

Who is afraid of “data dredging”?



Case Study 1

(Hollenbeck, 2017)

Design and data collection

- A graduate student, desperate to get a job, chooses 30 popular survey measures and partitions them into two questionnaires, each 15 questions.
- Recruits 2,000 Mechanical Turk workers for pennies per hour, and sends out the questionnaires separated in time by 2 months



Analysis and publication

- At the end of the 2 months, the 30×30 correlation matrix produces 20 significant correlations, some of which cross the Time 1 – Time 2 divide.
- Ignore significant correlations that make no theoretical sense.
- Others are HARK-ed into mediator role, woven into a **post-hoc** narrative, presented as an **a-priori** causal model, written up and published in a top journal.

Aftermath

- Several research teams fail to replicate the results
- Speculation whether the student was incompetent, unethical, or just very lucky,
- Everyone moves on, case forgotten

Case Study 2

(Hollenbeck, 2017)

Design and data collection

- Past literature suggests that a certain drug might cure a novel viral disease.
- Recruiting 100 patients, team launches RCT, two years of study find $\rho = 0.10, p = 0.32$

The discovery – validation dance

- Anecdotal rumours surface of women cured by the treatment
- Data stratified by gender: $\rho_{men} = 0.00, p > 0.1, \rho_{women} = 0.20, p_{women} = 0.08$.
- Speculations about the (insignificant) interaction between treatment and sex.
- Post-hoc literature survey suggests to deductively hypothesize interaction between treatment – estrogen level
- Same data is stratified by estrogen levels. $\rho_{estrogen} = 0.50, p_{estrogen} = 0.03$
- The **Discussion Section** of their manuscript the team notes these were result of an **exploratory analysis** of the data that **after** the main effects for the drug were found to be nonsignificant

Aftermath

- Large scale replication reveals a correlation of $\rho = .40$ with much larger sample size, highly statistically significant.
- “discovery saves thousands of lives”



Search life-sciences literature (42,454,870 articles, preprints and more)

[Advanced search](#)

[Abstract](#)

[Full text](#)

[References](#)

[Citations & impact](#)

[Similar Articles](#)

HARKing: hypothesizing after the results are known.

Kerr NL ¹

[Author information](#) ▶

Personality and Social Psychology Review : an Official Journal of the Society for

Personality and Social Psychology, Inc, 01 Jan 1998, 2(3):196-217

DOI: 10.1207/s15327957pspr0203_4 PMID: 15647155

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

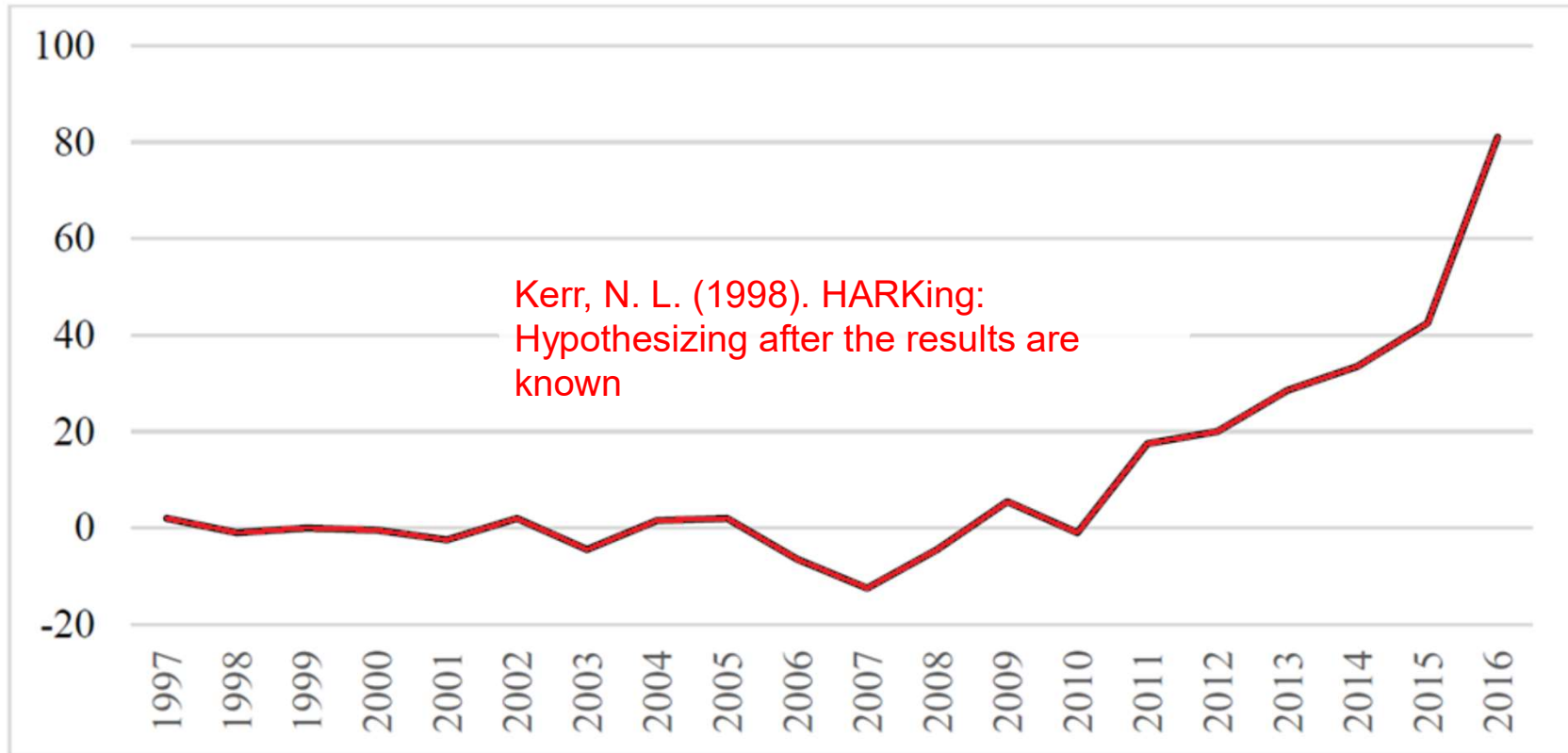


Figure 1. Difference in number citations per year to Kerr's (1998) seminal article on HARKing compared to the average number of citations to two other articles that were published in the same issue of the same journal. The difference in citations remained relatively low during the period 1997-2010 and then increased dramatically after 2011. Data is sourced from Google Scholar.

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

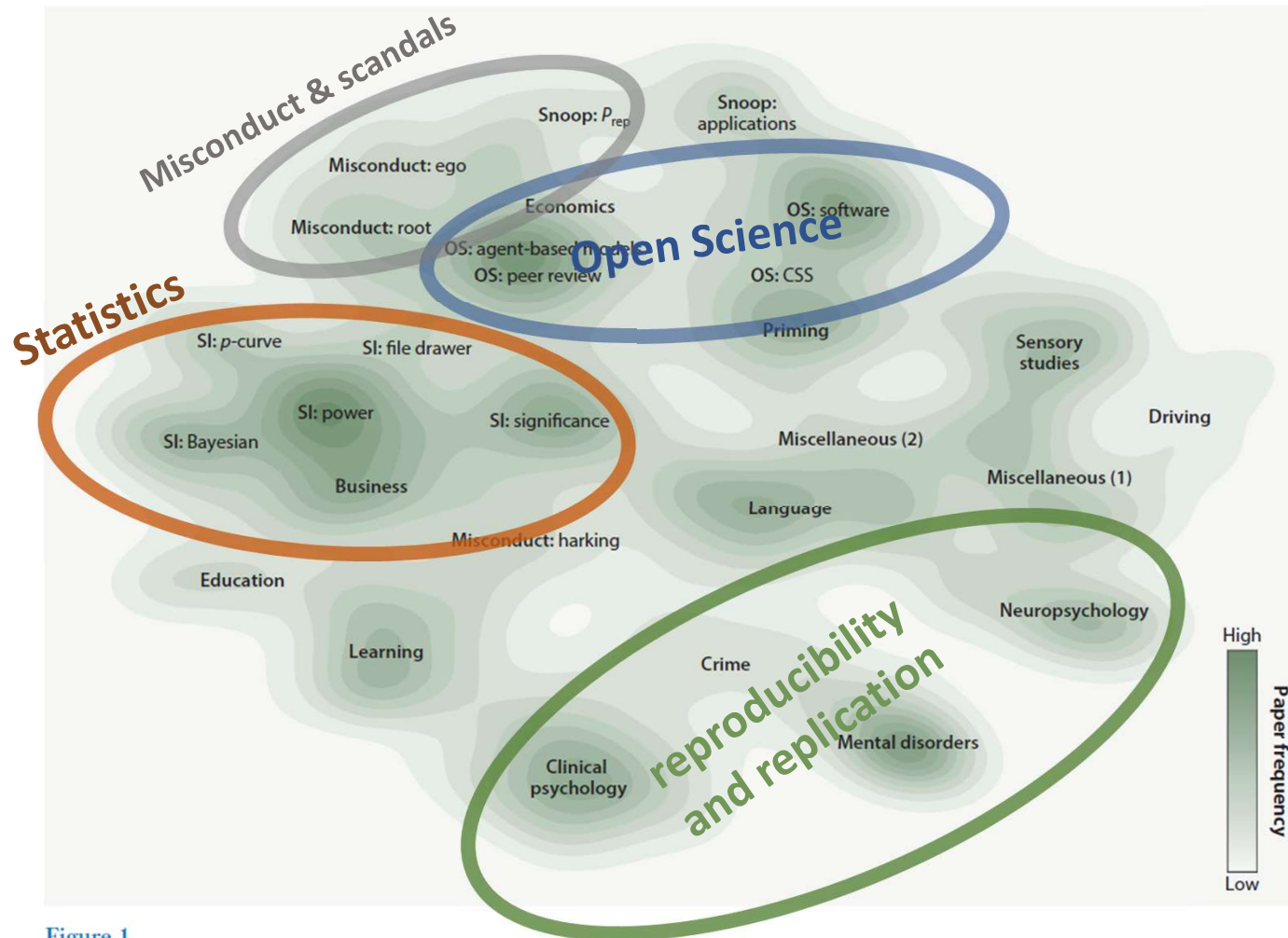


Figure 1

Replication research intellectual landscape. This is a graphical representation mapping the topic similarity network of 1,947 papers on replication in the social sciences. The space is constructed to maximize co-location of topically similar papers while placing dissimilar papers far apart, which results in topical clusters being close together. For visual clarity, we then fit a two-dimensional kernel-density estimate for the number of papers at each point in the space, so clusters emerge as hills in the map. Abbreviations: CSS, computational social science; OS, open science; SI, statistical issues.

Moody, J. W., Keister, L. A., & Ramos, M. C. (2022). Reproducibility in the social sciences. *Annual Review of Sociology*, 48, 65-85.

What is wrong?

- Replication crisis
- Culprit: bias in favor of publishing false positives
- Why now? If **big data** → **power** → **significance** too easy, why is there a replication crisis?
- Is it about intentions or consequences?

Costs and benefits?

- A modest proposal (Benjamin et al 2018)
- Prevention of HARK-ing: does it come at a cost?

The role of serendipity in drug discovery

Thomas A Ban ¹

Affiliations + expand

PMID: 17117615 PMID: [PMC3181823](#) DOI: [10.31887/DCNS.2006.8.3/tban](#)

[Free PMC article](#)

Abstract in English, [Spanish](#), [French](#)

Serendipity is one of the many factors that may contribute to drug discovery. It has played a role in the discovery of prototype psychotropic drugs that led to modern pharmacological treatment in psychiatry. It has also played a role in the discovery of several drugs that have had an impact on the development of psychiatry. "Serendipity" in drug discovery implies the finding of one thing while looking for something else. This was the case in six of the twelve serendipitous discoveries reviewed in this paper, i.e., aniline purple, penicillin, lysergic acid diethylamide, meprobamate, chlorpromazine, and imipramine. In the case of three drugs, i.e., potassium bromide, chloral hydrate, and lithium, the discovery was serendipitous because an utterly false rationale led to correct empirical results; and in case of two others, i.e., iproniazid and sildenafil, because valuable indications were found for these drugs which were not initially those sought. The discovery of one of the twelve drugs, chlordiazepoxide, was sheer luck.

Editorial Commentary

Harking, Sharking, and Tharking:
Making the Case *for* Post Hoc
Analysis of Scientific Data

John R. Hollenbeck
Michigan State University
Patrick M. Wright
University of South Carolina

Secretly HARK-ing in the Introduction section

HARK-*ing*

Hypothesizing After Results Are Known (HARK)

SHARK-*ing*

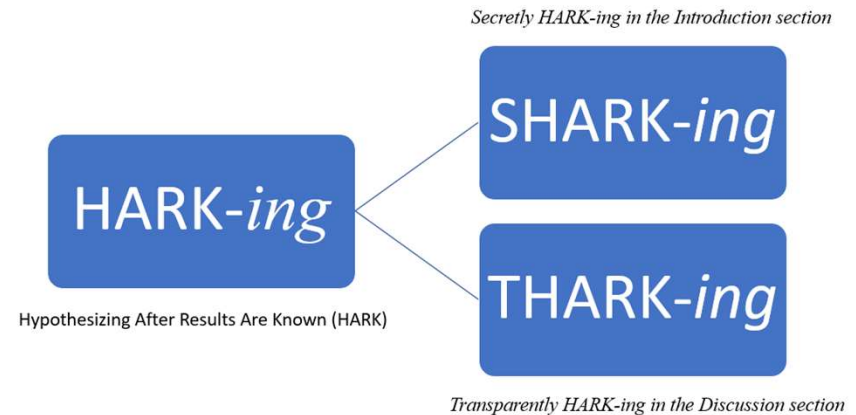
THARK-*ing*

Transparently HARK-ing in the Discussion section

Editorial Commentary

Harking, Sharking, and Tharking: Making the Case *for* Post Hoc Analysis of Scientific Data

John R. Hollenbeck
Michigan State University
Patrick M. Wright
University of South Carolina



We argue that the discussion sections of all empirical papers should include a subsection that reports post hoc exploratory data analysis. We explain how authors, reviewers, and editors can best leverage post hoc analyses in the spirit of scientific discovery in a way that does not bias parameter estimates and recognizes the lack of definitiveness associated with any single study or any single replication. We also discuss why the failure to Thark in high-stakes contexts where data is scarce and costly may also be unethical.

Takeaways

“Wicked problems”* have no silver bullets

“The root of all science is a dance between discovery and validation”**

*Churchman, C. W. (1967). Guest editorial: Wicked problems. *Management science*.

**Moody et al (2022) Reproducibility in the social sciences, *Annual Review of Science*