn for researchers across various psychological disciplines.

Limitations and Future Research

The articles included in this dissertation may advance theory and research on OSPs in Psychology. Nevertheless, they have some limitations. Rather than repeating the limitations of the individual studies, I will discuss overarching limitations that suggest avenues for future research in the following.

As a first limitation, there is still insufficient research on the benefits of OSPs in Psychology to conclusively answer *Question I*. Although the theoretical model in Article 1 provides an overview of how OSPs could help to restore Psychology's trustworthiness, and there is robust evidence on the usefulness of some OSPs (e.g., Registered Reports; Hummer et al., 2017; Scheel et al., 2021; Soderberg et al., 2021), more empirical evidence is needed on the actual usefulness of various OSPs. For instance, research on Open Science Badges found that data were more likely to be actually available, correct, usable, and complete when an article earned badges (Kidwell et al, 2016). However, another study with a small sample of $N = 14$ articles studying the effectiveness of badges found that only the results of four articles were essentially or exactly reproducible (Crüwell et al., 2023). Thus, future research is needed that examines OSPs' effectiveness in Psychology.

As a second limitation, it is unclear whether the suggested encouragement of OSPs by journals on their homepages—one of the proposed changes to external factors in Article 1—actually changes individual researchers' behavior. For example, while one study found that providing an Open Science Badge for data sharing resulted in more open data (Kidwell et al., 2016), another study found that the opportunity to receive an Open Data Badge did not increase data sharing (Rowhani-Farid et al., 2020). Furthermore, there is some first correlational evidence that journals supporting replication studies and preregistrations indeed publish more replication studies (see Article 4), and more preregistered studies (see the

additional analyses). However, as this is only correlational research, causal conclusions are not possible. It could therefore be interesting for an experimental study to examine, for example, whether perceiving preregistration as a shared norm (versus no shared norm) in a field, increases researchers' intention to preregister their own studies.

Lastly, the findings from this dissertation suggest that Psychology is currently not broadly implementing OSPs. Understanding the underlying barriers would be beneficial. This dissertation includes a study on perceived barriers among editors (i.e., Study 2 in Article 3). However, we learned that editors often lack authority to make changes to their journals' websites. Thus, it might also be interesting to examine reasons for not implementing OSPs among other important stakeholders, such as the journals' publishers. Additionally, we did not examine barriers faced by individual researchers. Future research could explore why researchers choose not to preregister their studies or not to share their data (for some first surveys, see Sarafoglou et al., 2022; Spitzer & Müller, 2023).

Conclusion

In my dissertation, I assessed the development of OSPs in Psychology in the wake of the replication crisis with two theoretical articles, three empirical studies, and additional analyses, to answer the question whether Psychology is now more of an open science than it used to be before the replication crisis. I found that the implementation of OSPs by stakeholders and individual researchers' promises to increase Psychology's trustworthiness (see Articles 1 and 2). Unfortunately, OSPs do neither seem to be broadly implemented by stakeholders (see Articles 3 and 4), nor by individual researchers (see additional analyses), yet. However, we found some initial movement in the direction of implementing OSPs, both among stakeholders like journals (see Article 4), and among individual researchers (see additional analyses). However, there is still a lot of room for further developments and improvement. Finally, I found some fields to be more proactive in implementing OSPs than others, while challenges in broadly adopting and adhering to these practices persist across

various areas of Psychology (see Article 3 and additional analyses). I hope that the newly gained insights from my dissertation may help to facilitate and accelerate the implementation of OSPS in Psychology and that, in another decade from now, Psychology will be a more open science than it is today.

References

Aguinis, H., Banks, G. C., Rogelberg, S. G., & Cascio, W. F. (2020a). 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

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. 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

Crüwell, S., Apthorp, D., Baker, B. J., Colling, L., Elson, M., Geiger, S. J., Lobentanzer, S., Monéger, J., Patterson, A., Schwarzkopf, D. S., Zaneva, M., & Brown, N. J. L. (2023). What's in a Badge? A computational reproducibility investigation of the Open Data Badge policy in one issue of *Psychological Science*. *Psychological Science*, *34*(4), 512–522. https://doi.org/10.1177/09567976221140828

Grand, J. A., Rogelberg, S. G., Allen, T. D., Landis, R. S., Reynolds, D. H., Scott, J. C., Tonidandel, S., & Truxillo, D. M. (2018a). 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. (2018b). 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

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

Heirene, R., LaPlante, D., Louderback, E. R., Keen, B., Bakker, M., Serafimovska, A., & Gainsbury, S. M. (2021). *Preregistration specificity & adherence: A review of preregistered gambling studies & cross-disciplinary comparison.* PsyArXiv. https://doi.org/10.31234/osf.io/nj4es

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

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

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik,A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, 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

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

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

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., . . .

Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/ 10.1126/science.aab2374

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600-2606. https://doi.org/10.1073/pnas.1708274114

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

Rowhani-Farid, A., Allen, M., & Barnett, A. G. (2020). What incentives increase data sharing in Health and Medical Research? A systematic review. *Research Integrity and Peer Review*, *2*, Article 4, 1-10. https://doi.org/10.1186/s41073-017-0028-9

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

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

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2021). Pre-registration is a game changer. But, like random assignment, it is neither necessary nor sufficient for credible science. *Journal of Consumer Psychology*, *31*(1), 177-180. https://doi.org/10.1002/jcpy.1207

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*(5), 559-569. https://doi.org/gfpn9t

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., Esterling, K. M., & 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

Spitzer, L., & Mueller, S. (2023). Registered Report: Survey on attitudes and experiences regarding preregistration in psychological research. *Plos one*, *18*(3), Article e0281086. https://doi.org/10.1371/journal.pone.0281086

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

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

Author Contributions

Article 1

Reference

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*, Article 104005. https://doi.org/10.1016/j.actpsy.2023.104005

Author Contributions

All authors contributed to the study conception and design. The first draft of the manuscript was written by Sheila K. Keener, Sven Kepes, and Ann-Kathrin Torka and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Article 2

Reference

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

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Joachim Hüffmeier and Ann-Kathrin Torka. The first draft of the manuscript was written by Joachim Hüffmeier and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Article 3

Reference

Torka, A. K., Mazei, J., Bosco, F. A., Cortina, J. M., Götz, M., Kepes, S., O'Boyle, E., & Hüffmeier, J. (2023). How well are Open Science Practices implemented in Industrial and Organizational Psychology and Management? *European Journal of Work and Organizational Psychology*, *32*(4), 461-475. https://doi.org/10.1080/1359432X.2023.2206571

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ann-Kathrin Torka, Jens Mazei and Joachim Hüffmeier. The first draft of the manuscript was written by Ann-Kathrin Torka and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Article 4

Reference

Torka, A. K., Mazei, J., & Hüffmeier, J. (2023). Are replications mainstream now? A comparison of support for replications expressed in the policies of Social Psychology journals in 2015 and 2022. *Social Psychological Bulletin*, *18*, 1-22. https://doi.org/10.32872/spb.9695

Author Contributions

All authors contributed to the study conception and design. Material preparation and data collection were performed by Ann-Kathrin Torka and Jens Mazei. Data analysis was performed by Ann-Kathrin Torka. The first draft of the manuscript was written by Ann-Kathrin Torka and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Additional Analyses

Reference

Torka, A. K., & Hüffmeier, J. (2023). Going off Script: Exploring the Reporting of Preregistration Deviations in Industrial and Organizational Psychology and Their Relationship with QRPs

Author Contributions

Both authors contributed to the study conception and design. Material preparation, data collection and data analysis were performed by Ann-Kathrin Torka. The first draft of the manuscript was written by Ann-Kathrin Torka. Joachim Hüffmeier commented on previous versions of the manuscript. Both authors read and approved the final manuscript.