Science and Engineering Ethics (2019) 25:1549–1588
https://doi.org/10.1007/s11948-018-0079-4

ORIGINAL PAPER

What Crisis? Management Researchers' Experiences with and Views of Scholarly Misconduct

Christian Hopp[1] · **Gary A. Hoover**[2]

Received: 13 August 2018 / Accepted: 1 December 2018 / Published online: 2 January 2019

Abstract

This research presents the results of a survey regarding scientific misconduct and questionable research practices elicited from a sample of 1215 management researchers. We find that misconduct (research that was either fabricated or falsified) is not encountered often by reviewers nor editors. Yet, there is a strong prevalence of misrepresentations (method inadequacy, omission or withholding of contradictory results, dropping of unsupported hypotheses). When it comes to potential methodological improvements, those that are skeptical about the empirical body of work being published see merit in replication studies. Yet, a sizeable majority of editors and authors eschew open data policies, which points to hidden costs and limited incentives for data sharing in management research.

Keywords Scientific misconduct · Data fabrication · Data misrepresentation · Ethics

JEL Classification K30 · A11

Introduction

Editors Harley et al. (2014, p. 1361) reflect on the "general unease in the scholarly community about academic misconduct". They indicate growing concerns about the extent of unethical practice at their journal, the *Journal of Management Studies*, and in management research in general as conveyed through private conversations with editors of other leading management journals. Coincidentally, the number of

✉ Christian Hopp
hopp@time.rwth-aachen.de

Gary A. Hoover
ghoover@ou.edu

[1] Technology Entrepreneurship, RWTH Aachen University, TIME Research Area, 52012 Aachen, Germany

[2] Department of Economics, University of Oklahoma, 308 Cate Center Drive, Norman, OK 73072, USA

retractions related to fraudulent behavior is on the rise (Azoulay et al. 2017; Steen 2011). One prominent management scholar has accumulated sixteen retractions from major management journals (Retraction Watch 2016a), and another has racked up seven (Retraction Watch 2016b).

Though the American sociologist Merton (1942, p. 276) once attested to "the virtual absence of fraud in the annals of science," there is currently an ongoing and lively debate about the replicability of published results in, among other fields, psychology, economics, and management research. While some errors derive from accidents or incompetent research, a growing body of work asserts that these errors may also result from academic dishonesty by scholars facing an increasingly competitive environment (Honig et al. 2014; Hubbard 2015).

Current discourse in the social science community maintains that transparency, reproducibility, and traceability of empirical research are important. These factors are necessarily linked to access to original data, estimation files, and research documents. The following pages shed light on contemporary ethics in empirical management research. The present work elicits the views of researchers involved in various levels of the publishing process in management research.

Theoretical Background

Questionable Research Practices and Reproducibility

Individual academics receive credit from and influence the thinking of their peers and the wider academic community through publications (Azoulay et al. 2015; Azoulay et al. 2010). Research topics discussed and analyzed within academic peer groups only reach broader acceptance in academia if they reach a critical mass of publications, citations, and general outreach.

However, the focus on publications as the yardstick to measure academic productivity may induce scientists to push the envelope and, in some cases, engage in unethical behavior (Honig et al. 2014; Chambers 2017). Reluctance to publish non-significant results exacerbates this problem, as scientists compete to publish statistically significant results, and in doing so, to be the first to publish them (Brodeur et al. 2016; O'Boyle et al. 2017).

In the system of "publish or perish" in scientific research, with an increased focus on theoretical contributions, it is beneficial to introduce new ideas, yet not necessarily to validate older ideas (Nosek et al. 2012). Authors may change an a priori hypothesis to a new hypothesis that better reflects the data. Kerr (1998) refers to this practice as "Hypothesizing After the Results are Known (HARKing)": Researchers look ex-post for hypotheses that they can confirm with their data. However, creating or changing a hypothesis after the researchers become aware of the results increases the chance of falsely rejecting a null hypothesis.

Honig et al. (2014) speculate that personal incentives induce scholars in the fast-paced academic environment to take shortcuts, succumb to intense competition, and falter because of high rejection rates along with individualized incentives for publishing. In other words, the incentive system pushes individuals to put personal gains

over scientific peers and society, endangering the veracity and legitimacy of scientific research.

Bedeian et al. (2010) contacted every tenured and tenure-track faculty member (1940 responses) of PhD-granting institutions accredited by the Association to Advance Collegiate Schools of Business concerning the perception of the prevalence of misconduct and misrepresentation in management research. The authors report that 80% of their respondents knew colleagues withholding methodological details and/or dropping unsupported hypotheses. Hence, researchers exert excessively wide discretion in their data preparation and analyses. John et al. (2012) argue that these minor misrepresentations may be both more common and more damaging to the profession than outright fraud.

Replication: Purpose and Perception

One of the best ways to reduce incidents of erroneous scholarship is replication because it double-checks published results (Evanschitzky et al. 2007). Replications may help to improve and extend existing work and operate under the assumption that the underlying research is replicable (Brandt et al. 2014; Open Science Collaboration 2012; Schmidt 2009). Two types of replication exist.

Crandall and Sherman (2016, p. 93) define conceptual replication as "the attempt to test the same theoretical process as an existing study, but that uses methods that vary in some way from the previous study." Conceptual replications cannot refute original findings (Koole and Lakens 2012). Providing evidence on underlying theoretical processes, estimating the extent of effect sizes and/or testing the robustness of initially discovered effects represent the main objectives of conceptual replications (Brandt et al. 2014).

The second type of replication is exact replication. Its main goal is "to recreate a study as closely as possible," though some parts of the study design (participants, time of investigation, cultural background, etc.) are likely to differ (Brandt et al. 2014, p. 218). As much as possible, an exact replication uses the same materials, manipulations, methods, dependent variable, etc., as the original study. In sum, an exact replication involves repetition of all relevant aspects of the original study while a conceptual replication tests previously developed hypotheses using a similar but somewhat different design (Schmidt 2009).

Replication serves as a best practice measure to protect against the uncritical assimilation of erroneous empirical results (Evanschitzky et al. 2007; Hubbard et al. 1998). Evidence-based management research demands systematic replication in support of good and accurate science (Rousseau 2006). Only careful and unbiased re-examination of published data through rigorous replication provides scholars with certainty that critical findings are trustworthy, and should thus be heeded by the general community. The Open Science Collaboration (2012) project describes a study's reproducibility as the demarcation between science and non-science.

Recent research casts doubts on the general replicability of research in psychology. For example, Hartshorne and Schachner (2012) could only successfully replicate every second research project. Camerer et al. (2016) replicated a large number

of published experiments in psychology, with success rates ranging from 60 to 80%. The Open Science Collaboration (2015) combined replications efforts from many researchers into a large-scaled scientific research project investigating the reproducibility of psychological research. Effect sizes were 50% lower in replications; replications could only corroborate two-thirds of the statistically significant effects.

Other disciplines start to realize similar problems of non-replicability. Chang and Li (2015) managed only to replicate approximately half—29 of 59 papers—of the studies published in leading economics journals, even with help from the authors of the original studies. They conclude that, by and large, economics research is not reproducible. Recent work by Camerer et al. (2018) replicates significant coefficients for 60% of published experiments in the social sciences. Yet the effect sizes are 50% weaker than in the published experiments. In management research, replications have a hard time finding acceptance in management research (McCullough et al. 2008).[1] The special issue of *Strategic Management Journal* is a noticeable exception.[2] Many of the replication studies published in this special issue found the same effects as the original articles.[3]

Failure to investigate and ignoring a low reproducibility rate are problematic. "Self-critique, and the promise of self-correction, is what makes science such an important part of humanity's effort to understand nature and ourselves" (Open Science Collaboration 2012, p. 659).

In the following, the study therefore elicits the experiences of management scholars with misconduct and misrepresentations in empirical management research. Moreover, it presents results on the extent to which management researchers believe that empirical findings are replicable and reliable. Lastly, respondents indicate possible methodological improvements they would deem necessary to restore credibility in empirical findings in management research.

Data and Methodology

In August 2016, to begin this study, emails went out to a total of 38,426 unique email addresses of individuals who participated in Academy of Management annual meetings between 2005 and 2015. A second wave of reminder emails went out to the same addresses 4 weeks later, in September 2016. From that, 1215 useable replies came back, which corresponds to a direct response rate of 4.47%. 8682 email

[1] An August 2017 search of the Web of Science for articles with the term "replication" in the title found only 125 articles in journals included in the FT 45 journal list.

[2] See http://onlinelibrary.wiley.com/doi/10.1002/smj.2016.37.issue-11/issuetoc.

[3] Not all replication studies in the SMJ special issue draw findings similar to those of the original work. Tsang and Yamanoi (2016) point out inconsistencies in hypothesis development along with a lack of generalizability based on a sample from Barkema and Vermeulen's (1998) study. Park et al. (2016) fail to replicate the major findings of three studies they sought to replicate.

Table 1 Number of respondents in differing roles

	Editors	Reviewers	Authors
Total respondents: 1215			
Academic	208	767	831
Female	56	280	361
Assistant Professor	10	192	200
Associate Professor	57	193	203
Full Professor	121	240	248

bounced back, and an additional 2575 yielded out-of-office replies and institution changes. To preserve anonymity, the survey requested no identifiable information.[4]

Table 1 reports the sample composition. Out of the 1215, the study eliminated non-academics, and created groups to distinguish the views of those in an editorial role, and those involved as either reviewers, authors, or both. In total, 208 editors responded to the questionnaire. Also, 767 individuals reported that they had acted as reviewers and 831 said they had had an article published in a peer-reviewed publication. The group of authors/reviewers poses a contrast in this work to the editors and reflects on the broader perspective of the field.

Individuals affirming a role as editor, associate, or department editor represent 19 out of the 45 journals included in the Financial Times's FT 45 journal ranking in our dataset.[5] Respondents reviewed in 39 of the FT 45 journals and some 60% of the reviewers indicated reviewing activities for these journals.

Editors, reviewers, and authors were asked about their experiences in publishing and their views on scholarly management research in general. In line with previous research, the present approach was indirect, asking whether or not respondents had encountered a case of scientific malpractice. This does not necessarily mean they engaged in the behavior themselves. In fact, work by Fanelli (2009) shows that direct self-report surveys generally return lower percentages of malpractice. It is important to note that, like other work in this area, these results may not necessarily reflect the true extent of scientific wrongdoing in the profession, given that researchers may have different perspectives on what constitutes scientific malpractice.

Malpractice

The material that follows explores incidences of scientific malpractice, as discussed in other scientific fields. This study focuses on misconduct (fabrication, data deletion, deceptive representation, data tampering) and misrepresentation (method inadequacy, omission or withholding of contradictory results, dropping of unsupported

[4] As the questionnaire design and analysis took place outside of the US, no university institutional review board has been involved in the oversight of this research. The US-based co-author was not involved in data collection and had no access to identifiable data.

[5] In 2016, the FT 45 added five journals to become the FT 50.

Table 2 Questionable research practices (misconduct)

	Never	Rarely	Sometimes	Often	Very often	Total
How often have you encountered the case that an author						
Fabricated or falsified data (editors)	125	53	11	6	1	196
	63.78%	27.04%	5.61%	3.06%	0.51%	
Fabricated or falsified data (reviewers)	648	87	18	7	4	764
	84.82%	11.39%	2.36%	0.92%	0.52%	
Deleted data in an unjustified way (editors)	99	58	25	11	3	196
	50.51%	29.59%	12.76%	5.61%	1.53%	
Deleted data in an unjustified way (reviewers)	537	156	47	17	5	762
	70.47%	20.47%	6.17%	2.23%	0.66%	
Deceptively or misleadingly reported study design (editors)	72	62	38	18	6	196
	36.73%	31.63%	19.39%	9.18%	3.06%	
Deceptively or misleadingly reported study design (reviewers)	388	215	120	30	8	761
	50.99%	28.25%	15.77%	3.94%	1.05%	
Changed or omitted data points in a study (editors)	103	48	27	14	3	195
	52.82%	24.62%	13.85%	7.18%	1.54%	
Changed or omitted data points in a study (reviewers)	511	162	58	21	7	759
	67.33%	21.34%	7.64%	2.77%	0.92%	

hypotheses). Results are tabulated in Table 2. The misconduct questions 1–3 are adapted from Gardner et al. (2005) and Kattenbraker (2007), as cited in Fanelli (2009). Question 4 is a variant of a question asked in Kattenbraker (2007), as cited in Fanelli (2009) and Eastwood et al. (1996). We also investigate the prevalence of misrepresentation, including method inadequacy, omission or withholding of contradictory results, as well as dropping of unsupported hypotheses. Questions 1 and 2 in Table 3 derive from Bebeau and Davis (1996) and Kattenbraker (2007), as cited in Fanelli (2009). Questions 3 and 4 are adapted from Martinson et al. (2005) and Bedeian et al. (2010). For both misconduct and misrepresentation, respondents said not only if they had encountered an incident, but also how often. It is possible that multiple respondents refer to the same incident.

Views on Empirical Management Research

To gain insight into the views of people within the discipline regarding the current state of empirical management research, the survey asked respondents to provide information as to the degree to which they agree or disagree with a series of statements that aims to capture malpractice reported in earlier research. Answers fell on a 5-point Likert scale—Strongly Disagree, Disagree, Neither, Agree, Strongly Agree. The survey focused on incentives to publish statistically significant results (similar to the questionnaire in Devaney (2001), and as reported in Brodeur et al (2016)); overreliance on supporting theory (based on the theoretical reasoning in Hambrick 2007, and the discussion in Leung 2011); the extent to which management results reflect true effect sizes (especially in light of dropping unsupported hypotheses: See O'Boyle et al 2017); over-representation of p-values just below 0.1 (O'Boyle et al. 2017; Schooler 2011); general replicability in the field (Devaney 2001; List et al. 2001; Open Science Collaboration 2015); and the prevalence of false positives (Schooler 2011; Simmons et al. 2011).

General Views on Replication Studies

Respondents also expressed their general views on replications. On a 5-point Likert scale, respondents indicated whether they saw a need to repeat research that others had already conducted (Open Science Collaboration 2015) and whether replications contribute in new ways to the field (Pashler and Harris 2012). Further questions asked whether respondents had reservations with respect to replication studies, which might eventually hamper the conduction and publication of replications, and whether respondents believed that replication studies find acceptance in the field. A final group of questions on replication studies asked about potential disincentives that might explain why replication studies may not receive as much attention as the original work (Pashler and Harris 2012; Stroebe and Strack 2014), whether they may lack creativity (Pashler and Harris 2012; Stroebe and Strack 2014), and whether or not one would recommend that replication studies be carried out by doctoral students (Pashler and Harris 2012; Pashler and Wagenmakers 2012).

Table 3 Questionable research practices (misrepresentation)

	Never	Rarely	Sometimes	Often	Very often	Total
How often have you encountered the case that an author						
Used inadequate or inappropriate research designs (editors)	21	30	62	60	24	197
	10.66%	15.23%	31.47%	30.46%	12.18%	
Used inadequate or inappropriate research designs (reviewers)	134	185	252	134	56	761
	17.61%	24.31%	33.11%	17.61%	7.36%	
Failed to present data that contradicts previous research (editors)	51	55	46	31	11	194
	26.29%	28.35%	23.71%	15.98%	5.67%	
Failed to present data that contradicts previous research (reviewers)	330	192	141	75	22	760
	43.42%	25.26%	18.55%	9.87%	2.89%	
Withheld details of methodology or results (editors)	35	51	55	33	22	196
	17.86%	26.02%	28.06%	16.84%	11.22%	
Withheld details of methodology or results (reviewers)	191	247	174	108	40	760
	25.13%	32.50%	22.89%	14.21%	5.26%	
Dropped unsupported hypotheses from an article (editors)	51	46	41	36	21	195
	26.15%	23.59%	21.03%	18.46%	10.77%	
Dropped unsupported hypotheses from an article (reviewers)	276	203	133	82	63	757
	36.46%	26.82%	17.57%	10.83%	8.32%	

Methodological Improvements in Empirical Research

As a result of their analyses of the prevalence of false-positive findings in psychology research, Simmons et al. (2011) suggest several specific requirements for researchers to restore confidence in the publication process. Building upon their work, a series of questions asked respondents whether they saw these suggestions as beneficial for improving the trustworthiness of findings from empirical management studies. The survey elicits on a 5-point Likert scale whether or not respondents acknowledge rules for data collection termination before the collection begins; whether researchers should list all variables collected for a study; whether reviewers should be tolerant of study imperfections; whether data should be made publicly available; whether all conditions of analyses should be reported; and whether researchers should provide robustness analyses of potential outlier omissions (All in Table 6).

Results and Discussion

In this section, the results of the survey appear with editors contrasting with those who have reviewing and/or authorial roles in the field.

Tables 2 and 3 report the findings concerning the occurrence of misconduct and misrepresentation, disaggregated by editors and authors/reviewers. Tables 4, 5 and 6 report the results concerning views on empirical management research in general, opinions of replication studies, and potential methodological improvements. Tables 4, 5 and 6 are split into the views of authors/reviewers and editors.

Misconduct and Misrepresentation

We first asked respondents (in their role as editor or as author/reviewer) whether they had encountered a case where colleagues had either fabricated or falsified data presented in research. In both groups, most respondents—64% of the editors and 85% of the authors/reviewers—reported that they had never encountered such behavior. Nevertheless, editors reported much higher instances. This is not surprising, given that editors deal with a greater number of manuscripts than reviewers, though presumably it would be reviewers who uncover and report such activities.

The survey asked whether respondents encountered a manuscript containing unjustifiable data deletion. Once again, the majority of respondents in both groups reported that they never encountered such behavior. However, this majority amounted to only 51.51% for editors. The remaining editors responded that they rarely, sometimes, often or very often faced this type of scientific malpractice. In all categories except for "rarely," the rate was twice as much for editors than authors/reviewers.

Another question investigated was whether respondents had encountered manuscripts with deceptively or misleadingly reported study designs. There exists

Table 4 Views on the reliability and replicability of empirical management research

	Strongly disagree	Disagree	Neither	Agree	Strongly agree	Total
To which extent do you agree with the following statements?						
There are strong incentives to publish statistically significant results in empirical management research (editors)	1	1	21	60	125	208
	0.48%	0.48%	10.10%	28.85%	60.10%	
There are strong incentives to publish statistically significant results in empirical management research (authors)	5	15	60	187	564	831
	0.60%	1.81%	7.22%	22.50%	67.87%	
The power of established theories is so great that only results that support them find their way into journals (editors)	17	53	56	54	28	208
	8.17%	25.48%	26.92%	25.96%	13.46%	
The power of established theories is so great that only results that support them find their way into journals (authors)	57	142	282	248	102	831
	6.86%	17.09%	33.94%	29.84%	12.27%	
Results published in management research reflect the true distribution of effect sizes (editors)	3	13	84	72	36	208
	1.44%	6.25%	40.38%	34.62%	17.31%	
Results published in management research reflect the true distribution of effect sizes (authors)	2	60	312	329	128	831
	0.24%	7.22%	37.55%	39.59%	15.40%	
There is an overrepresentation of p-values in the tail of the distribution just below $\alpha = 0.1$ (editors)	12	35	90	56	15	208
	5.77%	16.83%	43.27%	26.92%	7.21%	
There is an overrepresentation of p-values in the tail of the distribution just below $\alpha = 0.1$ (authors)	22	123	377	214	95	831
	2.65%	14.80%	45.37%	25.75%	11.43%	
Results published in management research are generally replicable (editors)	1	43	66	80	18	208
	0.48%	20.67%	31.73%	38.46%	8.65%	
Results published in management research are generally replicable (authors)	10	145	319	259	98	831
	1.20%	17.45%	38.39%	31.17%	11.79%	
False positive results in management research are very persistent (editors)	15	43	81	59	10	208
	7.21%	20.67%	38.94%	28.37%	4.81%	

Table 4 (continued)

	Strongly disagree	Disagree	Neither	Agree	Strongly agree	Total
False positive results in management research are very persistent (authors)	27	174	422	163	45	831
	3.25%	20.94%	50.78%	19.61%	5.42%	

Table 5 Views on replication studies

	Strongly disagree	Disagree	Neither	Agree	Strongly agree	Total
To which extent do you agree with the following statements?						
It is important to repeat research others have already conducted and published (editors)	2	15	42	75	74	208
	0.96%	7.21%	20.19%	36.06%	35.58%	
It is important to repeat research others have already conducted and published (authors)	14	49	164	302	302	831
	1.68%	5.90%	19.74%	36.34%	36.34%	
When management research is replicated, early positive studies often receive more attention than later negative ones (editors)	2	26	71	71	38	208
	0.96%	12.50%	34.13%	34.13%	18.27%	
When management research is replicated, early positive studies often receive more attention than later negative ones (authors)	13	60	251	282	225	831
	1.56%	7.22%	30.20%	33.94%	27.08%	
Replication studies lack originality and creativity (editors)	9	43	63	58	35	208
	4.33%	20.67%	30.29%	27.88%	16.83%	
Replication studies lack originality and creativity (authors)	57	169	222	252	131	831
	6.86%	20.34%	26.71%	30.32%	15.76%	
Replication studies contribute in new ways to the field (editors)	3	22	56	82	45	208
	1.44%	10.58%	26.92%	39.42%	21.63%	
Replication studies contribute in new ways to the field (authors)	18	82	236	333	162	831
	2.17%	9.87%	28.40%	40.07%	19.49%	
I would advise Ph.D. students not to launch their careers by conducting careful replications of published research (editors)	56	56	42	42	12	208
	26.92%	26.92%	20.19%	20.19%	5.77%	
I would advise Ph.D. students not to launch their careers by conducting careful replications of published research (authors)	245	237	193	101	55	831
	29.48%	28.52%	23.23%	12.15%	6.62%	

Table 6 Views on ensuring credibility in empirical management research

	Strongly disagree	Disagree	Neither	Agree	Strongly agree	Total
To ensure credibility in empirical management						
Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article (editors)	51	53	58	34	12	208
	24.52%	25.48%	27.88%	16.35%	5.77%	
Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article (authors)	158	219	284	107	63	831
	19.01%	26.35%	34.18%	12.88%	7.58%	
Authors must list all variables collected in a study (editors)	40	52	52	34	30	208
	19.23%	25.00%	25.00%	16.35%	14.42%	
Authors must list all variables collected in a study (authors)	173	223	173	183	79	831
	20.82%	26.84%	20.82%	22.02%	9.51%	
Reviewers should be more tolerant of imperfections in results (editors)	53	74	38	30	13	208
	25.48%	35.58%	18.27%	14.42%	6.25%	
Reviewers should be more tolerant of imperfections in results (authors)	237	299	156	94	45	831
	28.52%	35.98%	18.77%	11.31%	5.42%	
Journals should require authors to make their original materials and data publicly available (editors)	54	66	44	26	18	208
	25.96%	31.73%	21.15%	12.50%	8.65%	
Journals should require authors to make their original materials and data publicly available (authors)	177	229	198	147	80	831
	21.30%	27.56%	23.83%	17.69%	9.63%	
Authors must report all experimental conditions, including failed manipulation (editors)	75	80	40	12	1	208
	36.06%	38.46%	19.23%	5.77%	0.48%	
Authors must report all experimental conditions, including failed manipulation (authors)	295	309	142	62	23	831
	35.50%	37.18%	17.09%	7.46%	2.77%	
If observations are eliminated, authors must also report what the statistical results are if those observations are included (editors)	75	81	38	13	1	208
	36.06%	38.94%	18.27%	6.25%	0.48%	

Table 6 (continued)

	Strongly disagree	Disagree	Neither	Agree	Strongly agree	Total
If observations are eliminated, authors must also report what the statistical results are if those observations are included (authors)	272 32.73%	303 36.46%	175 21.06%	66 7.94%	15 1.81%	831

a clear separation between editors' and authors'/reviewers' results too. While a slim majority of authors/reviewers reported never encountering such actions, only 37% of editors reported the same. The largest separation between the two groups occurred in the "often" and "very often" category, where the ratio of editors reporting that they found this activity taking place tripled the ratio of authors/reviewers.

Finally, respondents weighed in on whether they had encountered cases where authors had changed or omitted data points from a study. While both respondent categories had majorities reporting that they never encountered such actions, the majority was again slimmer for editors. In addition, for all of the other categories (rarely, sometimes, often, very often), editors reported encountering this form of unethical behavior at much higher rates.

Table 3 covers issues revolving around misrepresenting findings in manuscripts. Unlike the results presented previously, the responses here show that these problems pervade the profession to a much higher degree. Editors are more likely to report having rarely, sometimes, often, or very often encountering such actions. Unlike the previous section, not one question in this series had respondents reporting a majority never encountering such actions.

The survey asked whether respondents had noticed cases where authors had used inadequate or inappropriate research designs. Over 40% of editors and 25% of authors/reviewers reported that they often or very often see colleagues engaging in such malpractice. This could of course also involve several manuscripts of a sole author, which may understate the evidence reported in here.

The survey asked whether respondents had encountered cases where manuscripts failed to present data that would have contradicted previous research. Presenting findings that do not conform with previously published works could lead to rejection or increased scrutiny. As a way of avoiding this, authors might be tempted to leave such findings out. Nearly half of the editors and over 30% of the authors/reviewers reported that they sometimes, often, or very often found manuscripts where this had taken place.

The survey asked whether respondents had encountered cases where the details of the methodology or results were withheld. As noted above, if results do not support the *a priori* assumptions of the model, scholars may simply withhold them from the manuscript. Only 18% of editors and 25% of authors/reviewers reported that they had never encountered this activity. On the other end of the spectrum, 11% of editors and 5% of authors/reviewers reported that they found this happening very often.

Finally, the survey asked whether respondents had encountered manuscripts where unsupported hypotheses had been dropped. A little over a quarter of editors reported that they had never encountered this behavior, and a little over a quarter reported encountering this behavior often or very often. The responses from authors/reviewers were similar, with more reporting that they had never encountered such behavior and fewer reporting that they had often or very often found this in manuscripts.

Views on the Reliability and Replicability of Empirical Management Research

The questions tabulated in Table 4 start by asking prospective groups whether they agreed that strong incentives exist to publish statistically significant results.. For both editors and authors/reviewers, nearly 90% agreed or strongly agreed.[6] As mentioned before, this could derive from authors not wanting to report findings that contradict published works. However, it goes against the scientific notion that researchers should allow the data to lead them. This is troubling and could cast doubt on the reliability of management research. Echoing Hambrick's (2007) discussion of management researchers' obsession with theory, Kwok Leung (2011) explains that these incentives might tempt scientists to engage in post-hoc theorizing—the phenomenon Kerr (1998) calls HARKing, Hypothesizing After the Results are Known. HARKing does not refer to a pattern of induction—inferring some relation or principle post-hoc from a pattern observed in the data—but is about changing an *a priori* hypothesis to a new hypothesis that better reflects the data, increasing the risk of falsely rejecting a null-hypothesis. In a similar fashion, O'Boyle et al. (2017) document how authors drop statistically non-significant hypotheses when papers get closer to publication.

Hence, another way of asking the same question is to ask if the respondents agreed that established theories possess so much power that only papers supporting those theories would find their way into journals. About 40% of editors and authors/reviewers agreed or strongly agreed with this statement. For editors, a bit over a quarter, and for authors/reviewers, about a third, neither agreed nor disagreed with this statement.

Strong and widespread engagement in questionable research practices may invalidate a large part of empirical results. Omitting non-significant variables or outliers, as well as dropping unsupported hypotheses, may falsely inflate the effect sizes reported in the study (Schooler 2011). Scholars are keenly aware of the distribution of the effect sizes. When asked whether published works in management research truly reflects these effects, less than 10% of each group disagreed or strongly disagreed. In other words, the overwhelming majority of respondents were neutral, agreed, or strongly agreed that the printed results indeed reflected the true distribution of effect sizes.

The survey asked respondents if there was an overrepresentation of p-values in the tail of the distribution just below $\alpha = 0.1$. The results were mixed. Almost half of both groups neither agreed nor disagreed with this statement. Among editors and authors/reviewers over 30% either agreed or strongly agreed with the statement. Taken as a whole, it appeared that respondents have concerns that there was indeed an overrepresentation in published works of results with high level of statistical significance.

[6] To reflect on the other side of the process, we also asked journal editors whether manuscripts *should* contain statistically significant results. This question bases on the work of Devaney (2001). Of the 191 respondents to this question, 131 responded that yes, manuscripts should indeed contain significant results. This, in part, may relate to the perceptions of authors/reviewers here.

The key to empirical work is that it is replicable and expanded upon. Some skepticism existed about whether published work in management research was generally replicable; however, the skepticism stayed within limits. For both editors and authors/reviewers, over 40% of the respondents agreed or strongly agreed that management research was generally replicable. It was troubling that for both groups, nearly a third responded that they neither agreed nor disagreed with the statement. Of course, this could mean that respondents did not have enough information to make an informed opinion on the matter, but it is more likely that they were not entirely sure regarding replicability.

Finally, respondents reported how they felt about the notion of persistence of false negative results in management research. In other words, did results that had been disproved or unsupported by data still appear in management manuscripts? For authors/reviewers, 50% neither agreed nor disagreed. Over 30% of journal editors and nearly 25% of authors/reviewers either agreed or strongly agreed with this notion.

Replication

As alluded to in the previous section, management scholars' opinions on the amount of replicable work varied greatly. Table 5 explores the idea of replication even further.

First, the survey questioned both groups about their opinion on the importance of repeating already conducted and/or published work. Among editors and authors/reviewers, over 70% agreed or strongly agreed that such activities should take place.[7] Conversely, both groups had less than 10% of respondents reporting that they disagreed or strongly disagreed.

The survey asked respondents if they agreed that after the replication of management research, early positive studies often receive more attention than later negative results. For editors, over 50% either agreed or strongly agreed with the statement, while more than 60% of authors/reviewers either agreed or strongly agreed. However, this does not settle the issue because for both groups a sizeable third neither agreed nor disagreed.

In the natural sciences, replication of previous studies has value in and of itself. The respondents to this current study, all from the field of management, gave varying responses when asked if they agreed that replication studies lacked creativity and originality. Over 40% of editors and authors/reviewers agreed or strongly agreed with the notion that replication studies were not creative nor original. However, for both groups over a quarter of respondents disagreed or strongly disagreed.

Following the previous question, respondents communicated whether they agreed with the notion that replication studies added to the field in new ways. Over 60% of editors and 70% of authors/reviewers agreed or strongly agreed. However, this

[7] One question (following Devaney 2001), which addressed those with editorial responsibilities only, asked whether replication studies were appropriate for publication. An overwhelming majority of editors, 84%, responded that these types of studies were appropriate.

seems to stand in contrast to the previous question, since so many felt that replication studies were not original or creative. The two notions do not have to be mutually exclusive of each other, but do highlight the lack of uniformity in regards to this important area.

Frank and Saxe (2012) point out that students may benefit from engaging in replication studies by witnessing the process of scientific discovery firsthand, as replicators, instead of only as passive readers. Students achieve this by moving beyond nicely written papers, paying careful attention to methods employed, and understanding and scrutinizing researchers' discretionary choices.

Finally, respondents indicated whether they agreed with the statement that they would not advise doctoral students to start their careers by carefully replicating previous research. Over half of the editors and authors/reviewers either disagreed or strongly disagreed with this statement. Thus, they felt it was a worthy endeavor for beginning scholars. In addition, over 20% of both groups neither agreed nor disagreed with the statement. In other words, a majority of management scholars see value in this type of exercise and felt that young scholars engage in worthy activities by doing so.

Ensuring Credibility

Table 6 addresses questions involving what should and could be done to ensure that management research is perceived as credible. In the end, if the output of scholarly efforts contains doubtful results, end users could view the field and those engaged in it with skepticism.

The first question asked respondents if authors must decide the rule for terminating data collection before data collection begins and report this rule in the manuscript. Editors and authors/reviewers disagreed or strongly disagreed with this notion by a wide margin. In fact, for both groups, only around 20% agreed or strongly agreed. Thus, management researchers will most likely not implement this solution.

The second question asked if authors must list all variables collected for a study. Responses were mixed. For editors, over 40% disagreed or strongly disagreed with the statement while approximately 30% agreed or strongly agreed. For authors/reviewers, the responses were about the same. Thus, this represents another area where no consensus of opinion on how to increase the credibility of management scholars' research findings existed.

The third question asked if reviewers should tolerate more imperfections in reported results (accepting non-significant findings, for example). The answer was a resounding *no*. For both editors and authors/reviewers, over 60% disagreed or strongly disagreed with this sentiment. Only 20% of editors and 16% of authors/reviewers agreed or strongly agreed.

Many journals are now requiring that the original data and materials are supplied at the time of manuscript submission. Not all will make the data publicly available, but some will. We asked respondents if journals should require the provision and publication of data and materials. A sizeable majority of editors disagreed or strongly disagreed. Nearly half of the authors/reviewers disagreed or strongly

disagreed. Conversely, less than 20% of editors agreed or strongly agreed with the statement. Given that the implementation of such a policy would come from the editors and editorial boards, the high rate of disagreement indicates the unlikeliness of a wide adoption, at least in the short term.

The fourth question asked whether authors must report all experimental conditions including failed manipulations. For both, editors and authors/reviewers, over 70% disagreed or strongly disagreed with this statement. There clearly existed little support among respondents for such a requirement for management scholars.

The fifth and final question asked if researchers eliminate observations, they should still report the original statistical significance as a robustness check. For editors, over 70% disagreed or strongly disagreed with this notion. For authors/reviewers, over 60% disagreed or strongly disagreed. Once again, this is a suggestion that will probably not catch on with management scholars, given the strong and overwhelming negative responses.

Post-hoc Analyses

To learn more about how often respondents reacted to cases of misconduct and misrepresentation, we engaged in several post-hoc tests. In doing so, Table 7 uses the variables *misconduct* and *misrepresentation* as the dependent variable. Here, a factor analysis reveals two unique factors on which the questions in Tables 2 and 3, respectively, load. Cronbach's alpha is 0.82 for misconduct and 0.80 for misrepresentation. Due to the nature of the variables (average of four individual items), we employ an OLS regression for each.

Concerning the explanatory variables, individuals can act as *editors* (Managing Editor, Editor-In-Chief = 1, 0 otherwise), or *department or associate editors* (Department/Associate Editor = 1, 0 otherwise). Their *reviewing* activities for journals enter here (Reviewed for FT 45 journal = 1, 0 otherwise).[8] Editors can also be authors/reviewers: categorizations are not mutually exclusive. Similarly, the study controls for the productivity of the scholars by including the *number of scientific publications* (1 = 0 publications, 2 = 1–5 publications, 3 = 6–10 publications, 4 = 11–20 publications, and 50 = > 20 (co)-publications during 2006 and 2010). In addition, the study breaks this variable down into the *number of FT45 journal list publications* with the equivalent coding scheme. Regarding personal characteristics, the study includes the *academic position* (1 = Doctoral/Ph.D student, 2 = PostDoc, 3 = Assistant Professor, 4 = Associate Professor, 5 = Full Professor and 0 = Other), *age* (real number), *gender* (female = 1, male = 0), whether they hold a Ph.D (Ph.D. completed = 1, 0 otherwise), and the location of the university that currently employs them (US = 1, Europe, South America, Asia Pacific, Africa = 0).

Overall, some 44% of respondents came from US-based institutions and 45% from European institutions. One-third were women; 208 respondents indicated editor-in-chief, department editor or associate editor responsibilities; about 50% reviewed for

[8] Respondents received the link to the journal list to corroborate that the journals they had published in appeared on the list in the study.

Table 7 Regression experience with misconduct and misrepresentation

	(1) Misconduct	(2) Misconduct	(3) Misconduct	(4) Misrepresentation	(5) Misrepresentation	(6) Misrepresentation
Academic and publishing experience						
Managing editor			0.213**			0.228*
			(0.040)			(0.078)
Department/associate editor			0.196**			0.254***
			(0.011)			(0.007)
Reviewer for FT 45 journal			0.113**			0.222***
			(0.021)			(0.002)
No. of scientific publications		0.065***	0.047**		0.112***	0.093***
		(0.003)	(0.045)		(0.000)	(0.004)
No. of FT-45 publications		0.103***	0.063*		0.113**	0.037
		(0.001)	(0.068)		(0.011)	(0.459)
Academic position		− 0.009	− 0.017		0.031	0.018
		(0.691)	(0.455)		(0.304)	(0.556)
Controls						
Age	0.003	0.000	− 0.000	− 0.003	− 0.009***	− 0.009***
	(0.131)	(0.906)	(0.901)	(0.236)	(0.005)	(0.003)
Holds a Ph.D.	− 0.010	− 0.122	− 0.120	0.136	− 0.075	− 0.086
	(0.917)	(0.220)	(0.226)	(0.261)	(0.548)	(0.486)
Female	− 0.051	− 0.018	− 0.018	− 0.061	− 0.020	− 0.024
	(0.258)	(0.697)	(0.686)	(0.357)	(0.762)	(0.712)
US scholar	− 0.048	− 0.048	− 0.046	− 0.025	0.003	0.008
	(0.298)	(0.290)	(0.302)	(0.700)	(0.960)	(0.900)
F-value	1.21	4.10	5.01	0.77	4.75	5.90

Table 7 (continued)

	(1) Misconduct	(2) Misconduct	(3) Misconduct	(4) Misrepresentation	(5) Misrepresentation	(6) Misrepresentation
$p >$ F-valued	0.304	0.000	0.000	0.548	0.000	0.000
Observations	938	894	894	938	894	894

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

FT 45 journals; on average, respondents reported 6–10 publications between 2006 and 2010; 40% reported zero FT 45 publications; almost 50% reported between 1 and 5 FT 45 publications; the average age was 47, with 23% being assistant professors, 29% associate professors, and 38% full professors.[9]

Table 7 shows the results. They paint a consistent picture that more experienced scholars experienced more instances of academic misconduct and misrepresentation. Yet, life experience measured in age and academic position did not drive these effects. Instead, experience in the academic publishing process, as authors and as reviewers, influenced researchers' perceptions of the prevalence of academic misconduct and misrepresentation. Hereby, individuals serving as editors or department/associate editors report the strongest effects. Interestingly, academics who reviewed for FT 45 journals experienced more instances of misrepresentation than those who served as reviewers at other journals.

Based on these findings, this study further investigated how experiences with misconduct and misrepresentation affected the views scholars held on the state of the discipline. The results appear in Table 8. Here, the two variables of misconduct and misrepresentation appear as explanatory variables, with the same control variables as in Table 7.

Academics reporting higher levels of misrepresentation generally attested to stronger misguided incentives to publish significant results and to stronger incentives to support theories. Among the control variables there results are mixed, with the only exception that US-based scholars were more likely to agree that incentives for reporting significant results existed. Furthermore, reviewers and department editors of FT 45 journals less often concurred with an overrepresentation of results slightly below the 10% cut-off value.

Interestingly, there existed mixed evidence regarding the trustworthiness of results. Those who experienced misconduct and misrepresentation attested to a stronger belief in the true effect sizes being reported, though they similarly attested to a high prevalence of false positives. Authors/reviewers at FT 45 journals indicated stronger beliefs in replicability, while women and US-scholars were more skeptical. However, whether or not someone considered results robust, or effect sizes accurate, strongly depended on the individuals' definition of misconduct. Observing something that one might not consider as problematic (for example, considering the dropping of unsupported hypotheses as inductive theory building) might not induce a revision of expectations regarding the replicability of published results.

This study therefore also conditioned its analysis on the experiences of the corresponding respondents. The study re-estimated Table 8 with a sample comprising only individuals in an editorial role (available upon request from the authors). What is interesting to note here is that the responses by editors painted a slightly different

[9] When it comes to differences across the groups, three out of four editors report at least one FT 45 publication, with one third of editors having more than five FT 45 publications. For those in non-editor roles, more than 40% report zero FT 45 publications. Among those reviewing for FT 45 journals 82% report at least one FT 45 publication; 20% have more than five FT 45 publications, while two out of three of those not reviewing for FT 45 journals report zero FT 45 publications.

Table 8 Regression views on general trustworthiness of empirical research

	(1)	(2)	(3)	(4)	(5)	(6)
	Publishing incentives			Trustworthiness		
	fishing for significance	Large power of theories	Overrepresentation of $p<0.1$	True effect sizes	Generally replicable	Prevalence of false positives
Academic and publishing experience						
Managing editor	− 0.061	− 0.132	− 0.016	− 0.132	− 0.150	− 0.005
	(0.518)	(0.356)	(0.895)	(0.246)	(0.201)	(0.972)
Department/associate editor	− 0.108	− 0.096	− 0.263***	− 0.150	− 0.044	− 0.174*
	(0.129)	(0.394)	(0.005)	(0.100)	(0.629)	(0.057)
Reviewer for FT 45 journal	− 0.050	− 0.154*	− 0.184**	0.097	0.159**	0.041
	(0.355)	(0.067)	(0.013)	(0.146)	(0.030)	(0.553)
No. of scientific publications	− 0.001	0.028	− 0.038	0.031	0.004	0.020
	(0.977)	(0.445)	(0.241)	(0.266)	(0.889)	(0.511)
No. of FT-45 publications	0.077**	− 0.055	0.094*	0.015	− 0.028	0.008
	(0.025)	(0.343)	(0.069)	(0.734)	(0.593)	(0.858)
Academic position	− 0.022	0.005	− 0.006	− 0.050**	− 0.042	− 0.041
	(0.396)	(0.879)	(0.812)	(0.041)	(0.144)	(0.143)
QRP's experienced						
Misconduct	− 0.008	0.081	0.059	0.204***	0.143**	0.198***
	(0.856)	(0.300)	(0.348)	(0.000)	(0.024)	(0.004)
Misrepresentation	0.122***	0.142***	0.150***	0.089**	0.060	0.098**
	(0.000)	(0.005)	(0.000)	(0.024)	(0.159)	(0.014)
Controls						
Age	− 0.002	− 0.002	− 0.006**	− 0.004	0.002	0.000
	(0.407)	(0.550)	(0.038)	(0.149)	(0.416)	(0.943)

Table 8 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)
	Publishing incentives			Trustworthiness		
	fishing for significance	Large power of theories	Overrepresentation of $p<0.1$	True effect sizes	Generally replicable	Prevalence of false positives
Holds a Ph.D.	− 0.030	0.271*	− 0.078	0.138	0.064	0.030
	(0.755)	(0.082)	(0.464)	(0.236)	(0.622)	(0.775)
Female	0.069	0.108	− 0.070	− 0.099*	− 0.198***	− 0.113*
	(0.183)	(0.168)	(0.286)	(0.092)	(0.002)	(0.063)
US scholar	0.240***	− 0.064	0.011	0.083	− 0.184***	0.022
	(0.000)	(0.416)	(0.864)	(0.157)	(0.004)	(0.735)
F-value	5.03	2.83	4.51	6.16	3.81	3.51
$p>$F-valued	0.000	0.001	0.000	0.000	0.000	0.000
Observations	893	893	893	893	893	893

*$p<0.1$; **$p<0.05$; ***$p<0.01$

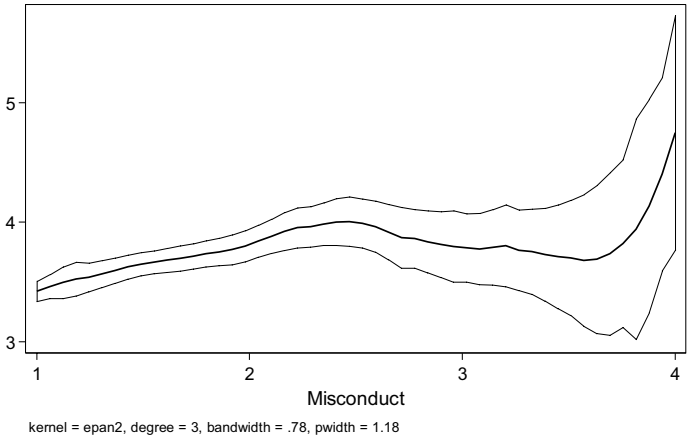

Fig. 1 Local polynomial smoother—effect of misconduct experienced on views about the true effect sizes reported

and more nuanced picture than the full sample analysis. Editors who witnessed more instances of misrepresentation generally attest to stronger misguided incentives to publish significant results and to a more pronounced over-representation of p-values just below the critical 10% cut-off. Those editors who had seen more instances of misconduct were more likely to attest to stronger incentives to support theories. Importantly, the effects of misrepresentation appear robust while the previously reported effects in this paper about misconduct do not hold up.

Subsequent questions therefore also explored how misconduct affected respondents' views using a non-linear smoothing estimator (Seifert and Gasser 2004) reported in Fig. 1. In fact, instances of misconduct were rare, and therefore effects differ slightly from those observed for misrepresentations.

For misconduct encountered, Fig. 1 reports a curvilinear impact on stated beliefs about effect sizes with widening confidence intervals for rarer instances. Those respondents who stated that they sometimes have encountered misconduct (in comparison with those respondents who had seen rare instances or had never observed it) believed the true effect sizes to be lower than the ones reported in published work. However, few respondents agreed that misconduct had happened often or very often. Hence, the large standard errors in this area increase the statistical difficulty of the depiction of the curvature.

Table 9 reveals the views that individuals held on replication studies. Importantly, those who experienced more instances of misreporting saw greater value in replication studies. Hence, scholars perhaps perceived that replication might safeguard against overstating results and other minor research offenses. Yet those who experienced more misconduct did not see much value in replication studies, which indicates that replications may be only part of the answer to misconduct.

Those who indicated that false positives were prevalent, that there was an over-representation of $p < 0.1$, and that there were strong incentives to publish something significant also believed that it is important to replicate the work of

Table 9 Regression—views on value of replication studies

	(1) Importance of repetition	(2) Less credit for replications	(3) Lack of originality	(4) Replications contribute in new ways	(5) Advise Ph.Ds against replications
Academic and publishing experience					
Managing editor	0.058	− 0.230**	− 0.020	0.021	− 0.043
	(0.572)	(0.039)	(0.879)	(0.845)	(0.788)
Department/associate editor	− 0.045	− 0.072	− 0.011	0.020	0.034
	(0.638)	(0.433)	(0.927)	(0.840)	(0.781)
Reviewer for FT 45 journal	− 0.053	− 0.079	0.002	− 0.065	− 0.010
	(0.476)	(0.265)	(0.985)	(0.394)	(0.912)
No. of scientific publications	0.036	− 0.018	0.070*	0.050	0.094**
	(0.243)	(0.569)	(0.077)	(0.138)	(0.019)
No. of FT-45 publications	0.054	0.001	− 0.111*	− 0.075	− 0.146**
	(0.274)	(0.990)	(0.073)	(0.140)	(0.026)
Academic position	− 0.001	0.017	− 0.065*	− 0.023	− 0.007
	(0.959)	(0.556)	(0.068)	(0.445)	(0.853)
QRP's experienced					
Misconduct	− 0.218***	− 0.071	− 0.068	− 0.006	0.042
	(0.000)	(0.245)	(0.398)	(0.931)	(0.600)
Misrepresentation	0.090**	0.069	− 0.045	− 0.019	− 0.097*
	(0.046)	(0.114)	(0.414)	(0.695)	(0.086)
Views about publishing incentives					
Incentives to publish significance	0.215***	0.098**	0.138**	0.115**	− 0.136**
	(0.000)	(0.035)	(0.022)	(0.022)	(0.016)

Table 9 (continued)

	(1) Importance of repetition	(2) Less credit for replications	(3) Lack of originality	(4) Replications contribute in new ways	(5) Advise Ph.Ds against replications
Large power of theories	− 0.007	0.102***	0.022	− 0.048	− 0.040
	(0.834)	(0.001)	(0.579)	(0.133)	(0.294)
Overrepresentation of $p<0.1$	0.066*	0.095**	0.004	0.059	− 0.018
	(0.076)	(0.015)	(0.935)	(0.122)	(0.705)
Views about empirical traceability					
Reflect true effect sizes	0.033	0.160***	0.002	0.055	− 0.054
	(0.418)	(0.000)	(0.965)	(0.220)	(0.282)
Results are generally replicable	0.050	− 0.010	0.096**	− 0.011	− 0.002
	(0.202)	(0.786)	(0.036)	(0.781)	(0.966)
Prevalence of False Positives	0.093**	0.119***	− 0.035	0.134***	− 0.054
	(0.028)	(0.003)	(0.505)	(0.002)	(0.286)
Controls					
Age	0.000	0.002	0.001	0.003	0.009**
	(0.942)	(0.443)	(0.766)	(0.315)	(0.013)
Holds a Ph.D.	− 0.105	0.012	0.085	0.004	0.059
	(0.413)	(0.919)	(0.599)	(0.974)	(0.689)
Female	− 0.184***	0.095	− 0.068	− 0.065	− 0.069
	(0.007)	(0.146)	(0.400)	(0.339)	(0.423)
US scholar	0.073	0.046	0.079	0.008	− 0.376***
	(0.272)	(0.492)	(0.332)	(0.903)	(0.000)
F-value	4.42	6.67	1.54	2.24	4.28
$p>$ F-valued	0.000	0.000	0.069	0.002	0.000

Table 9 (continued)

	(1) Importance of repetition	(2) Less credit for replications	(3) Lack of originality	(4) Replications contribute in new ways	(5) Advise Ph.Ds against replications
Observations	893	893	893	893	893

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

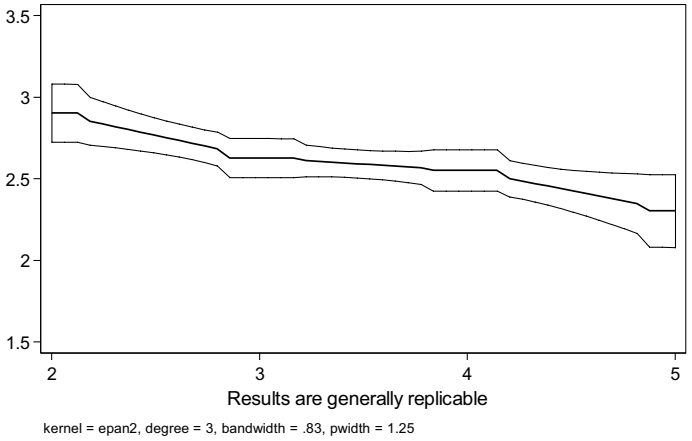

Fig. 2 Local polynomial smoother—effect of replicability views on agreement to open data sharing

others. In addition, those who think that false positives were prevalent and that there were strong incentives to publish something significant believed that replication studies contribute in new ways to the field. Finally, those who believed that there was an overrepresentation of $p < 0.1$, that strong incentives to publish significant results existed, and that management scholars over-relied on theories also believed that replication studies would have a difficult time finding acceptance. Not surprisingly, there still remains mismatch among misrepresentations observed, the value attributed to replications, and the number of replication studies published.

Scholars who published more generally viewed replication studies as less original. However, those who published more in FT 45 journals and those holding higher academic positions saw replication studies as somewhat original. Moreover, more productive scholars publishing in more prestigious journals would not discourage their Ph.D. students from engaging in replication studies, while productive scholars publishing in less prestigious journals would. Likewise, US-based scholars would be more likely to encourage Ph.D. students to engage in replication studies.

Lastly, those who have experienced more instances of misrepresentation did not see open data as beneficial measures to ensure credibility. Those respondents who thought that there existed an over-representation of $p < 0.1$ did not think that more data should be made available. Note again that these respondents were more in favor of replication studies, as reported in Table 9. Those who generally trusted the true effect sizes were generally less inclined to want more reviewer tolerance, or condition and outlier transparency. Similarly, those who thought that results were generally replicable did not see data sharing as important, while those who thought that results were generally not replicable were in favor of open data policies. Again, the study graphically depicted the effect of trust in replicability on the willingness to share data openly using a kernel-weighted local polynomial smoother in Fig. 2. The more trust respondents placed in results, the less they wanted open data policies. Those with little trust were more open to data sharing.

Table 10 Regression views on methodological improvements

	(1) Rules for termination	(2) List all variables collected	(3) More reviewer tolerance	(4) Data availability	(5) Condition transparency	(6) Outlier transparency
Academic and publishing experience						
Managing editor	− 0.117	0.058	0.170	− 0.300**	− 0.064	0.123
	(0.402)	(0.709)	(0.247)	(0.041)	(0.541)	(0.279)
Department/associate editor	− 0.089	0.114	− 0.106	− 0.082	0.074	− 0.154*
	(0.462)	(0.376)	(0.334)	(0.502)	(0.439)	(0.088)
Reviewer for FT 45 journal	− 0.005	0.154	− 0.325***	− 0.009	0.040	− 0.030
	(0.956)	(0.117)	(0.000)	(0.929)	(0.602)	(0.690)
No. of scientific publications	− 0.035	− 0.016	− 0.002	0.015	− 0.023	0.030
	(0.360)	(0.714)	(0.961)	(0.728)	(0.503)	(0.354)
No. of FT-45 publications	0.213***	0.095	0.082	− 0.012	0.059	0.074
	(0.001)	(0.155)	(0.155)	(0.855)	(0.229)	(0.154)
Academic position	0.033	0.020	− 0.022	0.074**	0.004	0.042
	(0.383)	(0.606)	(0.541)	(0.039)	(0.884)	(0.125)
QRP's experienced						
Misconduct	− 0.156*	− 0.232***	0.006	0.071	− 0.044	− 0.034
	(0.055)	(0.005)	(0.928)	(0.388)	(0.421)	(0.537)
Misrepresentation	0.054	0.018	− 0.034	− 0.088	− 0.084*	− 0.030
	(0.316)	(0.772)	(0.516)	(0.117)	(0.053)	(0.468)
Views about publishing incentives						
Incentives to publish significance	− 0.062	0.127**	− 0.238***	0.049	− 0.008	− 0.022
	(0.254)	(0.024)	(0.000)	(0.410)	(0.875)	(0.642)

Table 10 (continued)

	(1) Rules for termination	(2) List all variables collected	(3) More reviewer tolerance	(4) Data availability	(5) Condition transparency	(6) Outlier transparency
Large power of theories	− 0.053	− 0.051	− 0.015	− 0.017	0.023	0.002
	(0.158)	(0.216)	(0.679)	(0.688)	(0.486)	(0.946)
Overrepresentation of p < 0.1	− 0.129***	− 0.149***	− 0.087**	− 0.089*	− 0.117***	− 0.111***
	(0.003)	(0.002)	(0.043)	(0.068)	(0.002)	(0.004)
Views about empirical traceability						
Reflect true effect sizes	0.029	0.052	− 0.083*	− 0.067	− 0.097**	− 0.106**
	(0.558)	(0.336)	(0.072)	(0.193)	(0.022)	(0.011)
Results are generally replicable	0.041	0.005	0.024	− 0.116**	− 0.051	− 0.054
	(0.345)	(0.920)	(0.585)	(0.016)	(0.174)	(0.140)
Prevalence of false positives	− 0.067	− 0.021	0.011	− 0.050	− 0.027	− 0.054
	(0.184)	(0.694)	(0.814)	(0.364)	(0.525)	(0.173)
Controls						
Age	− 0.005	− 0.005	0.023***	− 0.011***	− 0.009***	− 0.010***
	(0.219)	(0.199)	(0.000)	(0.004)	(0.002)	(0.001)
Holds a Ph.D.	0.249	− 0.064	− 0.159	0.021	0.141	0.123
	(0.130)	(0.715)	(0.290)	(0.905)	(0.291)	(0.303)
Female	0.035	0.088	− 0.029	0.403***	0.162**	0.127*
	(0.671)	(0.337)	(0.702)	(0.000)	(0.028)	(0.071)
US scholar	− 0.027	0.224**	− 0.246***	0.203**	0.247***	0.153**
	(0.739)	(0.015)	(0.002)	(0.025)	(0.000)	(0.023)
F-value	3.21	2.97	7.07	4.36	4.25	3.62
p > F-valued	0.000	0.000	0.000	0.000	0.000	0.000

Table 10 (continued)

	(1) Rules for termination	(2) List all variables collected	(3) More reviewer tolerance	(4) Data availability	(5) Condition transparency	(6) Outlier transparency
Observations	893	893	893	893	893	893

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

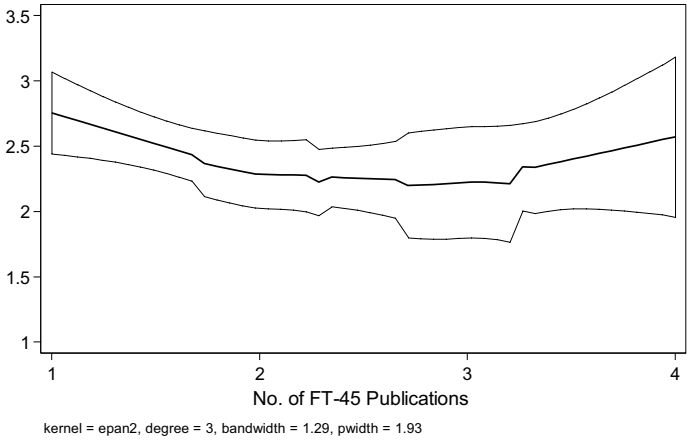

Fig. 3 Local polynomial smoother—impact of FT 45 publications on agreement to open data sharing

Surprisingly, Table 10 reports that editors generally oppose open data policies. Further explorations of these findings show that the study conditioned its analysis again on the academic experience of the editors. When inspecting the results graphically (using a local kernel-weighted local polynomial smoother reported in Fig. 3) in more detail for the editors, a curvilinear effect appears of the number of FT 45 publications on the assessment of open data policies. It seems that editors held generally a bifurcated view of open data policies. Those still in their early careers and those facing publishing pressure were less likely to embrace open data policies. We discuss the implications of these findings further in the conclusion and implications section.

In addition, those who had experienced misconduct did not regard more information on data collection and variables as a potential methodological improvement. More transparency may potentially help against misrepresentation but not against outright fraud. Among the control variables, we find that US-based scholars were substantially more in favor of open data policies and more transparency.

Limitations

The study is not without limitations. The data resembles the Academy of Management's composition fairly well in terms of US- and Europe-based scholars. However, a number of Asia-based scholars sent emails stating that they were not able to fill out the Google-based questionnaire. This restriction might limit the generalizability of the findings as they relate to non-US and non-European scholars. Similarly, the research touches upon misconduct and misrepresentation as they relate to quantitative empirical studies. These types of studies clearly represent only part of the empirical work carried out in management research. This study is not meant to be seen as exhaustive of other similar and potentially equally malicious wrongdoings as

they relate to qualitative research, case study based work, mathematical- and simulation-based modelling, and others.

Also, the low response rate of only 4.47% is somewhat concerning. Despite the large sample derived (1215 replies) the findings may overstate the extent of individual misconduct and misrepresentation in comparison to the main body of published management research work. In fact, qualitative researchers or researchers engaging in theory building will naturally not have witnessed data misconduct or data misrepresentation and might simply have abstained from the survey. Nonetheless, the sample size is similar in magnitude to prior work (Bedeian et al. 2010) and future research may extend this work by including other forms of empirical research and different forms of misconduct more explicitly.

Individuals indicated they have acted as reviewers for nearly all the journals in the FT 45 list. The results are therefore generalizable for the main body of management research. Yet only a very small number reported reviewing for journals focused on accounting, finance, and economics. Hence, the results may have limited value for these sub-disciplines. Similarly, the survey incorporated Marketing and Operations Management journals; they represent the minority. These journals might warrant a closer look into certain types of misbehavior that our questionnaire has not captured. However, the present results are, by and large, reflective of journals like journals like *Academy of Management Journal/Review/Perspectives, Journal of Business Ethics, Journal of Management Studies, Organization Studies, Management Science, Organization Science, Administrative Science Quarterly*, *Entrepreneurship Theory and Practice*, *Journal of Business Venturing and Strategic Management Journal.* For most of these journals, more than 10% of the respondents indicate they have carried out reviewing activities.

The finding that scholars believe published findings to reflect the true distribution of effect sizes and that results are mostly replicable, seems encouraging. Yet it does not follow directly from our results that incidences of misconduct occur only rarely. For example, the belief in the credibility of findings strongly hinges on the views that respondents hold about their discipline (Gigerenzer 2018). Definitions of misconduct strongly affect the level of misconduct reported for this study. As Fanelli (2013) points out, a low consensus about methods, and not necessarily the employing of poor methods or deliberate wrongdoing, might cause biased and not replicable findings. Respondents may not identify practices as misconduct that others may view as such. Therefore, these results may under-report the extent of misconduct if respondents regard part of the practices as acceptable. Similarly, editors and reviewers may not always be in a position where they suspect, let alone detect, misconduct. After all, the review process is not intended to replicate the work submitted by the authors.

This work also did not aim at weighing misbehaviors against each other. Importantly, an over-emphasis on misconduct as the most harmful practice may downplay how adversely misrepresentations may affect the credibility of the discipline (Bülow and Helgesson 2018). Accordingly, Zigmond and Fischer (2002) believe that minor misdemeanors can be equally harmful because they occur more often. While the present respondents are generally open to replication studies, it is important to acknowledge that not all published work should be replicated. As with all research,

the burden will ultimately rest on the replicator to indicate the gains from the replication study a priori.

Finally, the list of types of misrepresentation and misconduct is not exhaustive, nor does the study find evidence at the individual level. Further studies could directly ask scholars about their individual involvement in wrongdoings.

Conclusions and Implications

The survey addressed individuals attending Academy of Management meetings and asked about their experience with scientific malpractice. The results therefore take stock of the general views held by 1214 individual respondents about the current state of their discipline. The respondents report that they rarely encounter misconduct, i.e. fabricated or falsified research. However, they often encounter misrepresentation, which here includes method inadequacy, omission or withholding of contradictory results, and the dropping of unsupported hypotheses.

The present findings indicate that scientific misconduct and misrepresentation are quite common in empirical management research, but not excessive. Editors and authors/reviewers involved in management research report that misrepresentation is not uncommon in empirical management research.

Publishing is paramount for academic careers, yet competition creates incentives to cut corners. When it comes to fighting misrepresentation, method guidelines may help to pave the future of empirical research and may provide help on how to carry out and report research and reported (Antes et al. 2018). One solution would require authors to submit or register their a priori hypotheses before analyzing their data to avoid HARKing (Asendorpf et al. 2013; Chambers 2017). Preregistration is now being widely implemented in randomized medical trials. While some scholars regard post-hoc theorizing as inductive reasoning and not an offense, pre-registration could ensure that researchers follow their protocol and do not delete unsupported hypotheses or develop new hypotheses after the statistical analysis.

Because of the large variety of defensible assumptions applicable for the data analysis, other fields have started to explore different ways to overcome uncertainty about empirical research. In case of disputes about the drawing of conclusions being drawn, adversarial collaboration would be more fruitful (Mellers et al. 2001). This approach requires the involved parties to agree to work together on empirical tests with the help of an arbiter to resolve the dispute. For example, Silberzahn et al. (2014) have teamed up to reconcile ambiguous findings in Silberzahn and Uhlmann (2013) in a scientifically productive fashion.

Management researchers seem to be open to replication-based solutions that might expose mistakes in original findings or may extend previous work by applying newer or better suited methods that become available. Respondents who feel that reliability is endangered point to ways—replication and data sharing—that may help fix the problem. Respondents who are more skeptical about the current state of empirical management research—who say that there exist strong incentives to publish something significant, that false positives are prevalent, and that there is an

over-representation of $p < 0.1$—are more open than others to replication-based solutions that might potentially expose mistakes in original findings.

Especially in light of the pressure to search for significance, significance, and for tailoring theory to the findings, there is a need for scrutiny (Jasny et al. 2011; Karabag and Berggren 2016; Nosek et al. 2012; Open Science Collaboration 2012). Online appendices and repositories allow the documentation of the original working steps. Special sections of journals are devoted to book reviews and teaching cases, but not replication studies. Also, clearly not every paper needs replication, but for the most influential ones, those often cited and carrying implications for a great body of subsequent work, the need for replication is immanent (Ioannidis and Khoury 2014).

The present results also show that scholars agree on the importance of replication studies when dealing with misrepresentation but not necessarily fraud. Notwithstanding the merits of replication, fatal flaws with the original data collection and preparation will go undetected. In fact, the discovery of fraud in the works by Diederik Stapel (Levelt et al. 2012), Lichtenthaler (2010) nor Walumbwa et al. (2011) did not happen through replication.[10] Nevertheless data access and an exact replication would have worked to uncover the fraudulent asterisks that appeared for insignificant coefficients (Lichtenthaler), and it would have exposed instances in which data was not available/existent (Stapel).

Transparency, reproducibility and traceability in scientific research are almost impossible to implement without access to original data, estimation files, and research documents. However, when it comes to future extensions of current processes, both authors/reviewers and editors take a strong stance against making data publicly available in empirical research. In fact, the present research finds that scholars generally prefer to eschew open data. First, respondents who generally think that results are replicable do not view data sharing as important. Secondly, those who report feeling under higher pressure to publish may find open data an excessively heavy burden on the researcher. The research therefore supports an argument by Andreoli-Versbach and Mueller-Langer (2014) that while data sharing for replication is generally beneficial for scientific discourse, come along with hidden costs.

Our findings for the management discipline are in line with concurrent work that reports the open sharing of data as a hotly-debated topic (Borgman 2012, 2015; Gorman et al. 2017). A recent survey by Berghmans et al. (2017) found that while 73% of researchers agree that they would benefit from having access to other researchers' data, a third of the respondents would not share their own data. Respondents cite privacy, data ownership, and control/trust considerations as the primary reasons. Also, editors fear that researchers might gain unearned benefits from the work that the original authors put into data collection efforts and that open data might lead to unfair discredits of authors or the disruption of original publications (Devereaux et al. 2016; Longo and Drazen 2016). Data sharing might carry other hidden costs, too. It may involve curation, documentation, standardization, normalization, and

[10] Though, in some instances, replication has helped to identify fraudulent behavior, as evidenced in Broockman et al. (2015).

metadata association, processes that may put excessively heavy burdens on the researcher (Fecher et al. 2015). Similarly, it is not widely understood how the licensing policies that allow the free reuse of author-submitted data work when data and publication rights are with the publisher. Lastly, the processes of crediting, referencing and citing of original data in re-analyses are unclear.

Important scientific contributions should not rest on the assumption that researchers have done everything in their power to ensure the robustness of the initial findings. As the Open Science Collaboration (2015, p. 943) vividly states, "Innovation points out paths that are possible; replication points out paths that are likely; progress relies on both."

References

Andreoli-Versbach, P., & Mueller-Langer, F. (2014). Open access to data: An ideal professed but not practised. *Research Policy, 43*(9), 1621–1633.

Antes, A. L., English, T., Baldwin, K. A., & DuBois, J. M. (2018). The role of culture and acculturation in researchers' perceptions of rules in science. *Science and Engineering Ethics, 24*(2), 361–391.

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., et al. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*(2), 108–119.

Azoulay, P., Bonatti, A., & Krieger, J. L. (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy, 46*(9), 1552–1569.

Azoulay, P., Furman, J. L., Krieger, J. L., & Murray, F. (2015). Retractions. *Review of Economics and Statistics, 97*(5), 1118–1136.

Azoulay, P., Graff Zivin, J., & Wang, J. (2010). Superstar extincition. *The Quarterly Journal of Economics, 125*(2), 549–589.

Barkema, H. G., & Vermeulen, F. (1998). International expansion through start-up or acquisition: A learning perspective. *Academy of Management Journal, 41*(1), 7–26.

Bebeau, M., & Davis, E. (1996). Survey of ethical issues in dental research. *Journal of Dental Research, 75*(2), 845–855.

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 and Education, 9*(4), 715–725.

Berghmans, S., Cousijn, H., Deakin, G., Meijer, I., Mulligan, A., Plume, A., et al. (2017). *Open data: The researcher perspective-survey and case studies*. New York: Mendeley Data.

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology, 63*(6), 1059–1078.

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., et al. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology, 50*(1), 217–224.

Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics, 8*(1), 1–32.

Broockman, D., Kalla, J., & Aronow, P. (2015). Irregularities in LaCour (2014). In *Working paper*. Stanford University.

Bülow, W., & Helgesson, G. (2018). Criminalization of scientific misconduct. *Medicine, Health Care and Philosophy*. https://doi.org/10.1007/s11019-018-9865-7.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science, 351*(6280), 1433–1436.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T., Huber, J., Johannesson, M., et al. (2018). Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nature Human Behaviour, 2*(9), 637–644.

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton: Princeton University Press.
Chang, A. C., & Li, P. (2015). Is economics research replicable? Sixty published papers from thirteen journals say "Usually Not". In *Finance and Economics Discussion Series 2015-083*. Washington, D.C.: Board of Governors of the Federal Reserve System.
Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology, 66,* 93–99.
Devaney, T. A. (2001). Statistical significance, effect size, and replication: What do the journals say? *The Journal of Experimental Education, 69*(3), 310–320.
Devereaux, P. J., Guyatt, G., Gerstein, H., Connolly, S., & Yusuf, S. (2016). Toward fairness in data sharing. *The New England Journal of Medicine, 375*(5), 405–407.
Eastwood, S., Derish, P., Leash, E., & Ordway, S. (1996). Ethical issues in biomedical research: Perceptions and practices of postdoctoral research fellows responding to a survey. *Science and Engineering Ethics, 2*(1), 89–114.
Evanschitzky, H., Baumgarth, C., Hubbard, R., & Armstrong, J. S. (2007). Replication research's disturbing trend. *Journal of Business Research, 60*(4), 411–415.
Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE, 4*(5), e5738.
Fanelli, D. (2013). Only reporting guidelines can save (soft) science. *European Journal of Personality, 27*(2), 124–125.
Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing? *PLoS ONE, 10*(2), e0118053.
Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science, 7*(6), 600–604.
Gardner, W., Lidz, C. W., & Hartwig, K. C. (2005). Authors' reports about research integrity problems in clinical trials. *Contemporary Clinical Trials, 26*(2), 244–251.
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.
Gorman, D. M., Elkins, A. D., & Lawley, M. (2017). A systems approach to understanding and improving research integrity. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-017-9986-z.
Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal, 50*(6), 1346–1352.
Harley, B., Faems, D., & Corbett, A. (2014). A few bad apples or the tip of an iceberg? Academic misconduct in publishing. *Journal of Management Studies, 51*(8), 1361–1363.
Hartshorne, J. K., & Schachner, A. (2012). Tracking replicability as a method of post-publication open evaluation. *Frontiers in Computational Neuroscience, 6,* 8.
Honig, B., Lampel, J., Siegel, D., & Drnevich, P. (2014). Ethics in the production and dissemination of management research: Institutional failure or individual fallibility? *Journal of Management Studies, 51*(1), 118–142.
Hubbard, R. (2015). *Corrupt research: The case for reconceptualizing empirical management and social science*. Los Angeles: Sage Publications.
Hubbard, R., Vetter, D. E., & Little, E. L. (1998). Replication in strategic management: Scientific testing for validity, generalizability, and usefulness. *Strategic Management Journal, 19*(3), 243–254.
Ioannidis, J. P., & Khoury, M. J. (2014). Assessing value in biomedical research: The PQRST of appraisal and reward. *Journal of the American Medical Association, 312*(5), 483–484.
Jasny, B. R., Chin, G., Chong, L., & Vignieri, S. (2011). Again, and again, and again…. *Science, 334*(6060), 1225.
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.
Karabag, S. F., & Berggren, C. (2016). Misconduct, marginality and editorial practices in management, business and economics journals. *PLoS ONE, 11*(7), e0159492.
Kattenbraker, M. (2007). *Health education research and publication: Ethical considerations and the response of health educators*. Carbondale: Southern Illinois University. Doctoral thesis.
Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217.
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.

Leung, K. (2011). Presenting post hoc hypotheses as a priori: Ethical and theoretical issues. *Management and Organization Review, 7*(3), 471–479.
Levelt, C., Noort, C., & Drenth, C. (2012). *Falende wetenschap: De frauduleuze onderzoekspraktijken van sociaal-psycholoog Diederik Stapel*. Tilburg: Tilburg University.
Lichtenthaler, U. (2010). RETRACTED: Determinants of proactive and reactive technology licensing: A contingency perspective. *Research Policy, 39*(1), 55–66.
List, J. A., Bailey, C. D., Euzent, P. J., & Martin, T. L. (2001). Academic economists behaving badly? A survey on three areas of unethical behavior. *Economic Inquiry, 39*(1), 162–170.
Longo, D., & Drazen, J. (2016). Data sharing. *New England Journal of Medicine, 374*(1), 276–277.
Martinson, B. C., Anderson, M. S., & De Vries, R. (2005). Scientists behaving badly. *Nature, 435*(7043), 737–738.
McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2008). Do economics journal archives promote replicable research? *Canadian Journal of Economics/Revue canadienne d'économique, 41*(4), 1406–1420.
Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12*(4), 269–275.
Merton, R. K. (1942). A note on science and democracy. *Journal of Legal and Political Sociology, 1,* 115.
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.
O'Boyle, E. H., 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.
Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science, 7*(6), 657–660.
Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.
Park, U. D., Borah, A., & Kotha, S. (2016). Signaling revisited: The use of signals in the market for IPOs. *Strategic Management Journal, 37*(11), 2362–2377.
Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*(6), 531–536.
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.
Retraction Watch. (2016a). http://retractionwatch.com/category/by-author/ulrich-lichtenthaler/. Accessed 29 Nov. 2018.
Retraction Watch. (2016b). http://retractionwatch.com/category/by-author/walumbwa/. Accessed 29 Nov. 2018.
Rousseau, D. M. (2006). Is there such a thing as "evidence-based management"? *Academy of Management Review, 31*(2), 256–269.
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.
Schooler, J. (2011). Unpublished results hide the decline effect. *Nature, 470*(7335), 437.
Seifert, B., & Gasser, T. (2004). Local polynomial smoothing. In S. Kotz, C. B. Read, N. Balakrishan, B. Vidakovic, & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences*. Hoboken: Wiley.
Silberzahn, R., Simonsohn, U., & Uhlmann, E. L. (2014). Matched-names analysis reveals no evidence of name-meaning effects: A collaborative commentary on Silberzahn and Uhlmann (2013). *Psychological Science, 25*(7), 1504–1505.
Silberzahn, R., & Uhlmann, E. L. (2013). It pays to be Herr Kaiser: Germans with noble-sounding surnames more often work as managers than as employees. *Psychological Science, 24*(12), 2437–2444.
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.
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., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science, 9*(1), 59–71.

Tsang, E. W., & Yamanoi, J. (2016). International expansion through start-up or acquisition: A replication. *Strategic Management Journal, 37*(11), 2291–2306.

Walumbwa, F. O., Luthans, F., Avey, J. B., & Oke, A. (2011). Retracted: Authentically leading groups: The mediating role of collective psychological capital and trust. *Journal of Organizational Behavior, 32*(1), 4–24.

Zigmond, M. J., & Fischer, B. A. (2002). Beyond fabrication and plagiarism: The little murders of everyday science. *Science and Engineering Ethics, 8*(2), 229–234.