UNIVERSITY OF MARYLAND
ROBERT H. SMITH SCHOOL OF BUSINESS

The state of the OB science: How are we doing, and can we do better?

Gilad Chen
University of Maryland
gchen3@umd.edu

Are we facing a “scientific crisis”?!

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

FOCAL ARTICLE

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

SVEN KEPES AND MICHAEL A. McDANIEL
Virginia Commonwealth University

* Academy of Management Learning & Education, 2010, Vol. 9, No. 4, 715–725.

Management Science on the Credibility Bubble: Cardinal Sins and Various Misdemeanors

ARTHUR G. BEDEIAN
Louisiana State University

SHANNON G. TAYLOR
Northern Illinois University

ALAN N. MILLER
University of Nevada, Las Vegas

Journal of Management
Vol. 43 No. 2, February 2017 376–399
DOI: 10.1177/0149206314527133

PERSONNEL PSYCHOLOGY
2016, 69, 709–750

HARKING'S THREAT TO ORGANIZATIONAL RESEARCH: EVIDENCE FROM PRIMARY AND META-ANALYTIC SOURCES

FRANK A. BOSCO
Virginia Commonwealth University

HERMAN AGUINIS
Indiana University

JAMES G. FIELD
Virginia Commonwealth University

CHARLES A. PIERCE
University of Memphis

DAN R. DALTON
Indiana University

The Chrysalis Effect: How Ugly Initial Results Metamorphosize Into Beautiful Articles

Ernest Hugh O'Boyle Jr.
University of Iowa
George Christopher Banks
Longwood University
Erik Gonzalez-Mulé
University of Iowa

Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?

Harold Pashler[1] and Eric–Jan Wagenmakers[2]
[1]University of California, San Diego and [2]University of Amsterdam, The Netherlands

Perspectives on Psychological Science
7(6) 528–530

DOI: 10.1177/1745691612465253
http://pps.sagepub.com
SAGE

Are we facing a “scientific crisis”?!

SCIENCE

The New York Times

Many Psychology Findings Not as Strong as Claimed, Study Says

The New York Times Magazine

FEATURE

When the Revolution Came for Amy Cuddy

As a young social psychologist, she played by the rules and won big: an influential study, a viral TED talk, a prestigious job at Harvard. Then, suddenly, the rules changed.

Are we facing a “scientific crisis”?!

A Harvard dishonesty researcher was accused of fraud. Her defense is troubling.

The more we learn about Francesca Gino’s lawsuit, the more problems arise.

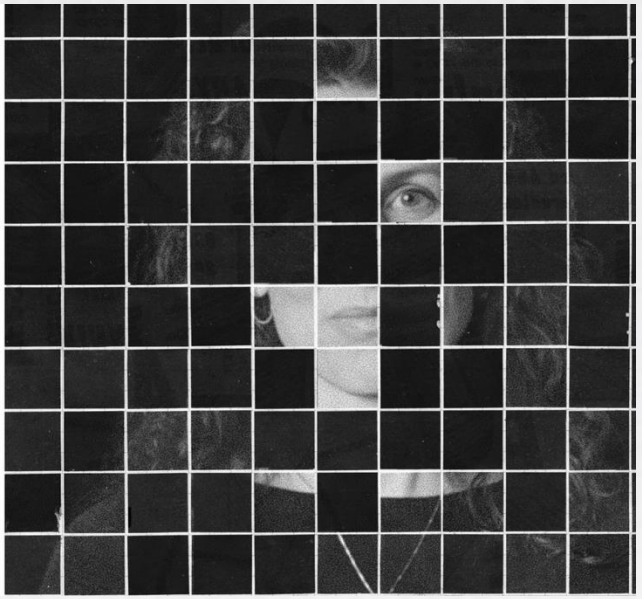
by **Kelsey Piper**
Mar 22, 2024 at 9:00 AM EDT

The New York Times

Account

The Harvard Professor and the Bloggers

When Francesca Gino, a rising academic star, was accused of falsifying data — about how to stop dishonesty — it didn’t just torch her career. It inflamed a crisis in behavioral science.



See: https://datacolada.org/

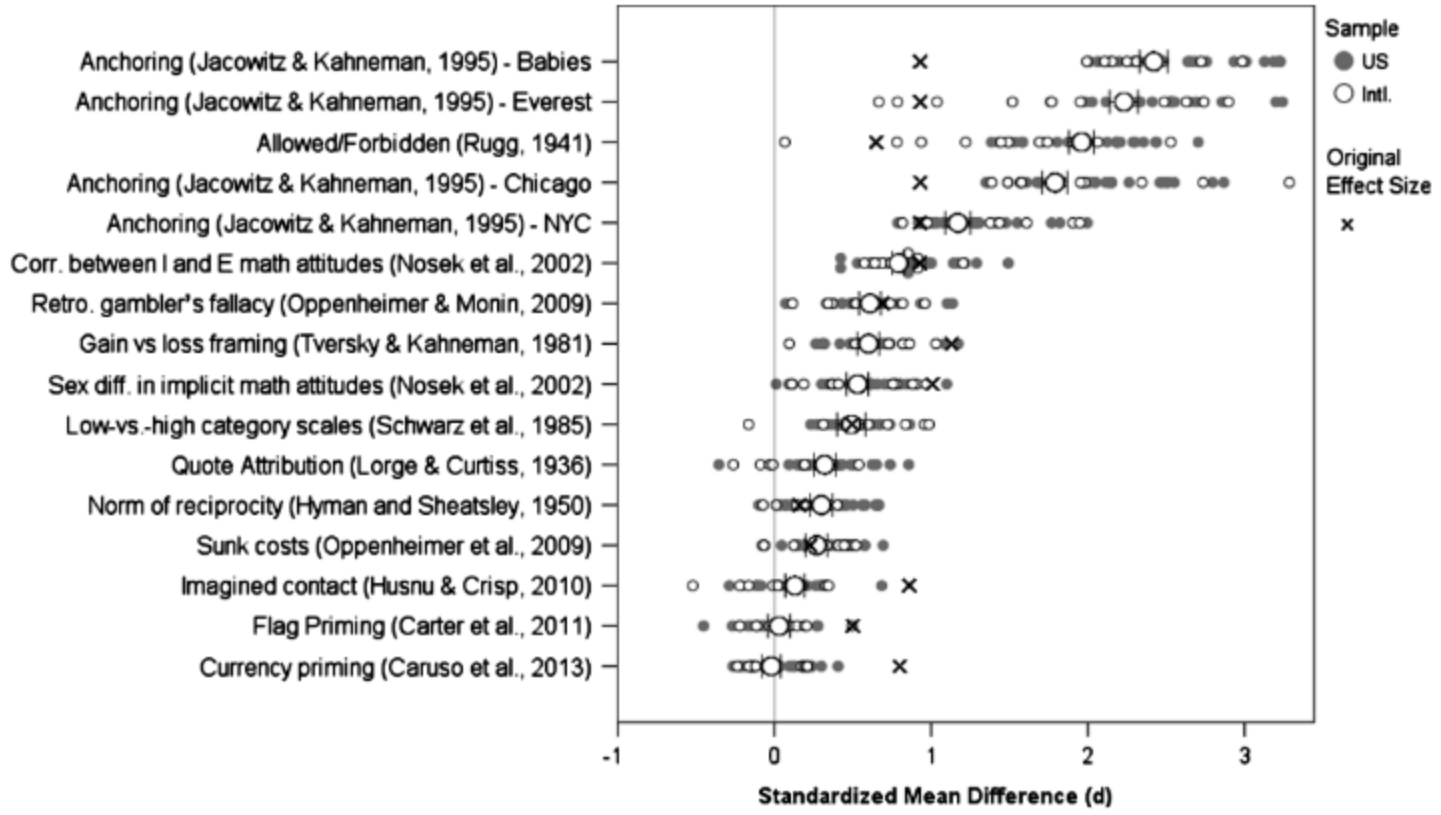
Are we facing a “scientific crisis”?!

Are we facing a “scientific crisis”?!

Can we Replicate Published Findings?

- **Nosek et al. (2015, Science):**
  - Large effort, attempting to replicate results from 100 psychology studies; depending on criteria used, only 36% to 47% of studies replicated
    - "collectively these results offer a clear conclusion: A large portion of replications produced weaker evidence for the original findings despite using materials provided by the original authors, review in advance for methodological fidelity, and high statistical power to detect the original effect sizes" (p. 943)
- **Gilbert et al. (2016, Science; see also Shrout & Rogers, 2018):**
  - Noted major problems with Nosek et al.'s procedures; e.g.:
    - Samples from different populations (e.g., Italians vs. Americans to study stereotypes RE African Americans; Americans asked about consequences of honeymoon vs. Israelis asked about consequences of military service)
    - Inconsistencies in procedures employed in original vs. replicated studies
  - When samples and procedures matched better (and were endorsed by original authors), replicate rates improved to 60-65%

Can we Replicate Published Findings?



Can we Replicate Published Findings?

- **Maxwell et al. (2015, American Psychologist):**
  - "the mere fact that a replication study yields a nonsignificant statistical result should not by itself lead to a conclusion that the corresponding original study was somehow deficient and should no longer be trusted, even if the replication study appears to have been adequately powered" (p. 495)
- **Anderson & Maxwell (2016, Psych Methods):**

Table 1
Six Replication Goals and Descriptions

No.	Goal	Recommended analysis	Success criterion
1	To inter the existence of a replication effect	Repeat analysis of original study	$p < .05$
2	To infer a null replication effect	Equivalence test	Confidence interval falls completely inside region of equivalence
3	To precisely estimate the replication effect size	AIPE, construct confidence interval for effect size	Effect size estimated with desired level of precision
4	To combine replication sample data with original results	Construct confidence interval for the average effect size of replication and original studies	Building on prior knowledge; more precise estimate of the effect of interest
5	To assess whether replication is clearly inconsistent with original	Construct confidence interval for the difference in effect sizes	Confidence interval for difference in effect sizes does not include 0
6	To assess whether replication is clearly consistent with original	Equivalence test, using confidence interval for the difference in effect sizes	Confidence interval for difference in effect sizes falls completely inside region of equivalence

Can we Replicate Published Findings (Cont'd)?

- **There is variance across sub-disciplines of psychology, when it comes to replications**
- **For example, Mitchell (2012, Perspectives on Psych Science):**

Table 2. Correlation of Lab-Field Effects by Subfield Classifications

PsycINFO classification (*n*)	*r*	*r*	Author's classification (*n*)
Social (80)	.53	.60	Social (79)
I-O (72)	.89	.82	I-O (98)
Personality (22)	.83	.84	Clinical (19)
Consumer (7)	.59	.59	Marketing (7)
Education (7)	.71	.87	Education (5)
Developmental (3)	−.82	−.88	Developmental (6)
Psychometrics/Statistics/Methods (19)	.61		
Human Experimental (5)	.61		

Note: Sample sizes reflect number of paired effect sizes. The PsycINFO classification excludes one pair of effects classified as "Environmental Psychology," and the author classification excludes two pairs of effects classified as "Health Psychology." Results exclude possible outlier effects from Mullen et al. (1991).

Can we Replicate Published Findings (Cont'd)?

- **Examples of published constructive replications from JAP:**
  - van Hooft & Noordzij (2009): replicated more basic research findings, showing that unemployed who attended a learning goal workshop were more likely to seek and find jobs than those who attended performance goal or control workshops
  - Schmidt & DeShon (2010): replicated Vancouver's findings of negative self-efficacy – performance relationship when performance ambiguity was high but not low
  - Schultze et al. (2012): failed to replicate biased information search as mechanism for EoC – but instead supported biased information evaluation

Suggestions for Improvement

- **How can we encourage more replications in our field?**

1. Appreciate value of empirical and practical contributions, as adding to and complementing theoretical contributions
   - Research Report sections can sometime help (cf. JAP vs. AMJ…)
2. Enhance transparency of methods reporting (e.g., RE sample and sampling, measures/experimental manipulations, etc.)
3. Training – e.g., ask PhD students to conduct replication study as a socialization tool
4. Perhaps consider new journals (or section of existing journals) devoted to replications (e.g., JPSP; JEP:G)
   - More likely in more basic than more applied research, given complexity of phenomena, context, sample, and data?

Can we Reproduce Published Findings?

- **Nuijten et al. (2016, Behavior Research Methods):**
  - Using R program, reported statistical reporting inconsistency in about half of ~30,000 articles examined in 8 psychology journals (including JAP); 13% of inconsistencies may have affected a statistical conclusion
  - Daniel Lakens and Thomas Schmidt took closer look at Nuijten et al.'s program and found considerable amount of "false positives"
- **PubPeer:**
  - A large-scale anonymous effort in recent years to flag inconsistencies in statistical reporting
  - Many false positives identified as well – e.g., false conclusions from effort to reproduce interaction effects
  - Hard to verify and engage in meaningful exchange, due to insistence of maintaining anonymity of so-called "whistle blowers"
  - But, many instances found to be "false positive"

Can we Reproduce Published Findings (Cont'd)?

- **Retractions at Leadership Quarterly and JAP:**
  - Common theme is inability to reproduce findings
- **Recent commentaries at JAP**
  - Served as vehicles to both (a) "correct the record" and (b) advance new scholarly insights regarding published findings
  - e.g., publication bias and stereotype threat research (Comment by Zigerell, and Reply by Ryan & Nguyen; 2017)
- **APA's Journal Article Reporting Standards (2018, American Psychologist)**

Suggestions for Improvement

- **How can we improve reproducibility?**
  - **Improve culture of statistical reporting** – e.g., better information of model specification, report S.E.s in addition to b-weights, provide basic results in addition to more complex (e.g., descriptives and correlations; simpler effects before more complex ones, etc.); see **APA's JARS**
  - **Encourage data sharing –** could be tricky in some applied contexts, but would enhance analytical diligence and accountability
  - **Enhance quality of training** – graduate students, junior faculty, as well as reviewers and AEs (e.g., PDWs for editors and reviewers?)
  - **Set better journal review processes** – e.g., Editor works with AEs, computer programs as aids, etc.
  - **Hold authors accountable** (e.g., corrections and retractions…)

Are Questionable Research Practices Prevalent?

- **Bedeian al. (2010, AMLE):**

TABLE 1
Percentage of Management Faculty Who Reported Knowledge of Faculty Engaging in the Listed Behavior Within the Previous Year

Behavior	All	Tenured	Nontenured
Category I – Fabrication, Falsification, and Plagiarism			
1. Withheld methodological details or results	79.2	79.7	78.4
2. Selected only those data that support a hypothesis and withheld the rest	77.6	77.9	77.1
3. Used another's ideas without permission or giving due credit	72.1	75.3	67.3
4. Dropped observations or data points from analyses based on a gut feeling that they were inaccurate	59.6	62.3	55.6
5. Withheld data that contradicted their previous research	49.5	50.6	47.7
6. Fabricated results	26.8	26.4	27.5
Category II – Questionable Research Practices			
7. Developed hypotheses after results were known	91.9	92.2	91.5
8. Published the same data or results in two or more publications	86.2	88.7	82.4
9. Developed "ins" with journal editors	83.3	82.3	85.0
10. Inappropriately accepted or assigned authorship credit	78.9	82.3	73.9
11. Circumvented aspects of human-subjects requirements	58.1	61.9	52.3

Are Questionable Research Practices Prevalent (Cont'd)?

- **Bosco et al. (2016, PPsych):**
  - Found evidence that hypothesized correlations are larger than non-hypothesized correlations, in articles published in JAP and PPsych (Study 1) and meta-analysis of the job satisfaction-job performance relationship (Study 2)

- **O'Boyle et al. (2017, JOM):**
  - Compared results reported in over 150 IO-OB-HR dissertations to those reported in resulting published journal articles
  - Found that supported dissertation hypotheses were more likely to reappear in published articles than non-supported ones (and non-supported hypotheses were more likely to be dropped)

Are Questionable Research Practices Prevalent (Cont'd)?

- **Fiedler & Schwarz (2016, Social Psychological & Personality Science)**
  - "one of the more widely cited publications on QRPs in psychological research (John et al., 2012) suffers from ambiguities that prohibit the damning conclusions drawn" (p. 50)
  - "The survey failed to give respondents the opportunity to clarify such ambiguities. Any clarifications respondents may have tried to offer in comment boxes were not considered in the analyses" (p. 50)
  - Revised study found less pessimistic evidence
- **Similar issues with Banks et al. (2016, JOM):**
  - "Have you ever "rounded off" a p-value (reporting that a p-value of .054 is <.05 rather than p =.05)?"
  - "Have you ever engaged in HARKing?"

Are Questionable Research Practices Prevalent (Cont'd)?

- **Fiedler & Schwarz (2016, Social Psychological & Personality Science):**

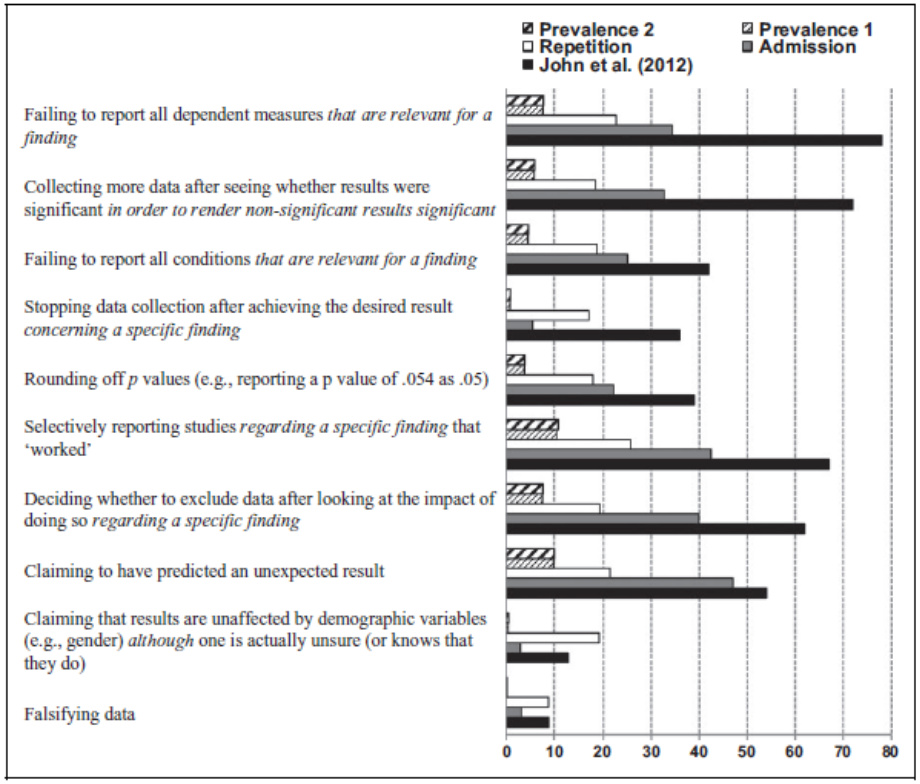


Figure 2. Prevalence indices (shaded bars) derived from admission rates of respondents committing questionable research practices at least once (gray bars) and repetition frequency (white bars), compared to the original John et al. (2012) data (black bars). Modified item wordings appear in italics.

Are Questionable Research Practices Prevalent (Cont'd)?

- HARKing assumes (a) one's a-priori theory and hypotheses are strong and valid, and (b) one adopts a deductive scientific approach – both of these assumptions have limits (see Hollenbeck & Wright, 2017, JOM)

- Many articles reporting full support for hypotheses get rejected – hard to evaluate fully in research on QRPs

- Examples of articles with null findings do exist (e.g., Chen, 2005, AMJ; Chen & Klimoski, 2003, AMJ; Sacco et al., 2003, JAP; Schultze et al., 2012, JAP)

Suggestions for Improvement

- **How can we reduce QRPs?**
  - Many solutions proposed, but not always simple or straightforward; e.g.:
    - "preliminary editorial decisions (i.e., accept or reject) could be formed prior to reviewers' and editors' knowledge of results" (Bosco et al., 2016, PPsych, p. 744)
    - Pre-registration of studies (e.g., Nosek et al., 2018, PANAS)
  - **Greater transparency** can help; e.g.:
    - Honest positioning and description of study's approach (inductive or semi-deductive, positioning of research questions vs. hypotheses, etc.)
    - Inclusion of auxiliary/post-hoc analyses and results
  - Appreciate **multiple forms of scientific discovery and contributions** (see Kilduff et al., 2011, AMR; Mathieu, 2016, JOB; Pillutla & Thau, 2013, OPM)
  - **Important roles of Editors and editorial teams – but also authors**!

Suggestions for Improvement

- **A "system-based" view (Chen, 2018)**
  - **Suggestions for improving scientific rigor (a) prereview practices, (b) review process practices, and (c) post-publication practices**

American Psychological Association

Journal of Applied Psychology

0021-9010/18/$12.00

2018, Vol. 103, No. 4, 359–361
http://dx.doi.org/10.1037/apl0000313

EDITORIAL

Supporting and Enhancing Scientific Rigor

Summary and Thoughts

- Overall, there is much to like about our field's self-reflection and systematic evaluation of scientific adequacy and rigor
- Some evidence certainly suggests problems exist – but disagreements exist on extent and scope of problems
- Let's approach this scientifically and methodically
  - How much of a role do statistical results play in ultimate publication decisions and downstream impact (vs. say, importance of questions, conceptual development, adequacy of methods)?
  - No one silver bullet when it comes to enhancing scientific rigor in the field – a system and culture of scientific rigor (Chen, 2018)