Is psychological science in a crisis?

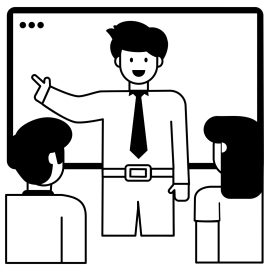
The Social Brain: Critical Perspectives on Science, Society and Neurodiversity

Richard Ramsey

Today

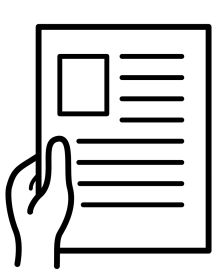
Part 1

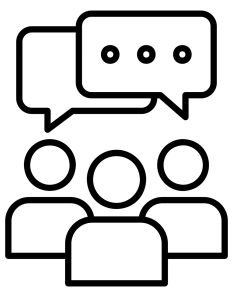
- Is psychological science in a crisis?



Part 2

- Read articles and discuss





Introduction

What is science?

Science

Not science

Estimating reproducibility

RESEARCH ARTICLE

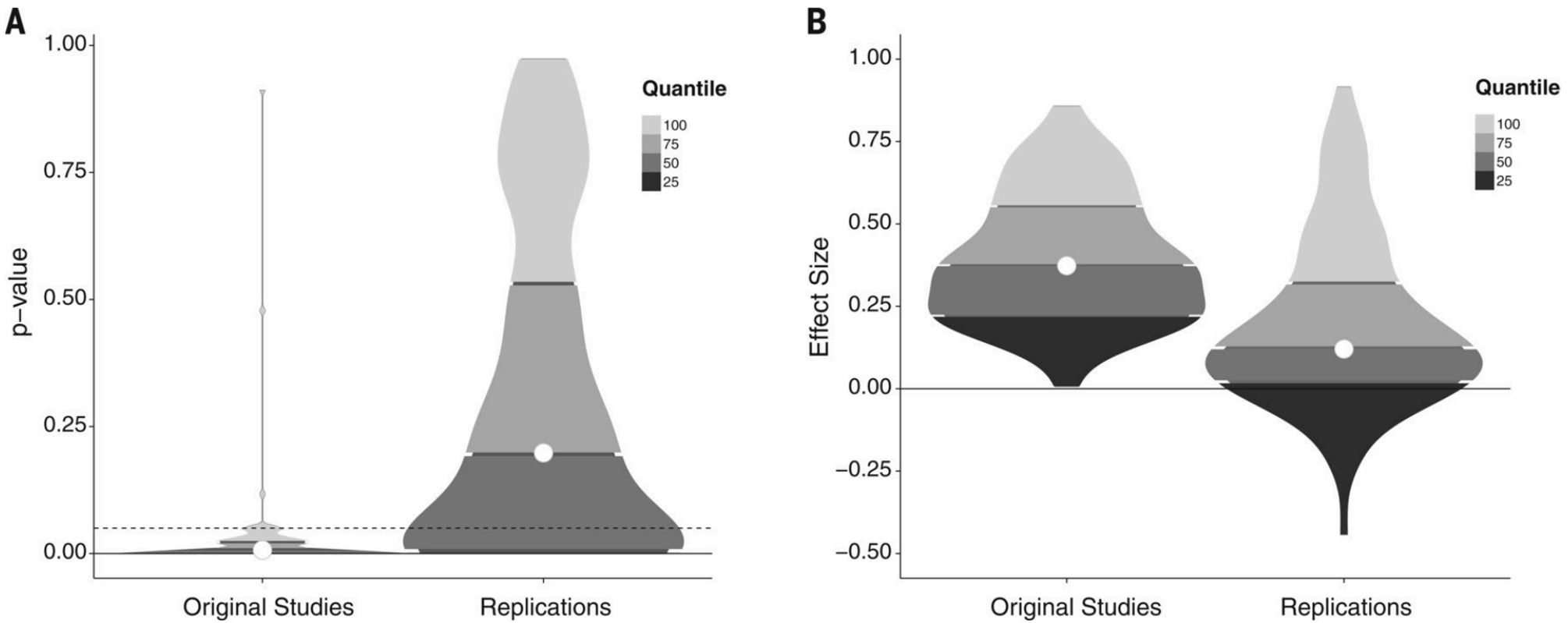
PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*†

- Attempted to replicate 100 experiments from 3 leading Psychology journals.
- contacted original authors for materials and hypotheses.
- pre-registered analysis plans.
- high power designs (e.g., much larger samples than the originals).

Estimating reproducibility



Estimating reproducibility

	Replications $P < 0.05$ in original direction	Percent
Overall	35/97	36
JPSP, social	7/31	23
JEP:LMC, cognitive	13/27	48
PSCI, social	7/24	29
PSCI, cognitive	8/15	53

ANALYSIS

Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button[1,2], John P. A. Ioannidis[3], Claire Mokrysz[1], Brian A. Nosek[4], Jonathan Flint[5], Emma S. J. Robinson[6] and Marcus R. Munafò[1]

IN FOCUS NEWS

CANCER BIOLOGY

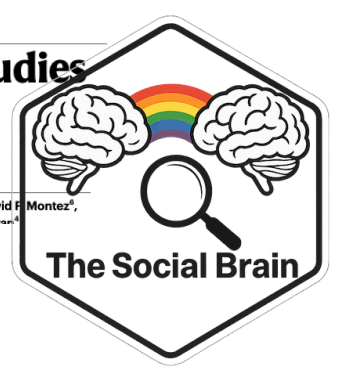
Reproducibility project yields muddy results

An ambitious effort to replicate cancer studies is provoking controversy.

Article

Reproducible brain-wide association studies require thousands of individuals

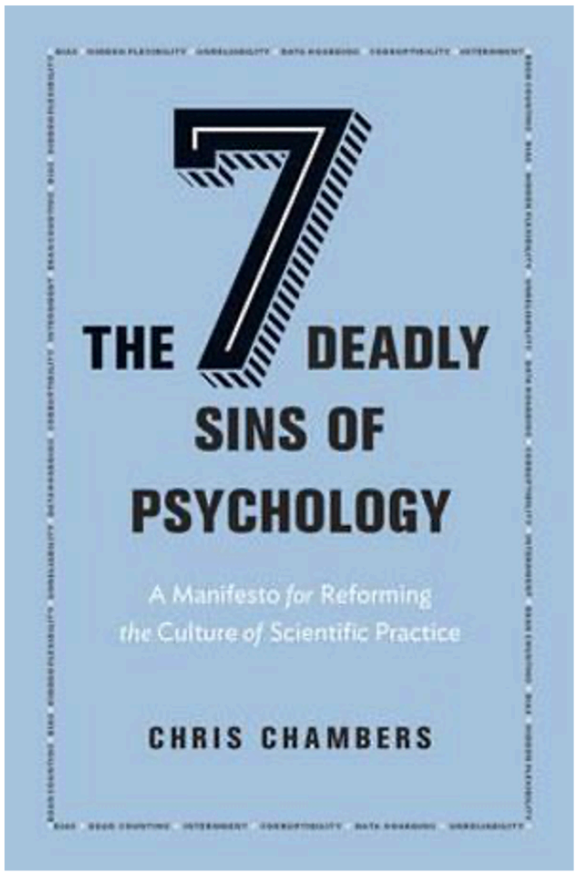
https://doi.org/10.1038/s41586-022-04492-9 Scott Marek[1,30], Brenden Tervo-Clemmens[2,3,30], Finnegan J. Calabro[4,5], David F. Montez[6], Benjamin P. Kay[6], Alexander S. Hatoum[1], Meghan Rose Donohue[1], William Foran[1]



7 Deadly Sins of Psychology

7 Deadly Sins of Psychology

1. Bias
2. Hidden Flexibility
3. Unreliability
4. Data Hoarding
5. Corruptibility
6. Internment
7. Bean Counting



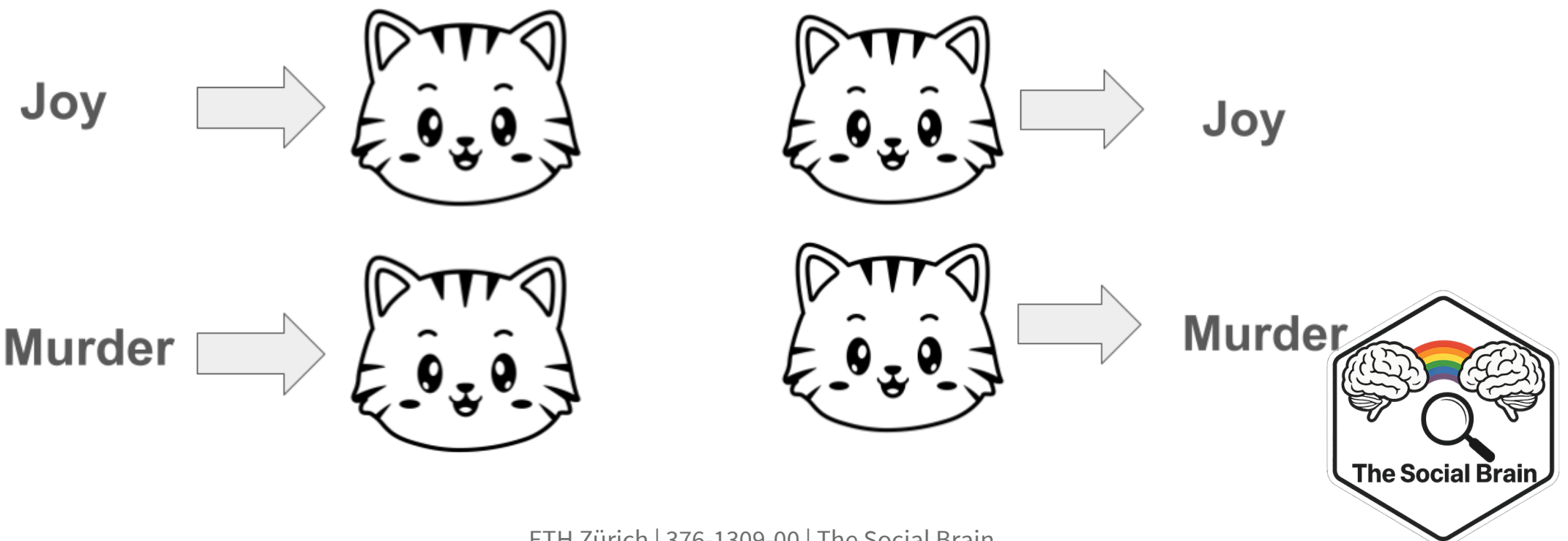
Bias

> The human understanding when it has once adopted an opinion … draws all else to support and agree with it.

Francis Bacon, 1620

Bias

- Positive and new trumps the negative but true (“A beautiful story undone by an ugly truth”)
- An extreme version: Daryl Bem and “precognition” – 2011.

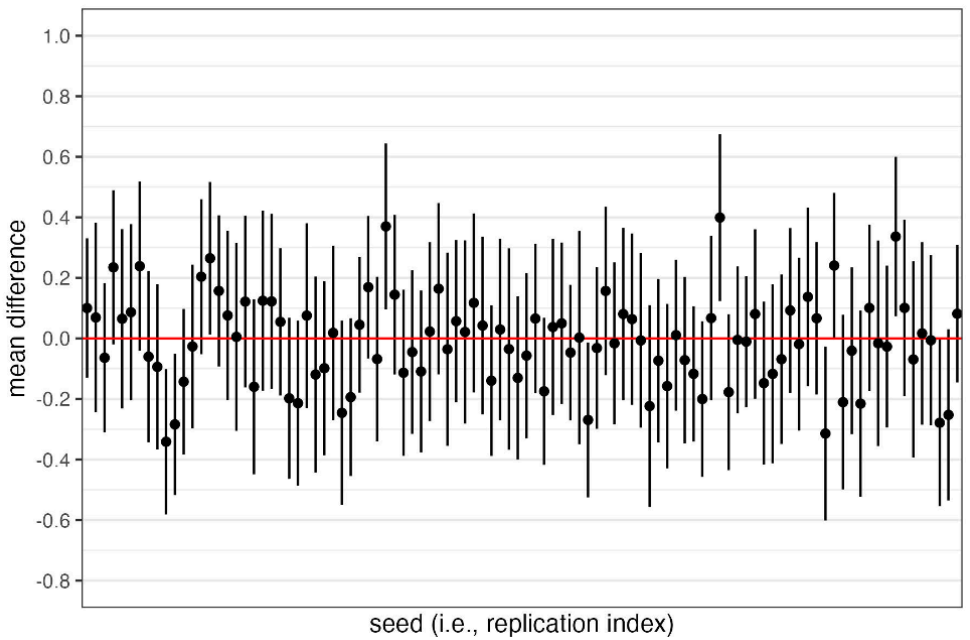


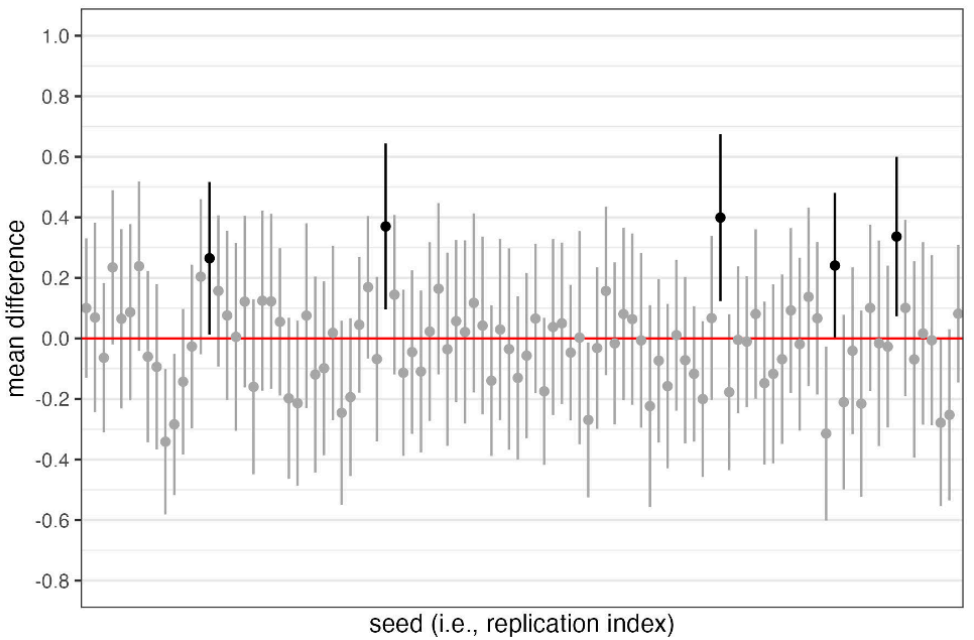
Bias

- Publication bias (or the “file-drawer effect”).
- Studies that fail to show a statistically significant effects, or that reproduce the work of others, have such low priority that they are effectively censored from the scientific record.
- They either end up in the file-drawer or never get conducted in the first place.

Bias

- Publication bias (or the “file-drawer effect”).





Bias

- Confirmation bias - we see what we expect and only check when there are deviations
- The scientific literature is dominated by results that confirm hypotheses, rather than fail to confirm.
- An arms-race has started to show new and “sexy” results, which confirm hypotheses, but ultimately bias the published literature.

Bias

- Hindsight bias
- HARKing (Hypothesising After the Results are Known)

Hidden Flexibility

> Torture numbers and they will confess to anything.

Gregg Easterbrook, 1999

Hidden Flexibility

Psychological Science
22(11) 1359–1366
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797611417632
http://pss.sagepub.com
SAGE

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

Abstract

In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings (≤ .05), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.

Study 2: musical contrast and chronological rejuvenation

Using the same method as in Study 1, we asked 20 University of Pennsylvania undergraduates to listen to either "When I'm Sixty-Four" by The Beatles or "Kalimba." Then, in an ostensibly unrelated task, they indicated their birth date (mm/dd/yyyy) and their father's age. We used father's age to control for variation in baseline age across participants.

An ANCOVA revealed the predicted effect: According to their birth dates, people were nearly a year-and-a-half younger after listening to "When I'm Sixty-Four" (adjusted $M = 20.1$ years) rather than to "Kalimba" (adjusted $M = 21.5$ years), $F(1, 17) = 4.92$, $p = .040$.

Hidden Flexibility

- p-hacking
- Peculiar patterns of p
- Ghost hunting (or lack of high-quality replication)
- Biased debugging
  - If the results turn out as expected, do no further checks. If the results are a surprise, then check the code for bugs.

Unreliability

> And it's this type of integrity, this kind of care not to fool yourself, that is missing to a large extent in much of the research in cargo cult science.

Richard Feynman, 1974

Unreliability

- Was Einstein wrong?
  - Particles break light-speed limit … September 2011

- Answer…. No.
- By June 2012, three independent teams had failed to replicate the original result. Neutrinos intheir experiments travelled at approximately the speed of light.
- One month later, the original team reported that their findings were caused by a loose fibre-optic cable!!
- But this is just what should happen in a good scientific environment, where human error will certainly happen.

Unreliability

- Sources:
  - disregard of direct replication
  - lack of power
  - failure to disclose methods
  - statistical fallacies
  - failure to retract

Data hoarding

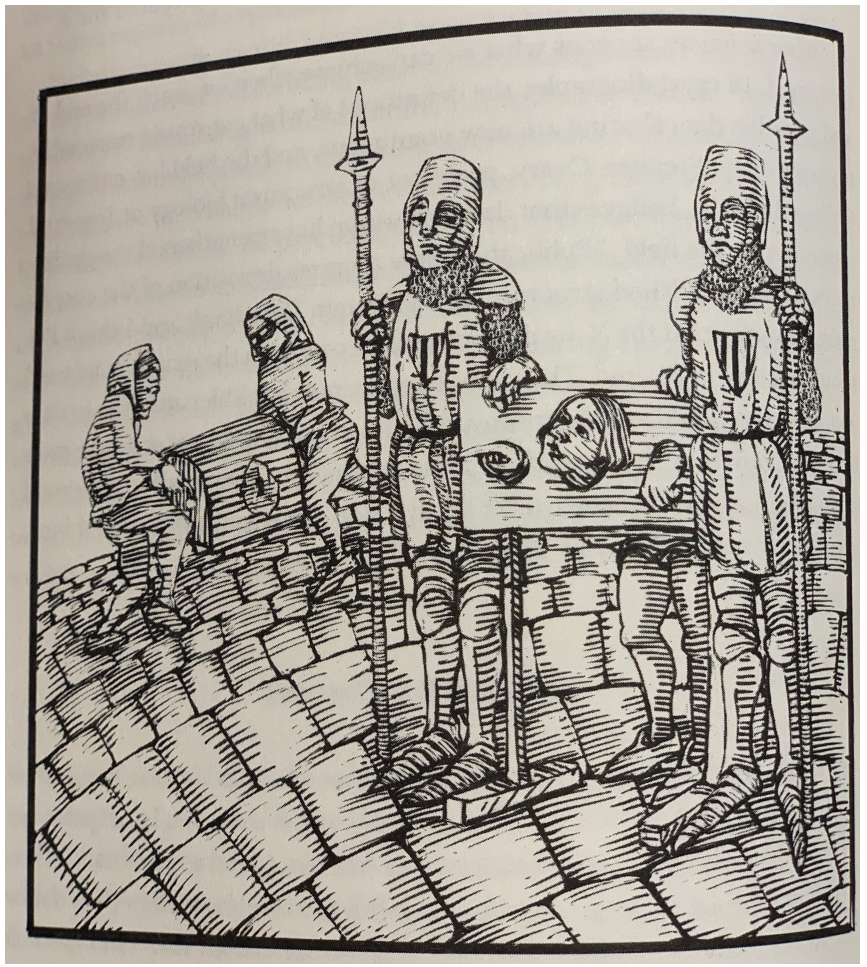
> Code and data or it didn’t happen.

Anon.

Data hoarding

- A story from my first paper submission (many years ago…).
- Failure to share
- Secret sharing
- Failure to share hides misconduct

Corruptibility



> I was doing fine, but then I became impatient, overambitious, reckless.

Diederik Stapel, 2012

Corruptibility

- Fraud examples
- https://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html
- 58 retracted papers
- Most (or all) for data fabrication
- https://retractionwatch.com

Corruptibility

- Science Fictions is another entertaining read
- It emphasizes that Fraud, Bias, Hype and Questionable Research Practices are not restricted to Psychological research.
- In fact, it is common in medical research and the drivers are largely the same
- This might be surprising given the consequences of medical research

The Social Brain

Internment

> Publish means “make public.”

Mike Taylor, 2012

Internment

- Not only are data not publicly avaiable (as we discussed earlier), but the article and conclusions are normally kept behind a paywall.
- This is morally questionable, given that the vast majority of research across the world is funded by the public via taxes.
- Univeristy libraries tend to have access via a subscription (so academics can read papers), but the general public do not.

Internment

- Why do Psychologists support barrier-based publishing?
  - Prestige associated with barrier-based journals
  - Incentives to publish in key journals for career progression
  - There was no real alternative (historically)

Bean counting

> Not everything that counts can be counted, and not everything that can be counted counts.

Gregg Easterbrook, 1999

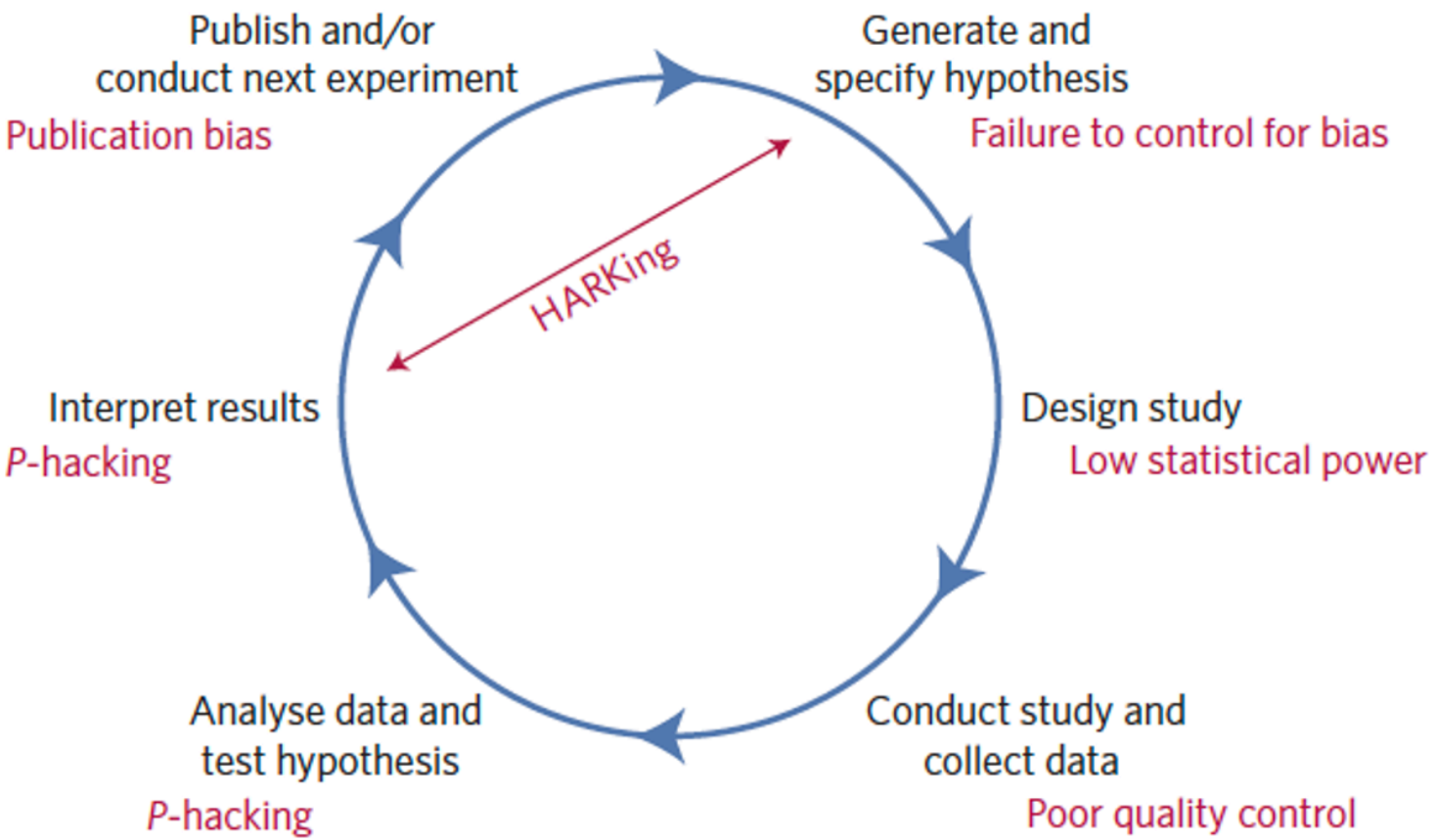
Bean counting

Can science measure itself?

- Metrics are used to measure the “quality” of science and scientists e.g., impact factor of a journal, grant capture, number of papers published, author position, number of citations.
- So folks shoot for these by trying to game the system in anyway they can, which tends to encourage poor research rather than better research.
- **Goodhart’s law:** when a measure becomes a target, it ceases to be a good measure.

Summary

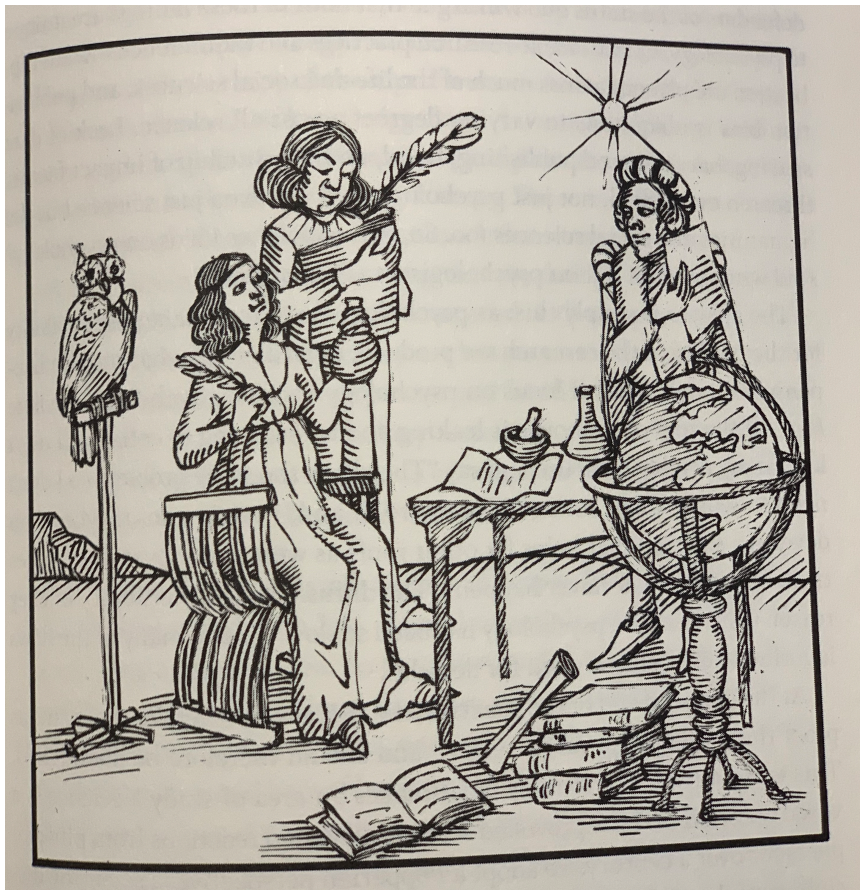
Where are we now?



Is psychological science in a crisis?

What do you think?

Redemption



> The method of science, as stodgy and grumpy as it may seem, is far more important than the findings of science.

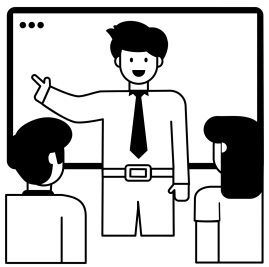
Carl Sagan, 1995

Note: This is a prelude to next week

Today

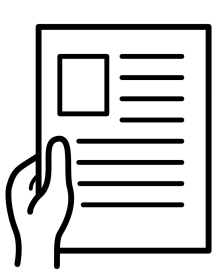
Part 1

- Is Psychological science in a crisis?



Part 2

- Read articles and discuss



Take a break

Part 2 - Read and discuss

Discussion material

- break into small groups (~ 5 per group)
- discuss aspects of the lecture – 7 deadly sins
- discuss aspects of the journal article: https://www.science.org/doi/full/10.1126/science.aac4716
- can you think of any solutions?

References

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

Munafʻo, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017 online 10). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021. https://doi.org/10.1038/s41562-016-0021

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Acknowledgements

- Unless otherwise specified, icons and images were used under license from The Noun Project
- Slides were created with Quarto and RevealJS