THE REPLICATION CRISIS

KEY CONCEPTS

"While collecting and analyzing data, researchers have many decisions to make, including whether to collect more data, which outliers to exclude, which measure(s) to analyze, which covariates to use, and so on. If these decisions are not made in advance but rather are made as the data are being analyzed, then researchers may make them in ways that self-servingly increase their odds of publishing (Kunda 1990). Thus, rather than placing entire studies in the file-drawer, researchers may file merely the subsets of analyses that produce nonsignificant results. We refer to such behavior as *p-hacking*" (Simonsohn, Simmons, and Nelson 2014, 534).

EARLY WARNING SIGNS

- Kunda's (1990) "The Case for Motivated Reasoning" *Psychol Bull* 108(3): 480–498.
- Kerr's (1998) "HARKing: Hypothesizing After the Results are Known" *Pers Soc Psychol Rev* 2(3): 196–217.

THE CONFLAGRATION ERUPTS

- Ioannidis' (2005) "Why Most Published Research Findings Are False" *PLoS Med* 2(8): 0696–0701.
- Simmons, Nelson, & Simonsohn's (2011) "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant" *Psychol Sci* 22(11): 1359–1366.
- Simonsohn, Simmons, & Nelson's (2014) "*P*-Curve: A Key to the File Drawer" *J Exp Psychol Gen* 143(2): 534–547.
- Wicherts et alia's (2016) "Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid *p*-Hacking" *Front Psychol* 7: 1832, 1–12.

DISCUSSION QUESTIONS

1. When an exciting new result is announced in your scientific field, how much credence do you tend to give it? Has this changed since the replication crisis?
2. Can you name an example of a researcher "degree of freedom" from your own area of work? This should be something which you can identify as an upstream choice which might later affect downstream estimations of the significance of your results.
3. What is p-hacking? Does it matter whether p-hacking occurs consciously or unconsciously in science?
4. Is the idea of comprehensive, non-binding research (pre)registration a good fit for your scientific field?
5. Without appealing to convention (or tradition, or any other similar conceptual analogue), can you provide a defense of the notion that a p-value of < 0.05 is meaningful, whereas a p-value of ≥ 0.05 is not meaningful?

IN SOCIAL CONTEXT

The Replication Crisis (or RepliGate, as some like to call it) has exposed not just one problem or instability within and outside of science but rather many. First, there is the concern it raises within science and among scientists about the reliability of their work and the literature so much of that work attempts to build on. Failures of replicability are also costly, and this is a problem not just for the scientists but also for their funders (public included). But wasting resources is not the only way to erode trust; trust can also be eroded when results are announced as fact and knowledge one day, only to be taken back the next. Scientists may know the difference between settled and unsettled science, but the public has not always been taught about science with that level of nuance.

FINE PRINT

In 1992–3, the National Academy of Sciences (NAS) published a pair of reports on *Responsible Science* (Vol. 1–2), and those reports ushered in an era of ethical oversight centered around the concept of the Responsible Conduct of Research (RCR) at federally-funded American research institutions across the nation. By 2009, the National Institutes of Health (NIH) had mandated that "all trainees, fellows, participants, and scholars receiving support through any NIH training, career development award (individual or institutional), research education grant, and dissertation research grant must receive instruction in responsible conduct of research" (NOT-OD-10-019). The National Science Foundation (NSF) recommends—though does not require—something similar. Both agencies suggest that satisfactory RCR instruction tends to cover: research misconduct; conflict of interest; human subjects research; animal subjects research; collaboration and interdisciplinarity; data acquisition and management; authorship, peer review, and publication; mentoring and being mentored; and the relationship between science and society. Because of its import for the health of the relationship between science and society, this handout introduces the topic of **the replication crisis.**

THE UNIVERSITY OF UTAH

THE OFFICE OF RESEARCH INTEGRITY

Sponsored by ORI grant IR200063-01