Using retracted journal articles in psychology to understand research misconduct in the social sciences: What is to be done?

Russell Craig[a], Adam Cox[1b], Dennis Tourish[c], Alistair Thorpe[d]

[1] Corresponding author. E-mail: Adam.Cox@port.ac.uk

[a] Durham University, Durham University Business School, Mill Hill Lane, Durham, DH1 3LB
[b] University of Portsmouth, Portsmouth Business School, Portland Street, Portsmouth, PO1 3DE, UK.
[c] University of Sussex, Falmer, Brighton, BN1 9RH, UK.
[d] University of Essex, Colchester, CO4 3SQ, UK.

Abstract

This paper explores the nature and impact of research misconduct in psychology by analyzing 160 articles that were retracted from prominent scholarly journals between 1998 and 2017. We compare findings with recent studies of retracted papers in economics, and business and management, to profile practices that are likely to be problematic in cognate social science disciplines. In psychology, the principal reason for retraction was data fabrication. Retractions took longer to make, and generally were from higher ranked and more prestigious journals, than in the two cognate disciplines. We recommend that journal editors should be more forthcoming in the reasons they provide for article retractions. We also recommend that the discipline of psychology gives a greater priority to the publication of replication studies; initiates a debate about how to respond to failed replications; adopts a more critical attitude to the importance of attaining statistical significance; discourages *p*-hacking and Hypothesizing After Results are Known (HARKing); assesses the long-term effects of pre-registering research; and supports stronger procedures to attest to the authenticity of data in research papers. Our contribution locates these issues in the context of a growing crisis of confidence in the value of social science research. We also challenge individual researchers to reassert the primacy of disinterested academic inquiry above pressures that can lead to an erosion of scholarly integrity.

Keywords
Misconduct, psychology, research, replication, retractions

JEL Codes
K10, K11, K42, Z0

1. Introduction

There are at least 456 academic journals with the word psychology in their titles. In the UK alone, about 100,000 students enrol in psychology courses each year.[1] Psychologists employ many advanced research techniques, and the discipline has long presented itself as a bastion of rigorous thinking and good scientific practice.[2] The integrity of its research thus matters for its own sake. It may also be a bellwether of problems, and possible solutions, within the wider social science community. Our paper therefore positions its critique of research within psychology in the growing and wide ranging volume of studies into research malpractice across disciplines.

Increasingly, the public image of psychology and the integrity of its peer review processes have been seriously damaged by well-publicized cases of research fraud, failures to replicate important studies, and the growing use of questionable research practices [QRPs] (Pashler & Wagenmakers, 2012; Simmons et al., 2011; Wicherts, 2011). For example, an extensive fraud perpetrated by a social psychologist, Diederik Stapel, was highlighted in *Nature* in 2012 (Callaway, 2011; Yong, 2012a) and in the *New York Times* in 2013.[3] Stapel's research fraud generated strong skepticism regarding the robustness of formerly well-established findings in social and experimental psychology (e.g., those of Bargh et al., 1996; and Doyen et al., 2012). The four following studies testify to some of the fundamental problems involved.

An analysis of approximately 250 research papers in psychology revealed that about 12% reported incorrect *p*-values, that the errors strongly favored the researchers' expectations, and that this led to non-significant findings being reported as significant (Bakker & Wicherts,

---

[1] https://www.studyinternational.com/news/the-increasing-popularity-of-psychology-in-uk-higher-education/

[2] As one indicator of this, the American Psychological Association published a three volume series of handbooks on research methods in psychology in 2012 that comprised 2074 pages (Cooper, 2012).

[3] See https://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html, accessed 4 October 2018.

2011). John et al. (2012) examined responses from 2,155 psychology researchers in major US universities. They reported that 91% of respondents had engaged in at least one QRP; 63% admitted they failed to report all dependent measures; 56% said they collected more data after determining whether results were significant; 22% rounded down p-values; 46% selectively reported studies that "worked"; and 38% excluded data after exploring the impact of doing so. Using a translated version of the John et al. (2012) questionnaire, Agnoli et al. (2017) found that 88% of 277 members of the Italian Association of Psychologists admitted to engaging in at least one QRP; 2.3% admitted to falsifying data; and 53% collected more data after exploring whether results were significant at the 0.05 level. A study of 1138 members of the German Psychological Association by Fiedler and Schwartz (2016) reported that 47% admitted to Hypothesizing After the Results are Known (a practice known as HARKing). We acknowledge that opinion is divided within the research community on the ethical gravity of each of these practices. Nevertheless, the results of these studies suggest a widespread tolerance for, and willingness to participate in, many practices that are increasingly queried by researchers into research integrity.

Robust research practices can be viewed as part of a mind-set in which the pursuit of knowledge and truth takes precedence over a desire to publish mainly (or only) to build an academic career (Tourish, 2019). Tolerance of the practices we discuss above suggests that the imperative to publish has increasingly distorted academic life, and much more widely than in pyschology. Consistent with such a view, Alvesson et al. (2017, p.5) draw attention to

> *… a widespread cynicism among academics on the merits of academic research, sometimes including their own. Publishing comes to be seen as a game of hit and miss, of targets and rankings, crucial for the fashioning of academic careers and institutional prestige but devoid of intrinsic meaning and value, and of no wider social uses whatsoever.*

In that context, this paper has three major aims. First, we seek to enhance understanding of how various forms of misconduct and/ or QRPs have contributed to retractions in

psychology. Second, we propose future pathways that will strengthen the validity and reliability of research in the discipline and the social sciences more widely. Third, we challenge individual academics to become conscious of, and resist, performative pressures that prioritize publication for its own sake above producing robust research that has something genuinely important to say.

Our conclusions stem from an analysis of 160 journal articles that were retracted between 1 January 1998 and 31 December 2017 from psychology journals listed in both the *PsycINFO* and *SCImago* journal databases. We explore reasons for the retractions and juxtapose results with those reported for retracted papers in economics by Cox et al. (2018) and business and management by Tourish and Craig (2018).

We address three research questions (RQ):

RQ 1: What forms of research misconduct in psychology have led to retractions of journal articles?

RQ 2: How does the nature and pattern of retraction in psychology vary from that found in economics and business and management?

RQ 3: What research policies and practices would strengthen the integrity of research in psychology and the wider social sciences?

Although the overall number of retractions across *all* disciplines increased approximately tenfold between 2000 and 2010, the total number of papers published increased at just under half this rate (Steen, 2011; Van Noorden, 2011). Fang et al. (2012, p. 17028) describe this as an "ongoing retraction epidemic." Of the retractions indexed on the *Web of Science,* the rate of retractions per 100,000 papers was 2 in the social sciences (including psychology), 1 in arts and humanities, 14 in biomedicine, 14 in multidisciplinary journals, and 6 in other science fields (Lu et al., 2013). Despite the growing rate of retractions, there has been little attempt to study retractions as a means to better understand research misconduct in psychology. One exception is a study by Stricker and Günther (2019). Their injunction that all inter-disciplinary

comparison of rates of retraction "should be interpreted with caution due to differences in methods and covered time periods" (p.59) is particularly relevant to our study. This viewpoint echoes Cox et al. (2018, p. 928) who argue that:

> *Precise calculation of retraction rates and comparisons of them across, and within, disciplines is an exercise fraught with difficulty. Caution should be exercised in reading purported retraction rates to ensure 'comparing like for like.' Quite often the calculation of individual retraction rates is based on differing methods and underlying assumptions, and is affected by double counting, omissions and other errors.*

Mindful of these issues, care has to be taken in making inferences about levels of research misconduct based on comparison of reported rates of retraction in the three disciplines of interest here.

Kullgren and Carter (2015), co-authors of a retracted paper in psychology,[4] have suggested that most retractions, including their own, are due to "honest scientific mistake." Greitemeyer (2014), drawing on Fang et al. (2012), disagrees. He claims that most retracted papers result from misconduct "whereas error accounts for only about 21.3% of retracted scientific articles" (p. 557). Such conflicting views highlight the need for closer study of the reasons for retraction and for developing a better understanding of the types and severity of research misconduct. We address this need by analyzing research misconduct in psychology through the lens of retracted journal articles.

The literature reviewed in Section 2 highlights contentious behaviors that constitute critically important QRPs, outlines the crisis of research misconduct in psychology, and discusses how misconduct contributes to the lack of reproducibility of research in the field. Section 3 describes the data sources and research methods adopted. Ensuing sections present findings, proffer some policy suggestions, and outline avenues for future research.

---

[4] Their paper (Kullgren, K.A., Tsang, K.K, Ernst, M.M, Carter, B.D., Scott, E.L., & Sullivan, S.K. in *Clinical Practice in Paediatric Psychology,* 2015) was retracted at the authors' request because of errors in "data scoring."

2. Literature Review

2.1 Questionable research practices (QRPs) and the crisis of research misconduct

QRPs typically involve researchers taking advantage of ambivalent attitudes regarding the acceptability of various commonly adopted research practices. For example, one widely used but questionable practice is HARKing. This occurs where a *post-hoc* hypothesis is presented as though it had been generated *a priori*. In this context, "it reinforces the tendency to confirm hypotheses and further reduces the possibility of falsifying them" (Rubin, 2017, p. 308). Rubin points out that such selective reporting of results, and the consequent misrepresentation of methods, studies and results may have contributed to the perception of a replicability crisis within psychology. Given the bias of journals in favour of publishing only positive results, HARKing increases the frequency of type I errors (falsely rejecting the null hypothesis) in the literature. This encourages the adoption of theories that lack robust empirical support (Kerr, 1998). As Bishop (2019, p. 435) argues in relation to HARKing: "*P* values are meaningless when taken out of the context of all the analyses performed to get them."

P-hacking is an umbrella term for a number of practices that, collectively, seek to increase a researcher's chances of obtaining desired *p*-values, such as excluding experimental conditions or dependent measures that do not produce a statistically significant *p*-value (Simonsohn et al., 2014). *P*-hacking also includes either stopping or continuing data collection based on whether a required *p*-value has been reached, rounding *p*-values down to achieve statistical significance, and excluding data after assessing their effect on the prospect of obtaining statistically significant results (Head et al., 2015). Ferguson and Heene (2012, p. 556) describe the publication bias that ensues from *p*-hacking as "an 800-lb [363 kg] gorilla in psychology's living room." Biagioli et al. (2019, p. 403) include *p*-hacking and HARKing in a range of activities that they contend are "aimed at 'puffing up' results to increase their impact." Notably, the American Statistical Association [ASA] has cautioned against what they termed

the widespread "misunderstanding or misuse of statistical inference" (Wasserstein & Lazar, 2016, p. 131). In a specific criticism of *p*-hacking, a statement by the ASA points out that "Conducting multiple analyses of the data and reporting only those with certain *p*-values… renders the reported *p*-values essentially uninterpretable" (Wasserstein & Lazar, 2016, p. 132).

The use of QRPs and the credibility of academic research have become critical issues in science generally (Stroebe et al., 2012). A social psychologist, Stapel, published 58 (now retracted) articles with fraudulent datasets. These appeared in prestigious psychology journals such as *Journal of Personality and Social Psychology* (n = 14) and *European Journal of Social Psychology* (n = 8). Stapel's exposure "came as a shattering blow to the scientific community of psychologists and damaged both their image in the media and their collective self-esteem" (Stroebe et al., 2012, p. 670).

A university committee of investigation into Stapel's activities reported that many of his co-authors had been pressured by journal reviewers and editors to invent retrospective pilot studies, omit analyses that did not lead to expected results, and delete variables that led to inconsistencies with their overall research narrative (Levelt Committee, 2012). This suggests widespread institutional pressure on researchers to report their research dishonestly. We have no reason to imagine that such conduct prevails only within psychology. Ironically, in view of the stress to achieve positive results that arises from these pressures, the consequences for the health of any field are unlikely to be positive. This point was emphasized by Chambers (2017, p. 33) who argued: "If virtually all psychologists engage in *p*-hacking (even unconsciously) at least some of the time, and if *p*-hacking increases the rate of false positives to 50% or higher, then much of the psychological literature will be false to some degree."

Attempts to restore the integrity of research in psychology have focused on reforming elements of the scientific process. The reform agenda has involved re-evaluating methodological and statistical procedures (Morey et al., 2014), requiring authors to submit their

raw data (Chambers, 2017)[5], and agitating for the publication of more replication studies in scholarly journals (Maner, 2014; Martin & Clarke, 2017). There have been many calls too for peer review of research papers to be reconceived to avoid a bias in favor of publishing only positive results. Registered Reports (RRs), a new article format adopted now by several journals,[6] has been proposed as a way forward (Chambers, 2017; Frankenhuis & Nettle, 2018). RRs are intended to emphasize robust and replicable studies, rather than merely novel and/or significant results. They involve pre-registration of research hypotheses and peer-review of the intended research method prior to data collection. This appears to be a low cost strategy intended to lessen the incidence of researchers engaging in QRPs such as *p*-hacking and HARKing.

These issues have come into sharp focus within psychology by what has been dubbed a "crisis of reproducibility". Replications within the discipline are rare (Makel et al., 2012), while many attempts at replication have failed (Ritchie et al., 2010; Francis, 2012; Doyen et al., 2012; McCarthy et al., 2018; Rohrer et al., 2015; Hagger et al., 2016). The field may be rife with Type 1 errors and theories that claim greater empirical support than is justified by the evidence. Our study of retractions seeks to illuminate some of the research weaknesses within the field that contribute to this significant problem.

3. Method

We searched the *PsycINFO* database of approximately 2,300 journals for words with the stem root "retract" on February 8, 2018. At that date, this database provided access to 2,327,172 articles published between 1998 and 2017. The search returned 2,094 items. We removed 1,701 of these items because they drew attention to the retraction notice itself (rather than to a newly

---

[5] This requirement has not garnered widespread support (see Carter et al., 2017; Gabriel & Wessel, 2013).

[6] For a full list of journals using registered reports see: https://cos.io/rr/

retracted paper), to "retraction" as the subject matter of an article, or to retracted comments, letters, editorials, and book reviews. The remaining 393 items represented journal articles that had been retracted in the analysis period.

Journals listed in *PsycINFO* include those with a minor classification as "psychology." Many such journals publish articles on psychology-related topics occasionally and are not usually regarded as being primarily in the field. Examples include *Academy of Management Journal, Journal of Management Studies, Journal of Consumer Research*, *Molecular Pain*. To ensure our focus was on retractions from journals unambiguously in psychology, we required articles identified using *PsycINFO* to also be included in the *SCImago* listing of 1,144 psychology journals. This filtering process reduced our database of retracted articles from 393 to 148. However, *The Leadership Quarterly* (*LQ*) (*n* = 5) was classified as a psychology journal in both databases. Although *LQ* has an interdisciplinary base, its universal prime categorization is as a business and management journal.[7] Accordingly, we regarded *LQ* as a business and management journal, and eliminated its five retractions from our database, leaving 143 for analysis.

In October 2018, the website *Retraction Watch* announced the release of a comprehensive database of 18,000 retractions across all disciplines. We checked the completeness of the database of 143 retractions in psychology we had compiled at that time, against this new listing. We found that *Retraction Watch* listed 201 retractions in psychology during our analysis period. However, it included articles retracted from journals with a third and fourth tier classification in a branch of psychology.[8] We eliminated 82 articles from the *Retraction Watch*

---

[7] For example, the Australian Business Deans' Council classifies *LQ* as in the Australian Research Council field of research 1503, Business and Management (http://www.abdc.edu.au/master-journal-list.php?title=the+leadership+quarterly&issn=&for=0&rating=).

[8] Also included were journals that do not fall within the discipline of psychology: *International Journal of Wrestling Science, International Journal of Nursing Practice, Romanian Journal of English Studies* and highly ranked journals in the discipline of accounting such as *The Accounting Review, Journal of Accounting Research,* and *Accounting, Organizations and Society*.

list because they were not listed also by *SCImago* in psychology, were duplicated items, or were items inaccurately listed as research articles. Thus, our cross-check revealed 119 retractions in the *Retraction Watch* list that satisfied our selection parameters. Of these, 102 were already in our database. The remaining 17 were added to our [then] pre-existing database of 143 retractions, yielding a total of 160 retractions for analysis.[9]

We profiled each of the 160 retracted papers in terms of article title, authors, journal quality, date of publication, date of retraction, principal reason for retraction, and additional information. We classified the reasons for retraction in seven categories: "data fabrication/falsification/fraud", "self-plagiarism", "plagiarism", "flaws in analysis reasoning or reporting", "fake peer review", "other" and "no reason". Journal quality indicators were taken from Thomson Reuters Journal Impact Factor (JIF) scores. *SCImago* Journal Rankings were expressed in four quartile classifications (Q1, Q2, Q3, Q4). We acknowledge that JIF scores have major limitations (Osterloh & Frey, 2019) and should be regarded as "merely indicative of the citation behaviour of researchers" (Eston, 2005, p.1). Nonetheless, we explore them because of their widespread use as a metric to assess the quality of research published in journal articles.

A file describing the 160 retracted papers is published separately in the journal *Data in Brief.* Principal (but not exclusive) responsibility for the reliability of the data in this file, and the process of identifying retracted papers, vests with the second and fourth authors. The reasons for retraction and other pertinent data for each retracted article were obtained from retraction notices on the *PsycINFO* database or the website of the retracting journal; information on the *Retraction Watch* website; or formal reports of various committees of

---

[9] Seven retractions in our database were not listed by *Retraction Watch.* A further 31 retractions in our database are listed by *Retraction Watch* under fields outside of psychology. Our reconciliation revealed that the *Retraction Watch* database duplicates some items and mis-categorises others (e.g., some editorial comments were labelled as research articles).

investigation into research misconduct (such as those conducted into the fraud of Stapel by the Levelt Committee, 2012).

We also juxtaposed our profile of retracted papers in psychology against profiles reported for economics by Cox et al. (2018) and for business and management by Tourish and Craig (2018). These juxtapositions revealed some distinctive characteristics of research (mis)conduct in psychology. To facilitate the drawing of inferences, we restricted comparisons to retraction notices that were dated in the 20-year period from 1 January 1998 to 31 December 2017. This was a period of heightened sensitivity to research misconduct and one in which database technology facilitated systematic search.[10]

4.   Results

The journals with the most retractions were *Journal of Personality and Social Psychology* (*n* = 17), *Personality and Social Psychology Bulletin* (*n* = 11), *Psychological Science* (*n* = 10), and *Journal of Experimental and Social Psychology* (*n* = 9). A high proportion of retractions were due to data fabrication by the serial fraudster, Stapel (mentioned earlier). He was a co-author of 58 retracted papers, 55 of which are analyzed here. The remaining three were included in the retractions from business and management journals that were analyzed by Tourish and Craig (2018). Stapel's papers were published variously with 35 individual co-authors and (with three exceptions) were all from *SCImago* Q1 journals. The high level of citations of his work shows an alarming potential for the fraudulent practices of one individual to have a damaging effect.

Stapel is not the only psychology scholar with multiple retractions. Sanna was the co-author of five papers (with six other co-authors, but principally Parks and Chang) that were

---

[10] To ensure a common period of analysis across disciplines, we slightly adjusted the business and management data reported by Tourish & Craig (2018) by removing a paper by Abbott, retracted by the *Harvard Business Review* in 1898; and the economics data reported by Cox et al. (2018) by removing a paper by Nath and Enns retracted by the *Journal of Applied Probability* in 1982.

retracted at his request on the grounds "the data reported were invalid." In the relevant retraction notices, responsibility was attributed to Sanna, and all co-authors were absolved.[11] In 2011, concerns were raised by Uri Simonsohn about Sanna's published papers because the results displayed "odd patterns" and were consistently "super-significant" with "very large effects." The University of North Carolina [where Sanna was employed at the time] investigated these concerns but did not publish its findings. (For a fuller account see Yong, 2012b).

In this case, the lack of transparency and the researchers' apparent desire to report statistically significant findings is striking. The case also highlights the difficulties journal editors face in eliciting informative reasons for retraction and the problems with attempts to categorize various dimensions of research misconduct. In a retraction notice in *Psychological Science* (23 (8), 2012)[12], the editor stated that he urged Sanna to carefully follow the retraction guidelines developed by the Committee on Publication Ethics (COPE) when drafting his retraction notice, "particularly with respect to stating the reasons for the retraction..." Sanna was provided with a "model" retraction notice and advised to "specify clearly the reasons for the retraction." In reply, Sanna indicated he regretted the "research errors" that prompted him to request retraction, and that legal advice precluded him from "say[ing] anything further ..." The Editor retracted the paper immediately because of uncertainty about whether details of the "research errors would be forthcoming … even though [the published retraction] notice does not reflect COPE guidelines or journal policy."

Similarly, there is little transparency in the case of five papers, all co-authored by Lozano, Hard and Tversky (but with different author sequences) that were retracted because Hard and Tversky believed the research results "cannot be relied upon", and with Lozano taking full

---

[11] See, for example, *Group Dynamics: Theory, Research, and Practice*, 16(3), 2012, 205.

[12] see https://journals-sagepub-com.ezproxy.canterbury.ac.nz/doi/pdf/10.1177/0956797612457151

responsibility.[13] The *Retraction Watch* website reports that Hard and Tversky have declined to comment and that "no address details" can be found for Lozano. Again, the lack of transparency is striking. Why can't the results of the research at issue in the retracted papers be relied upon? Readers are none the wiser. Those who have cited the results will not know whether their work has been compromised, and if so, to what extent.

4.1 Reasons for retraction

The two far right columns of Table 1 report the reasons cited for retraction of journal articles in psychology. The other columns juxtapose reasons reported for retraction in economics and in business and management.

- ***Table 1 here -***

Data fabrication/falsification/fraud is much more prominent as a reason for retraction in psychology (48%) than in business and management (33%) and economics (0%).[14] However, these results are distorted by the volume of sudden retractions attributed to two serial fraudsters: Stapel (55 papers) in psychology (plus three in business and management), and Hunton (37 papers) in business and management. When those of their papers that were retracted for data fabrication/falsification/fraud are removed from the analysis, the level of retraction for this reason is reduced to 21% in psychology and 10% in business and management. This reason remains much higher in psychology than in the other disciplines identified in Table 1. The data provide no obvious reason for this. Two possible explanations suggest themselves. The first is that the level of retractions for data fabrication, falsification and fraud in psychology is due to a more rigorous approach to investigating suspicions of

---

[13] See, for example, *Journal of Experimental Psychology: General*, 137(4), 672.

[14] The non-reporting of data fabrication/falsification/fraud as a reason for any retraction in economics is surprising, particularly given the high level of reliance on extensive data fields in many broad areas of applied economics. A mitigating factor might be the high degree of reliance in some areas of economics (such as modelling) on publicly available data.

misconduct. Other disciplines may simply be less conscientious when it comes to these issues. Retractions only deal with problems that have come to light. They are therefore likely to under-state the prevalence of malpractice that actually occurs. The second explanation is that researchers in psychology are much less ethical than their colleagues in other disciplines. These possibilities warrant further research. Clearly, the high level of retraction for data fabrication/falsification/fraud is a troubling matter that requires further attention.[15]

There was only one single authored paper (1.2%) among the 77 retracted for reasons of data fabrication/falsification/fraud. This was despite 13 (8%) of all retracted papers having only one author. The fact that 82 co-authors (some on multiple occasions) claimed to be implicated innocently in instances of data fabrication is disturbing. The reason for co-authors being duped about the provenance and authenticity of research data is a matter for further research. For example, such research could explore the influence of power relations in research teams, together with the size and diffuseness of such teams. Later, we offer suggestions on how this problem can be addressed.

The lower incidence of "no reason" retraction in psychology (5%) as opposed to business and management (8%) and economics (52%) appears to reflect a stronger commitment by editors of psychology journals to publish informative reasons for retractions. It is a commitment that we commend to other disciplines. Possibly, they see this as a civic duty, given the likelihood that defective research in the discipline could affect human behavior and have harmful consequences for the health and wellbeing of individuals. Editors of psychology journals also seem to have greater respect for the editor guidelines produced by COPE. These

---

[15] Retractions in psychology are likely to grow rapidly. Scholars are campaigning for the retraction of at least 61 publications by the renowned British psychologist Hans Eysenck (Pelosi, 2019; Marks, 2019). *The International Journal of Sports Psychology* has now retracted one of his papers, but after our analysis was completed – see http://www.ijsp-online.com. A recent investigation has also strongly suggested that a seminal study into pseudo patients seeking admission to psychiatric hospitals for investigative purposes (Rosenhan, 1973) contains some fraudulent data, and may warrant retraction (Cahalan, 2019).

guidelines advise that when an error or errors in a paper "renders the work or substantial parts of it invalid, the paper should be retracted *with an explanation as to the reason for retraction…*" (italics applied).[16]

The combined levels for plagiarism and self-plagiarism are much lower in psychology (13%) than in business and management (25%) and economics (22%). There is no immediately obvious reason for this, although it is consistent with findings of a high incidence of text recycling in economics (Horbach & Halffman, 2017). Whilst "fake peer review" is a reason for retraction only in economics, this practice seems likely to exist undetected in business and management and psychology too.[17] The "other reasons" reported for retraction (12%) included publishing without the agreement of all named authors; making substantial changes to an article after acceptance; violating ethical, privacy or intellectual property protocols; and making administrative errors.

We also note that no paper in our database has been retracted because the failure of other researchers to replicate its findings. Yet there are many published papers in the public domain that may warrant scrutiny for this reason. For example, Alter et al. (2007) concluded that subjects found it easier to solve difficult maths problems when a fuzzy font was used to present them. In essence, this result was attributed to perceptual challenges provoking people to think more deeply. According to *Google Scholar*, the paper has been cited 837 times. However, a later study that combined the original data with 13 new replications and the results of three others found no such effect (Meyer et al., 2015). As Loken (2019) has pointed out, the original study "had both the smallest sample size and the most extreme result", a common issue in

---

[16] See https://publicationethics.org/files/International%20standard_editors_for%20website_11_Nov_2011.pdf. Accessed 9 November 2018.

[17] Many psychology journals offer submitting authors the opportunity to nominate reviewers (e.g., the Association for Psychological Science's flagship journal, *Psychological Science*). However, they do not commit to always inviting those nominated to review the papers in question.

failed replications. This suggests, at a minimum, that Alter et al.'s (2007) study was under-powered, and that its findings may therefore be both invalid and unreliable. Nevertheless, it remains in the scientific literature with no suggestion of further examination or retraction. If and when other disciplines also attempt replications more frequently, similar problems seem likely to arise.

4.2 Nature and distinctive features of retractions in psychology

Table 2 compares the time taken to retract and widely used measures of the retracting journal's impact and SCImago journal quartile rankings.

- Table 2 here -

The top band of Table 2 reveals the relative slowness with which articles in psychology are retracted. Sixty-five percent of the retractions in psychology occurred within the first four years of publication, compared with 71% in business and management and 94% in economics. One explanation for this slowness to retract is that researchers in psychology are particularly sensitive to the negative stigma associated with retractions and are discouraged from admitting and correcting mistakes (Kullgren & Carter, 2015).

However, the apparent slowness to retract in psychology is distorted by the high numbers of papers that were retracted suddenly in the exceptional Stapel case, long after the date of initial publication. Stapel's (co-authored) research seemed exemplary for a long time until exposed suddenly by whistle-blowers. A similar situation applied with a spate of 37 retractions between 2012 and 2017 of papers by Hunton in business and management; and to a lesser degree in economics where 12 papers co-authored by Zaman were retracted for "fake peer review" in 2015. When the effect of these mass retractions for individual authors is removed, the results are more favorable for psychology. They show 82% of papers were retracted in the first four years after publication, compared with 66% in business and management and 93% in economics.

Slowness to retract is an important issue because it means that defective work remains in circulation for longer and provides increased opportunity to be read, cited, and to steer the work of other researchers in unproductive directions. The impact is exacerbated because many retracted papers continue to be cited (often with approbation) long after retraction occurs.[18] Such problems are not necessarily avoided by retraction. Experimental research has reported

---

[18] For example, a paper by Hauser et al. titled "Rule learning by cotton top tamarinds," published in *Cognition* in 2002, was retracted in 2010 following findings of an internal report by Harvard University that "the data do not support the reported findings." As of 12 October 2018, this retracted paper had garnered 160 citations, including 36 after the year of retraction. A search within these post retraction citations revealed that only 12 cited the fact of retraction.

that retraction of an article for data fabrication is not sufficient to correct the beliefs of readers regarding the overall soundness of an article (Greitemeyer, 2014). In similar vein, Lu et al. (2013) report that "half or more of the future citations [of retracted papers] continue to accept the original claims [meaning that] … false results can live on, even after formal retraction, magnifying the consequences…" (p. 1).

The middle band of Table 2 discloses details of the incidence, within each of the three discipline areas of interest, of retractions according to the *Normalized* JIF score of the retracting journal. We report normalized JIF scores because *un-*normalized JIF scores are not distributed uniformly across discipline areas. The component factors comprising *un-*normalized JIF scores vary across disciplines because of differences in publication and citation cultures (Dorta-González & Dorta-González, 2013, p. 645).[19] *Un*-normalized JIF scores should only be compared *within* subject fields, preferably by comparing a journal's JIF score against the Aggregate Impact Factor [AIF][20] for all other journals in the subject field as a whole. We use the method presented by Dorta-González and Dorta-González (2013) to allow for the normalization of systematic differences in JIF components between subject fields, thereby rendering the normalized measure more useful for comparing across fields.[21]

---

[19] Sources of variance in JIF scores include differences between subject fields' rate of growth (or contraction) of a field; average number of references cited in published research; ratio of references occurring in the (two year) target window to total references; ratio of items includable in the JIF to total references; and proportion of cited to citing items in the target window. Some fields are characterized by a higher proportion of items that are not included in the JIF (citations of books, conference proceedings, working papers) than of items that are included (such as references to journal articles, notes, reviews, letters, editorials).

[20] The AIF for a subject field is calculated in the same way as the Impact Factor for a journal. However, it takes into account the number of citations to *all* journals in the subject field and the number of articles from *all* journals in the subject field. An AIF of 1.0 means that that, on average, the articles in the subject field published one or two years ago have been cited one time. (https://libguides.bodleian.ox.ac.uk/c.php?g=422992&p=2890486)

[21] Using JIF scores from JCR 2017, each journal JIF score is weighted by the Aggregate Impact Factor of the field with which the journal in question is listed compared to the AIF for all JCR listed fields. An assumption is made in this calculation that each journal was listed in just one field.

Retractions from journals with a Normalized JIF score > 5.0 accounted for 29% of all retractions in psychology, 2% of all retractions in business and management, and 2% of all retractions in economics. The highest Normalized JIF of any psychology journal retracting a paper was 19.236 (*Psychological Inquiry*), whereas in economics it was 5.447 (*American Economic Review*), and in business and management it was 7.056 (*Academy of Management Journal*).

These results strengthen the need for further enquiry into the impact of research misconduct in particular subject fields. Is research misconduct more harmful in psychology because it appears in journals with higher impact and greater prestige? High impact journals have a high level of desirability for authors, given the benefits that accrue from publishing in them. This may tempt some authors to take unethical shortcuts. Editors of such journals may be more motivated to act on suggestions of misconduct than those of lower impact journals. More research on this matter is required.

4.3 Classification case examples

The following two case examples illustrate the fraught nature of the difficulties involved in analyzing and classifying actual retractions. The first example is an article published in the *Canadian Journal of Behavioural Science* in 2000 and retracted in 2012.[22] When one attempts to access the published article, a "correction notice" appears indicating "an erratum for this article was reported" [in the retraction notice of 2012]. Both the correction notice and the retraction notice repeat the published abstract of the retracted paper. The retraction notice says that the retracted paper "reports an error" but does not elaborate on what that error was.

---

[22] The original article appears as Maio, G.R., Esses, V.M., & Bell, D. W. (2000). Examining conflict between components of attitudes: Ambivalence and inconsistency are distinct constructs. *Canadian Journal of Behavioural Science*, 32(1), 58-70. The retraction notice appears at Vol 44 (2), p. 92 of the journal.

A search using *Google*, *Google Scholar* and *Google Citations* failed to reveal any further information about the retraction. In November 2018, *Google Citations* listed 121 citations for the article, including 46 after the year of retraction. We searched within these 46 post-retraction works for further information about the retraction, but found none. The *Retraction Watch* website reported that "unreliable data" and "unreliable results" were the reason for retraction but did not disclose the source for these reasons or provide other information. The website of the publisher of the retracting journal, the Canadian Psychological Association, did not offer any clarifying information.

So, in this case, we have a paper that is retracted for "an error" of unspecified type. However, we have to guess whether this is an error in statistical analysis, reasoning, data transcription, ethical behavior, or an administrative glitch. The absence of publicly available information makes it impossible to classify the source of any research misconduct involved or the severity to those affected by the research misconduct. Moreover, the absence of information about the nature of the error concerned makes it difficult to determine whether other researchers should reject the paper's theoretical framework, methodological approach, statistical analysis, conclusions, or all of the above. Research into the subject matter of the paper (the role of attitudes in conflict) is thus rendered more difficult.

Even when public information is available about a retracted paper, determining the precise reasons for retraction can be difficult. A case in point is a paper retracted by the *Scandinavian Journal of Psychology* in 2013[23] because "certain data within the published work has been used without confirming with Dr. Helge Molde that the data could be utilized … as it had been used in previous studies." The *Retraction Watch* website reports that the lead author, Myrseth, claimed "I had permission to use the data, but at the final stage… I found Molde disagreed. As

---

[23] Myrseth, H., Tvera, R., Hagatun, S., & Lindgren, C. (2012). Impulsivity and sensation seeking in pathological gamblers and skydivers. *Scandinavian Journal of Psychology*, *53*, 340-346.

the disputed data were of no importance for the results and conclusions, I suggested removal of these data ... A slightly modified version was published shortly after retraction."[24]

Jakobsen (2012) reported that Myrseth removed the disputed data to avoid a prolonged conflict, and that Molde declined to comment on why the data could not be used. The non-author complainant, Molde, and the lead author, Myrseth, were both associate professors in the Faculty of Psychology at the University of Bergen. They were two of four people in the Faculty who were reported to be conducting research into "gambling and addiction." They co-authored a paper in the *Journal of Gambling Studies* in 2010. Myrseth's three co-authors of the retracted paper were students. The nature of the professional, social and "power" relationship between Molde and Myrseth in this case could not be determined from publicly available sources. Nor could a reliable determination be made of where the truth lay in terms of the claim and counter claim or whether permission had been given, implicitly or explicitly, to use the data.

The principal reason for the retraction appears to be a violation of Molde's intellectual property rights. It is unclear to what extent the rules on intellectual property are known to researchers, and the rules also vary from country to country. For her part, Myrseth seems likely to contend that she complied with the rules because her permission to use Molde's intellectual property was, at least, tacit – or that confirmation was simply overlooked. If Molde's account is true, the behavior exposed constituted inappropriate conduct because of a violation of known rules.

Determining the source of any research misconduct (such as in this case) is challenging. Drawing on the classificatory terminology advanced by Hall and Martin (2018), it is difficult to determine whether this case is one of what they describe as an "honest mistake". [25]

---

[24] https://retractionwatch.com/2012/06/04/psychologists-take-a-gamble-on-using-data-about-risky-behavior-and-are-forced-to-retract-a-paper/

[25] The use of such a term might simply enable editors to avail themselves of it, rather than probe more deeply into what has gone wrong. For further discussion of such issues, see Resnik and Stewart (2012) and Kullgren and Carter (2015).

Continuing with their terminology in the following sentences, does this case simply instance "ignorance and sloppiness" arising from the inexperience of student co-authors or does it reflect "complexity and ambiguity"? With four authors, perhaps Myrseth assumed that someone else would check whether formal permission had been obtained to use Molde's data? Does this case represent a "bending or gaming the rules" in that the authors were aware of the intellectual property rules, but exploited the lack of clarity and inconsistency in those rules for personal gain? Is this a case of "premeditated dishonesty" featuring the breaking of known rules because the likelihood of being exposed was slight? On balance, we expect that a reasonable reader would probably plump for "bending or gaming the rules."

5. Discussion and Recommendations

Most of the journal articles in psychology that we have analysed were retracted because of data fabrication. However, this result was skewed by the mass exposure of the activities of a serial fraudster, Stapel. In comparison to economics and business and management, retractions in psychology were slower to be made (although this result is again skewed by exposure of the research fraud of Stapel), and were generally from journals with higher prestige and greater impact.

Below, we make seven recommendations regarding research policies and practices that warrant closer consideration or adoption. In making them, we draw on the preceding analysis of retractions within psychology, and on studies of retractions in economics and in business and management. The following recommendations are likely to have a much wider disciplinary applicability:

1. The discipline of psychology should encourage journals to provide clearer and fuller reasons for retraction. Retraction notices rarely give enough information to enable readers to understand the nature of the flaws that have led to a paper's withdrawal from

the scientific record. Greater transparency would provide comparability across journals and disciplines, and help researchers to learn from whatever mistakes are said to have occurred.

2. Stronger support is needed for the publication of replication studies to address the high level of retractions in psychology that is due to fabrication, falsification and fraud (48%); and flaws in statistical analysis reasoning or reporting (22%) (Table 1). A higher level of replication would help to identify problematic research at an earlier stage and reduce the citation and circulation of fundamentally flawed research. Research mind-sets, and institutional and performative pressures that prioritize positive and novel results above replication, are an obstacle to the construction of a reliable knowledge base in the discipline.

3. We encourage debate throughout the social and physical sciences about how to respond to failed replications. Many papers in psychology that have been impossible to replicate have assumed canonical status within the discipline, and are often included in textbooks without any discussion of their methodological limitations (Hughes, 2018). Replications fail for many reasons, sometimes including insufficient expertise on the part of those conducting them. It is therefore clear that investigation and possible retraction should not be a default position in the event of failed replication. However, as noted above, no papers in our database have been retracted because failed replications initiated action. There are now multiple papers in the public domain where many attempts at replication have failed. This contributes to the perception of a crisis of research within psychology. Given the crisis of reproducibility, and the discipline's claims to the status of a science, we encourage the various scholarly academies and professional associations within the field to discuss under what conditions, and after what threshold of failed replication, original studies should be investigated with a view to possible retraction.

4. While the reasons provided for retractions in many disciplines are often vague, it is reasonable to infer that a high proportion of the problems attributed to flawed analysis and reasoning arise from *p*-hacking. These problems could be ameliorated by adopting a more critical attitude to the current pre-occupation with attaining statistical significance; and to the tolerance that researchers have developed for *p*-hacking and HARKing. Psychology is far from unique in suffering from these predilections. We argue that researchers, reviewers and editors should refuse to condone QRPs that have unintended and harmful effects. They should discourage *p*-hacking. This would reduce the likelihood of researchers "playing" with their data until desired *p*-levels are obtained (Ferguson & Heene, 2012).[26] A revisionist stance on these matters should involve stronger commitment to publishing papers that support the null hypothesis or that offer mixed and tentative conclusions — contrary to current publication practices (Grand et al., 2018). This would further challenge the practice of favoring only positive results, and help to lessen the harmful ensuing effect on the conduct of ethical research.

5. One increasingly popular response within psychology to the problems caused by such issues as *p*-hacking is the pre-registration of research projects and the use of Registered Reports (RRs) (Chambers, 2017; Frankenhuis & Nettle, 2018). Chambers (2019, p. 24) notes some early indications that "registered reports are more likely to reveal evidence that is inconsistent with the authors' pre-specified hypotheses (a possible indicator of reduced confirmation bias)." He concludes that RRs also contain more reproducible data. Nevertheless, we recommend that the long-term costs, benefits and unintended consequences of this approach should be carefully researched. There are grounds for caution. A study by Claesen et al. (2019) of 27 pre-registration plans filed by

---

[26] Research endeavours driven by "statistical significance should be a tiny part of an inquiry concerned with the size and importance of relationships" (Ziliak & McCloskey, 2008, p.2). Bayesian hypothesis testing offers a much better approach (Dienes, 2011) because it seems less likely to generate false positives.

psychologists for papers eventually published in *Psychological Science* revealed all researchers deviated from their registered plans, but that the authors of only one paper disclosed doing so. While Claesen and her colleagues acknowledge that deviations may often be inevitable and appropriate, they rightly criticise the failure to disclose them.[27]

6. We have identified data fabrication, falsification and fraud as the primary reason for the retraction of papers in psychology. Clearly, this is also an issue in other disciplines. We have also noted that 82 co-authors of these papers were absolved in retraction notices of responsibility for the unacceptable data issues involved. This demonstrates the need for more rigorous procedures to authenticate the data on which publications depend. A personal guarantee from authors that their work is genuine is no longer sufficient. The onus for validating research data has to shift from those who raise questions to those who advance the original research claims. This problem has been highlighted by the fact that many retracted papers contained no legitimate data at all. Others contained statistical flaws which might have been detected earlier had data been available more readily for re-analysis. Presently, data are shared infrequently (Houtkoop et al., 2018). Although there is considerable skepticism about the value of making data sharing mandatory (Washburn et al., 2018), the many reasons advanced for this reluctance (including confidentiality, restrictions on data usage, and non-data sharing norms) present problems that can, and must, be overcome. Some journals in psychology already do this. The *British Journal of Health Psychology* and the *British Journal of Psychology* require those who submit multi-authored papers to provide a clear statement of which author(s) take responsibility for the data. We suggest that this should be standard

---

[27] For example, one of the papers Claesen et al. (2019) examined had committed to 600 participants but the paper eventually reported a sample of 616. They claimed that this increase "leaves open the possibility that the authors stopped data collection at 600 participants and then used optional stopping to arrive at a favorable outcome with 616 participants" (p.13).

practice. We also support the view that de-identified raw data should be submitted with papers.

7. We urge the social science community to redefine *p*-hacking as a series of *deceptive* research practices rather than ones that are merely *questionable*. Many of the retracted articles in our database, and the wider problem of reproducibility of psychological research, are likely to arise because of *p*-hacking, and the consequent unreliability of many reported findings. This is particularly the case where an unspecified problem with statistical analysis is adduced as the justification for a retraction. Earlier, we noted the American Statistical Association's strong statement urging caution in the use of *p*-values, and their critique of *p*-hacking. Redefining *p*-hacking in the terms that we suggest here would further signal to the research community, including doctoral students, the importance of a more rigorous and transparent approach to statistical analyses. A greater reliance on effect sizes would also be a useful corrective response to the over-statements that have often characterized research in psychology. At a minimum, greater clarity is needed on the conditions under which null hypothesis significance testing (NHST) may be helpful rather than harmful. We endorse the suggestion by Hubbard (2019) that the American Statistical Association should be more forthcoming on this issue. In the interim, the discipline of psychology and its component associations and journals should devote more attention to the role of NHST in research.[28]

---

[28] We also acknowledge the need for some caution and for further enquiry into the importance and role of *p*-values in research. Some have urged the abandonment entirely of NHST (e.g. McShane et al., 2019). The journal, *Basic and Applied Social Psychology*, has taken this approach. However, its ban on NHST may already have had unintended consequences in view of the conclusions of an analysis of papers in that journal in the period following the ban. This found that authors were much more likely to over-state their conclusions beyond what the data would support (Fricker et al., 2019). While this is precisely the problem caused by *p*-hacking, it remains to be seen whether a wholesale ban on NHST provides a solution.

6. Conclusions

The legitimacy of research in psychology is being challenged as never before. Growing recognition of problems has led to exaggerated and premature claims that the field is now undergoing a "renaissance" (Nelson et al., 2018, p. 514). Although progress has been made in enhancing the research integrity of the discipline, there is still widespread reluctance to be fully open about the nature of data and the analytical methods used. If psychology is to contribute to human well-being in the way it is capable of, then greater transparency is vital. The recommendations we make above are intended to facilitate such transparency and help the field achieve its full potential.

We also suggest that individual academics need to look critically at the motivations for publication that have come to exercise strong influence on them. Increasingly, researchers often publish to satisfy the regulatory regimes and perverse audit cultures within which they operate, thereby limiting their disinterested pursuit of knowledge (Craig et al., 2014; Tourish, 2019). As these pressures intensify, and the career penalties for perceived failure to accommodate them become manifest, the temptation to cut corners facilitates many of the practices that we have criticized in this paper. Nevertheless, we do not subscribe to any suggestion that this renders academics powerless or free of any obligation to act. While it would be naïve to suggest that performative pressures can be escaped entirely, it is also naïve to suggest that we have no choice but to adopt dubious practices if we are to play and win the publication "game." Rather, it is vital to reassert the priority of such traditional academic values as the disinterested pursuit of knowledge for its own sake. Doing so requires us to resist talking about publishing as a "game" whose idiosyncratic rules can be bent to make it easier to score publication "hits."

We need a more serious approach to academic integrity, particularly in such disciplines as psychology, where research findings are often seized upon by the media, have implications for public policy and human health, and attract significant forms of public investment. The analysis

and recommendations offered here are intended to improve the practice of research in psychology (and all disciplines) and strengthen public confidence in the value of published research findings.

References

Agnoli, F., Wicherts, J. M., Veldkamp, C. L., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PLoS One*, 12(3), e0172792.

Alter, A., Oppenheimer, D., Epley, N., & Eyre (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General,* 136(4), 569-576.

Alvesson, M., Gabriel, Y., & Paulsen, R. (2017). *Return to meaning: A social science with something to say*. Oxford: Oxford University Press.

Bakker, M., & Wicherts, J.M. (2011). The (mis) reporting of statistical results in psychology journals. *Behavior Research Methods,* 43(3)**,** 666–678.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology*, 71(2), 230-244.

Baumeister, R., Bratlavsky, E., Muraven, M., & Tice, D. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology,* 74(5), 1252-1265.

Biagioli, M., Kenney, M, Martin, B.R., & Walsh, J.P. (2019). Academic misconduct, misrepresentation and gaming: A reassessment. *Research Policy,* 48(2), 401-413.

Bishop, D. (2019). Rein in the four horsemen of irreproducibility, *Nature,* 568(7753), 435.

Cahalan, S. (2019). *The great pretender*. Edinburgh: Canongate Books.

Callaway, E. (2011). Report finds massive fraud at Dutch universities. *Nature,* 479, 15.

Carney, D., Cuddy, A., & Yap, A. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science,* 21(10), 1363-1368.

Carter, E., Schönbrodt, F., Gervais, W.M., & Hilgard, J. (2017). Correcting for bias in psychology: A comparison of meta-analytic methods. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/9h3nu.

Chambers, C. (2019). The registered reports revolution: Lessons in cultural reform. *Significance,* 16(4), 23-27.

Chambers, C. (2017). *The 7 deadly sins of psychology: A manifesto for reforming the culture of scientific practice.* Princeton, NJ: Princeton University Press.

Claesen, A., Gomes, S., Tuerlinckx, F., Vanpaemel, W., & Leuven, K. (2019). Preregistration: Comparing dream to reality. Preprint *PsyArXiv.*

Cooper, H. (Ed.) (2012). *APA handbook of research methods in psychology.* Washington, DC: American Psychological Association.

Cox, A., Craig, R., & Tourish, D. (2018). Retraction statements and research misconduct in economics. *Research Policy,* 47(5), 924-935.

Craig, R., Amernic, J., & Tourish, D. (2014). Perverse audit culture and accountability of the modern public university. *Financial Accountability & Management*, 30(1), 1-24.

Dienes, Z. (2011) Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science,* 6(3), 274-290.

Dorta-González, P., & Dorta-González, M.I. (2013). Comparing journals from different fields of science and social science through a JCR subject categories normalized impact factor. *Scientometrics,* 95(2), 645-672.

Doyen, S., Klein, O., Pichon, C., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS One,* 7, e29081.

Eston, R. (2005). The impact factor: A misleading and flawed measure of research quality. *Journal of Sports Science,* 23(1)*,* 1-3.

Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences*, 109(42), 17028-17033.

Ferguson, C., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science,* 7(6), 555-561.

Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science*, 7(1), 45-52.

Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, 19(2), 151-156.

Frankenhuis, W., & Nettle, D. (2018). Open science is liberating and can foster creativity. *Perspectives on Psychological Science,* 13(4), 439-447.

Fricker, R., Burke, K. Han, X., & Woodall, W. (2019). Assessing the statistical analyses used in *Basic and Applied Psychology* after their *p*-value ban. *The American Statistician, 73:* Supp. 1, 374-384.

Gabriel, A. S., & Wessel, J. L. (2013). A step too far? Why publishing raw datasets may hinder data collection. *Industrial and Organizational Psychology*, 6(3), 287-290.

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

Greitemeyer, T. (2014). Article retracted, but the message lives on. *Psychonomic Bulletin & Review,* 21(2), 557-561.

Hagger, M., Chatzisarantis, Alberts, H., Anggono, C., Batailler, C., Birt, A.R. et al. (2016). A multilab preregistration replication of the ego-depletion effect. *Perspectives on Psychological Science,* 11(4), 546-573.

Hall, J., & Martin, B.R. (2019). Towards a taxonomy of research misconduct: The case of business school research. *Research Policy,* 48(2), 414-427.

Head, M.L., Holman, L., Lanfear, R., Kahn, A.T., & Jennions, M.D. (2015). The extent and consequences of *p*-hacking in science. *PLoS biology*, 13(3), e1002106.

Horbach, S.P.J.M., & Halffman, W. (2017). The extent and causes of academic text recycling or "self-plagiarism." *Research Policy,* 48(2)*,* 492-502.

Houtkoop, B., Chambers, C., Macleod, M., Bishop, D., Nichols, T., & Wagenmakers, E. (2018). Data sharing in psychology: A survey on barriers and preconditions, *Advances in Methods and Practices in Psychological Science,* 1(1), 7085.

Hubbard, R. (2019). Will the ASA's efforts to improve statistical practice be successful? Some evidence to the contrary. *The American Statistician,* 73, Supp. 1*,* 31-35.

Hughes, M. (2018). *Psychology in crisis*. London: Palgrave MacMillan.

Jakobsen, H.O. (2012). Gambler study retracted. *Forskning.no* [Research.no], 7 June.

Jellison, S., Roberts, W., Bowers, A., Combs, T., Beaman, J., Wayant, C., & Vassar (2019, August 5). Evaluation of spin in abstracts of papers in psychiatry and psychology journals, *MJ Evidence-based Medicine*, DOI: 10.1136/bmjebm-2019-111176.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truthtelling. *Psychological Science*, 23(5), 524–532.

Kahneman, D. (2012). Letter. Retrieved on 31 August 2018 from https://www.nature.com/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman%20Letter.pdf.

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

Kullgren, K.A., & Carter, B.D. (2015). Retraction experience, lessons learned, and recommendations for clinician researchers. *Clinical Practice in Pediatric Psychology,* 3(4), 352-357.

Kullgren, K.A., Tsang, K.K., Ernst, M.M, Carter, B.D., Scott, E.L., & Sullivan, S.K. (2015). Inpatient pediatric psychology consultation-liaison practice survey. *Clinical Practice in Pediatric Psychology*, 3(1), 37-47.

Levelt Committee, Noort Committee, Drenth Committee (2012). *Flawed science: The fraudulent research practices of social psychologist Diederik Stapel*, http://pubman.mpdl.mpg.de/pubman/item/escidoc:1569964:7/component/escidoc:156996 6/Stapel_Investigation_Final_report.pdf, Accessed 20 November 2015.

Loken, E. (2019). The replication crisis is good for science, *The Conversation,* https://theconversation.com/the-replication-crisis-is-good-for-science-103736, Accessed 25 September 2019.

Lu, S. F., Jin, G. Z., Uzzi, B., & Jones, B. (2013). The retraction penalty: Evidence from the Web of Science. *Scientific Reports,* 3, 3146.

Makel, M., Plucker, J., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science,* 7(6), 537-542.

Maner, J. K. (2014). Let's put our money where our mouth is: If authors are to change their ways, reviewers (and editors) must change with them. *Perspectives on Psychological Science*, 9(3), 343-351.

Marks, D. (2019). The Hans Eysenck affair: Time to correct the scientific record. *Journal of Health Psychology,* 24(4), 421-439.

Martin, G.N., & Clarke, R.M. (2017). Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology.* 11 April. https://doi.org/10.3389/fpsyg.2017.00523

McCarthy, R., Skowronski, J., & Verschuere, B. (plus 67 others) (2018). Registered replication report on Srull and Wyer (1979). *Advances in Methods and Practices in Psychological Science,* 1(3), 321-336.

McShane, B., Gal, D., Gelman, A., Robert, C., & Tackett, J. (2019). Abandon statistical significance. *The American Statistician,* 73, Supp. 1, 235-245.

Meyer, A., Frederick, S., Burnham, T., Pinto, J., Boyer, T., Ball, L., Pennycook, G., Ackerman, R., & Thompson, V. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General,* 144(2), e16-e30.

Molden, D. (2014). Understanding priming effects in social psychology: What is "social priming" and how does it occur? *Social Cognition,* 32 (SI), 1-11.

Morey, R.D., Rouder, J. N., Verhagen, J., & Wagenmakers, E.J. (2014). Why hypothesis tests are essential for psychological science a comment on Cumming (2014). *Psychological Science*, 25(6), 1289-1290.

Nelson, L., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance, *Annual Review of Psychology,* 69, 511-534.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716-1- aac4716-8.

Osterloh, M., & Frey, B. (2019). How to avoid borrowed plumes in academia. *Research Policy,* doi.org/10.1016/j.respol.2019.103831.

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.

Pelosi, A. (2019). Personality and fatal diseases: Revisiting a scientific scandal. *Journal of Health Psychology,* 24 (4), 409-420.

Ranehill, E., Dreber, A., Johannesson, M., Leiberg, S., Sul, S., & Weber, R. (2015). Assessing the robustness of power posing: No effect on hormones and risk tolerance in a large sample of men and women. *Psychological Science,* 26(5), 653-656.

Resnik, D.B., & Stewart, C.N. (2012). Misconduct versus honest error and scientific disagreement. *Accountability in Research Policies and Quality Assurance,* 19(1), 56-63.

Ritchie, S., Wiseman, R., & French, C. (2012). Failing the future: Three unsuccessful attempts to replicate Bem's "retroactive facilitation of recall" effect. *PLOS One*, 7, e33423.

Rohrer, D., Pashler, H., & Harris, C. (2015). Do subtle reminders of money change people's political views? *Journal of Experimental Psychology: General*, 144, e-73-e85.

Rosenhan, D. (1973). On being sane in insane places. *Science*, 179, 4070, 250-258.

Rubin, M. (2017). When does HARKing hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. *Review of General Psychology,* 21(4)*,* 308-320.

Silberzahn, R., Uhlmann, E., Martin, D., Anselmi, F., Aust, F., Awtrey, E. et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science,* 1(3), 337-356.

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.

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.

Stapel, D. (2014) *Faking Science: A True Story of Academic Fraud,* Trans. N. Brown, Available at http://nick.brown.free.fr/stapel/FakingScience-20141214.pdf. Accessed 16 January 2016.

Steen, R. G. (2011). Retractions in the scientific literature: is the incidence of research fraud increasing? *Journal of Medical Ethics*, 37(4), 249-253.

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.

Stricker, J., & Günther, A. (2018). Scientific misconduct in psychology: A systematic review of prevalence estimates and new empirical data. *Zeitschrift fürPsychologie,* 227, 53-63.

Tourish, D. (2019). *Management studies in crisis: Fraud, deception and meaningless research.* Cambridge: Cambridge University Press.

Tourish, D., & Craig, R. (2018). Research fraud and malpractice in business and management studies. *Journal of Management Inquiry,* DOI: 10.1177/1056492618792621.

Van Noorden, R. (2011). The trouble with retractions. *Nature*, 478(7367), 26-28.

Washburn, A., Hanson, B., Motyl, M., Skitka, L., Yantis, C., Wong, K. et al. (2018). Why do some psychology researchers resist adopting proposed reforms to research practices? A description of researchers' rationales. *Advances in Methods and Practices in Psychological Science,* 1(2), 166-173.

Wasserstein, R., & Lazar, N. (2016). The ASA's Statement on *p*-Values: Context, Process, and Purpose, *The American Statistician*, *70*, Supp 1*,* 129-133.

Wicherts, J. (2011). Psychology must learn a lesson from fraud case. *Nature*, 480(7375), 7.

Yong, E. (2012). Nobel laureate challenges psychologists to clean up their act. *Nature*, https://www.nature.com/news/nobel-laureate-challenges-psychologists-to-clean-up-their-act-1.11535. Accessed 23 September 2019.

Yong, E. (2012b). Uncertainty shrouds psychologist's resignation. *Nature.* 12 July. https://www.nature.com/news/uncertainty-shrouds-psychologist-s-resignation-1.10968

Ziliak, S., & McCloskey, D. (2008). *The cult of statistical significance,* Ann Arbor, MI: University of Michigan Press.