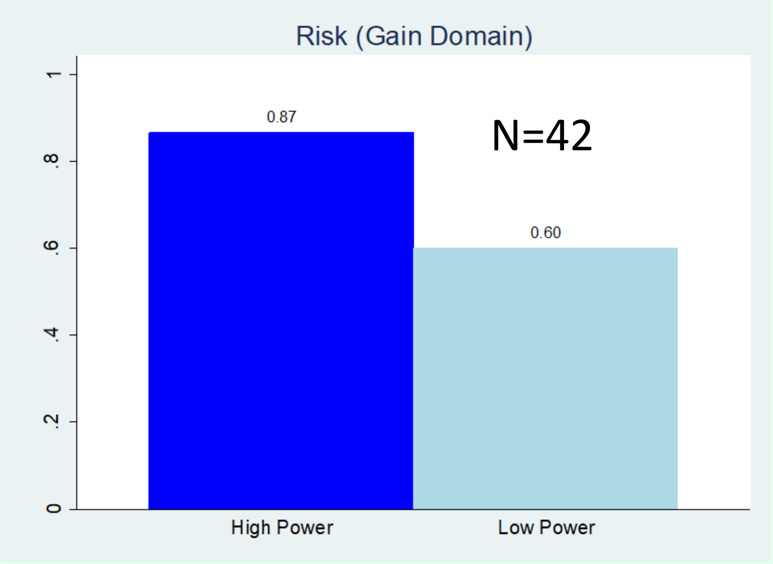
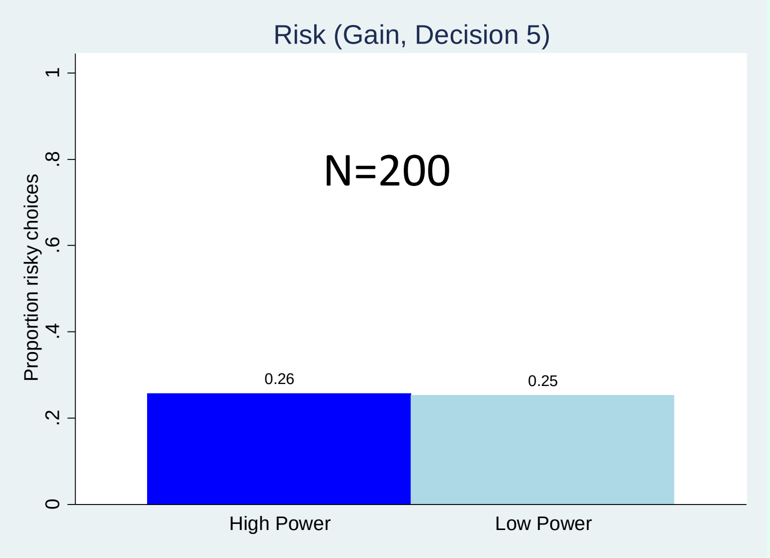
Funding flexibility, replicability and team science

Anna Dreber
Stockholm School of Economics

International Conference on Science and Technology for Sustainability 2025, 2026-02-11

Power posing





Carney et al. 2010, Ranehill et al. 2015

This talk

- Curiosity-driven research
- We learn new things during projects and change focus - how should funders react?
  - Particularly important for early career reseracher?

- My experiences on how funding can facilitate risk seeking and novel research
  - As researcher
  - As evaluator

- Team science and need to think about incentives

Different funding styles

- Project/program with detailed budget that exactly needs to be followed
- …
- No requirements

- Accountability and predictability
- Person vs project
  - Biases, information
- Incentive effects

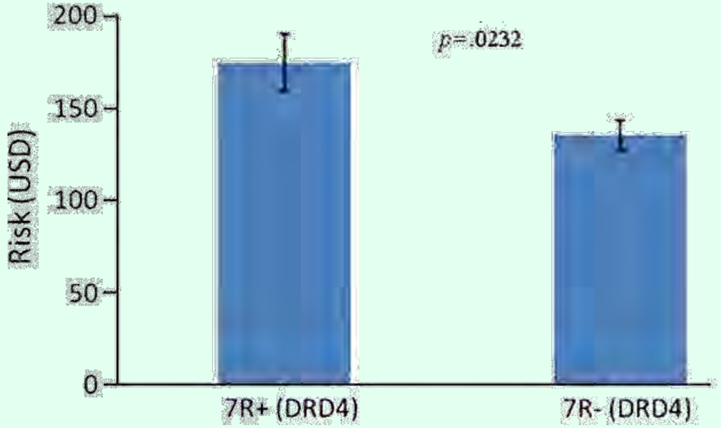
Example

- Wallenberg Academy Fellows
- Funded in 2013 (2nd cohort)
  - Detailed application
  - Behavioral economics
    - Hormones, genes and economic decision making
    - Cooperation in the repeated prisoner's dilemma

Knut and Alice Wallenberg Foundation

- Also other project funding for gene and risk taking project

Problems and opportunities



NEWSFOCUS

Psychology's Bold Initiative

In an unusual attempt at scientific self-examination, psychology researchers are scrutinizing their field's reproducibility

PICK UP THE JANUARY 2008 ISSUE OF *Psychological Science*, turn to page 49, and you'll find a study showing that people are more likely to cheat on a simple laboratory task if they have just read an essay arguing that free will is an illusion. It was a striking study that drew widespread attention both from psychologists and from many media outlets. But should you believe the result?

There's no reason to think that the study, conducted by psychologists Kathleen Vohs of the University of Minnesota Carlson School of Management in Minneapolis, and Jonathan Schooler, who is now at the University of California, Santa Barbara (UCSB), is incorrect. Yet according to many psychologists, their field has a credibility problem at the moment, and it affects thousands of studies like this one.

Part of the angst stems from recent high-profile cases of scientific misconduct, most dramatically the extensive fraud perpetrated by Dutch social psychologist Diederik Stapel (*Science*, 4 November 2011, p. 579), that have cast a harsh light on psychological science. Yet there is no evidence that psychology is more prone to fraud than any other field of science. The greater concern arises from several recent studies that have broadly critiqued psychological research practices, highlighting lax data collection, analysis, and reporting, and decrying a scientific culture that too heavily favors new and counterintuitive ideas over the confirmation of existing results. Some psychology researchers argue that this has led to too many findings that are striking for their novelty and published in respected journals—but are nonetheless false.

As a step toward testing that disturbing idea, one project begun this year offers an online site (PsychFileDrawer.org) where psychologists can quickly and easily post, in brief form, the results of replications of experiments—whether they succeed or fail. University of California, San Diego, psychologist Hal Pashler, one of the project's developers, says the goal is to counteract the "file drawer problem" that plagues all of science, including psychology; researchers usually just file away straightforward replication studies because most journals decline to publish such work.

Double trouble? Brian Nosek leads a large-scale effort to replicate recent psychology studies

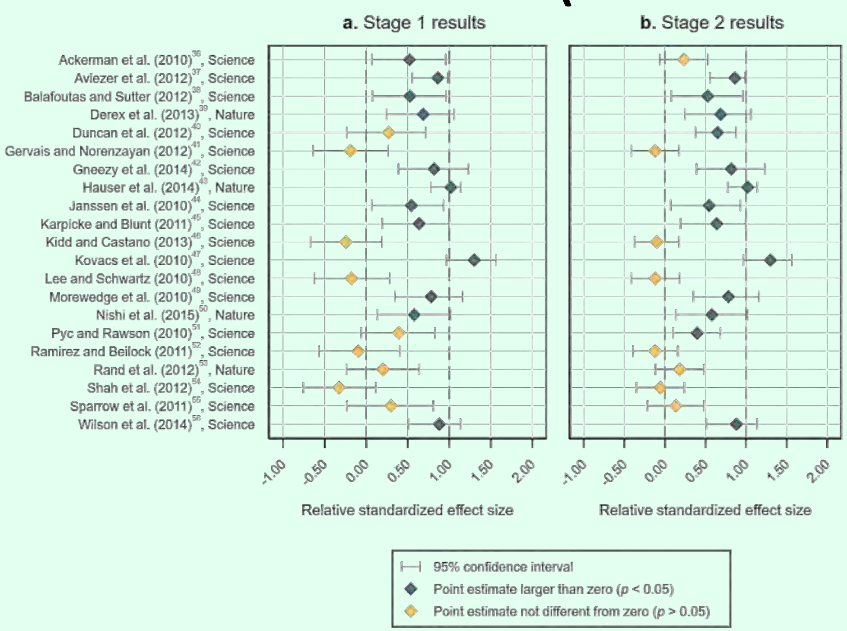
In an even more daring effort, a group of more than 50 academic psychologists, which calls itself the Open Science Collaboration (OSC), has begun an unprecedented, large-scale project to systematically replicate psychological experiments recently published in leading journals. "We're wringing our hands worrying about whether reproducibility is a problem or not," says psychologist Brian Nosek of the University of Virginia in Charlottesville, who is coordinating the effort. "If there is a problem, we're going to find out, and then we'll figure out how to fix it."

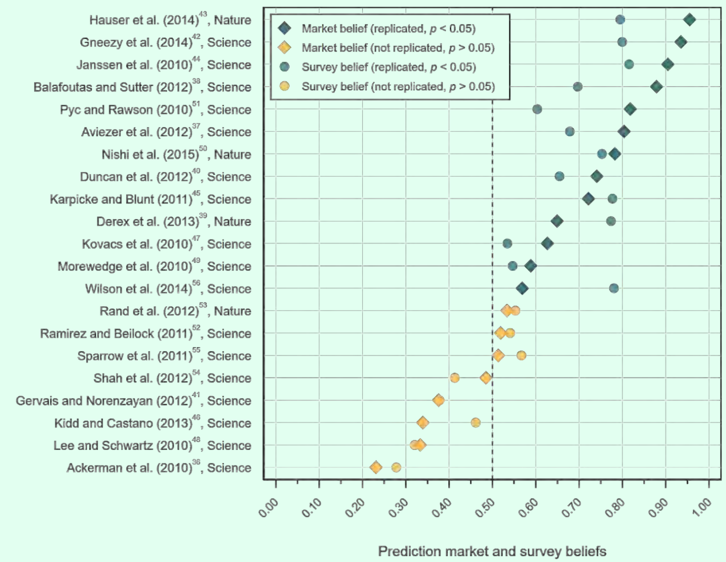
Robert Kail, a Purdue University developmental psychologist and editor of *Psychological Science*—one of the three journals whose papers the OSC is attempting to replicate—is optimistic that a high percentage of published findings will be replicated. Nonetheless, he views the field's recent attention to the issue of false positives as healthy. "There has been a lot of speculation about the extent to which it's a problem," he says. "But nobody has actually set it up as an empirical project. It's a great thing for somebody to actually do that."

Schooler, who is not directly involved with the project but whose free will study

Outcomes

- Other grant: Returned everything
- WAF: Continued with a new focus, prolonged and extended (Wallenberg Scholar)





Dreber et al. 2015 *PNAS*, Camerer et al. 2016 *Science*, Camerer et al. 2018 *Nature Human Behaviour,* Botvinik-Nezer et al. 2020 *Nature,* Huber et al. 2023 *PNAS,* Menkveld et al. 2024 *Journal of Finance,* Holzmeister et al. 2025 *Nature Human Behaviour,* Dreber et al. 2025 *PNAS,* etc.

What we learned from the new WAF projects

- There is something systematic about results that fail to replicate
- Importance of team science
- Pre-analysis plans one way forward
  - Confirmatory vs exploratory tests

- Pre-analysis plan shows one potential fork in the data
  - With meaningful p-values, but many forks possible
- Different researchers might choose different forks/pre-analysis plans
  - Should not result in systematic bias in effect sizes, but will underestimate the standard error in the statistical test
- <u>How large is the variation in results?</u>

Many analysts vs multiverse

- What is the natural variation in analyses and results in a given data set?
  - Can be explored with multi-analyst approach
    - Silberzahn et al. 2018 landmark paper
- What is the theoretically justified set of analyses and results in a given data set?
  - Can be explored with multiverse analyses, vibration of effects, specification curve analysis

Neuroimaging Analysis Replication and Prediction Study (NARPS)

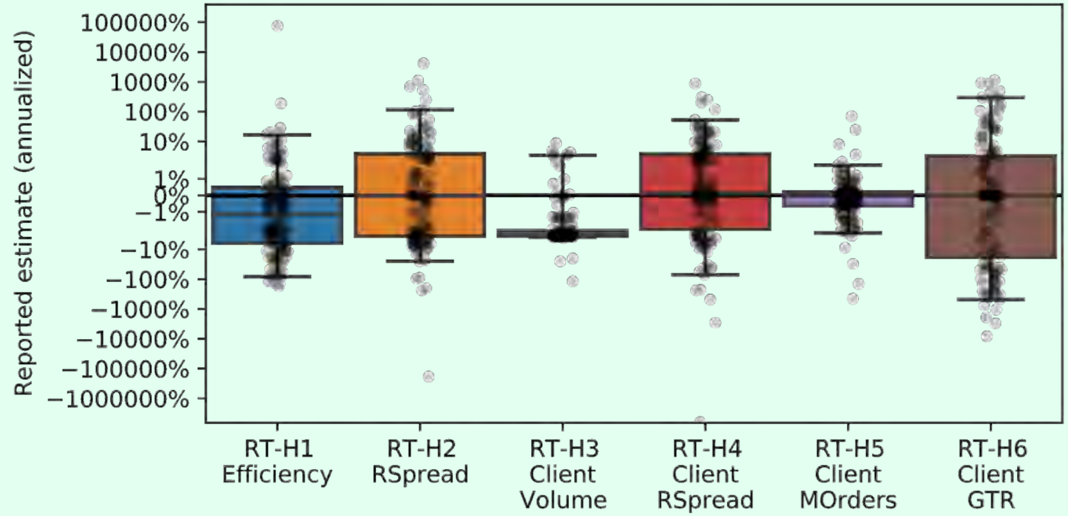
- Focus on variability of neuroimaging results across analysis teams
- We collected fMRI data from 108 participants performing two versions of a mixed gambles task
- 70 independent analysis teams received this dataset, including raw and preprocessed data. Asked to test 9 directional hypotheses on activations in specific contrasts and brain regions based on previous studies with this task
- Main outcome variable: fraction of teams reporting a significant result (yes/no), based on their own criteria, for each hypothesis

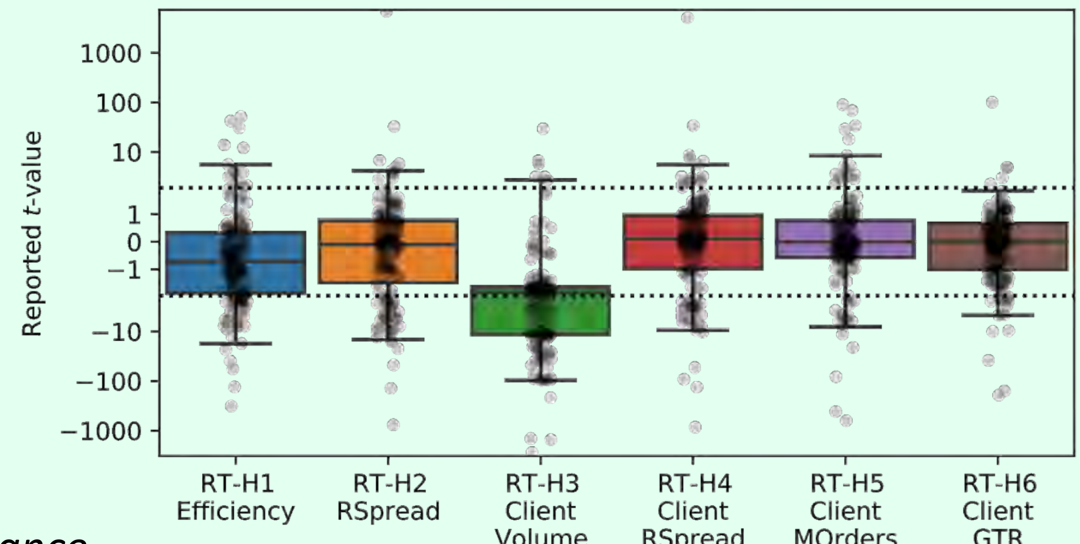
	Hypothesis description	**Fraction of teams reporting a significant result**	**Median confidence level**	**Median similarity estimation**
#1	Positive parametric effect of gains in the vmPFC (equal indifference group)	0.371	7 (2)	7 (1.5)
#2	Positive parametric effect of gains in the vmPFC (equal range group)	0.214	7 (1.5)	7 (1)
#3	Positive parametric effect of gains in the ventral striatum (equal indifference group)	0.229	6 (1)	7 (1)
#4	Positive parametric effect of gains in the ventral striatum (equal range group)	0.329	6 (1)	7 (1)
#5	Negative parametric effect of losses in the vmPFC (equal indifference group)	0.843	8 (1)	8 (1)
#6	Negative parametric effect of losses in the vmPFC (equal range group)	0.329	7 (1)	7 (1)
#7	Positive parametric effect of losses in the amygdala (equal indifference group)	0.057	7 (1)	8 (1)
#8	Positive parametric effect of losses in the amygdala (equal range group)	0.057	7 (1)	8 (1)
#9	Greater positive response to losses in amygdala for equal range group vs. equal indifference group	0.057	6 (1)	7 (1)

Botvinik-Nezer et al. *Nature* 2020

#fincap

- 164 research teams, 6 hypotheses, same big data
  - 720 million trade records over 17 years in EuroStoxx 50 index futures
- Review element/several stages



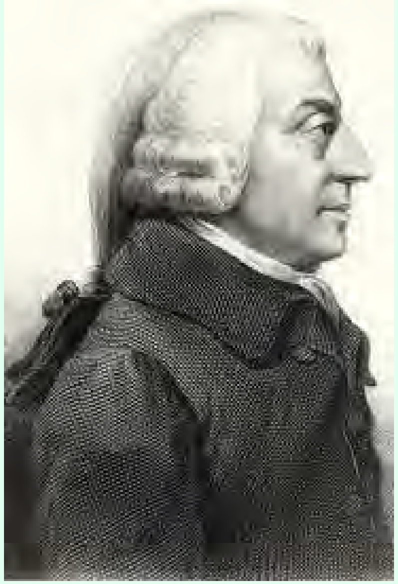


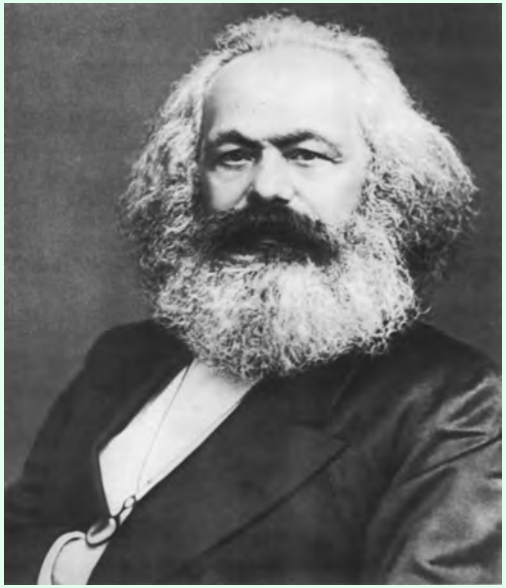
Menkveld et al. 2024 *Journal of Finance*

The boxes span the the first to the third quartile with the median as the interior horizontal line. The whiskers span 95% of the observations, starting from the 2.5% quantile to the 97.5% quantile.

Uncertainty from design heterogeneity

- Same hypothesis, different designs, different researchers, same analysis
- Does competition affect moral behavior?
- 43 research teams
- n>18k

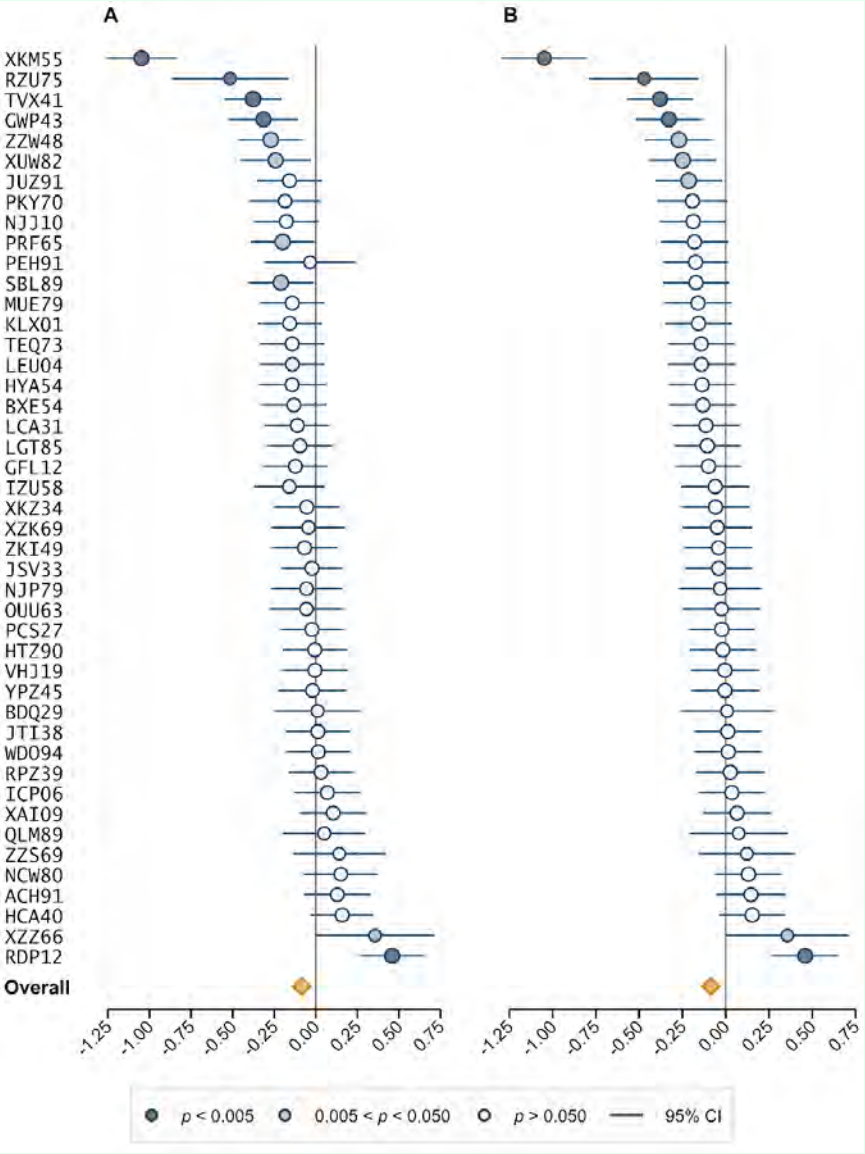




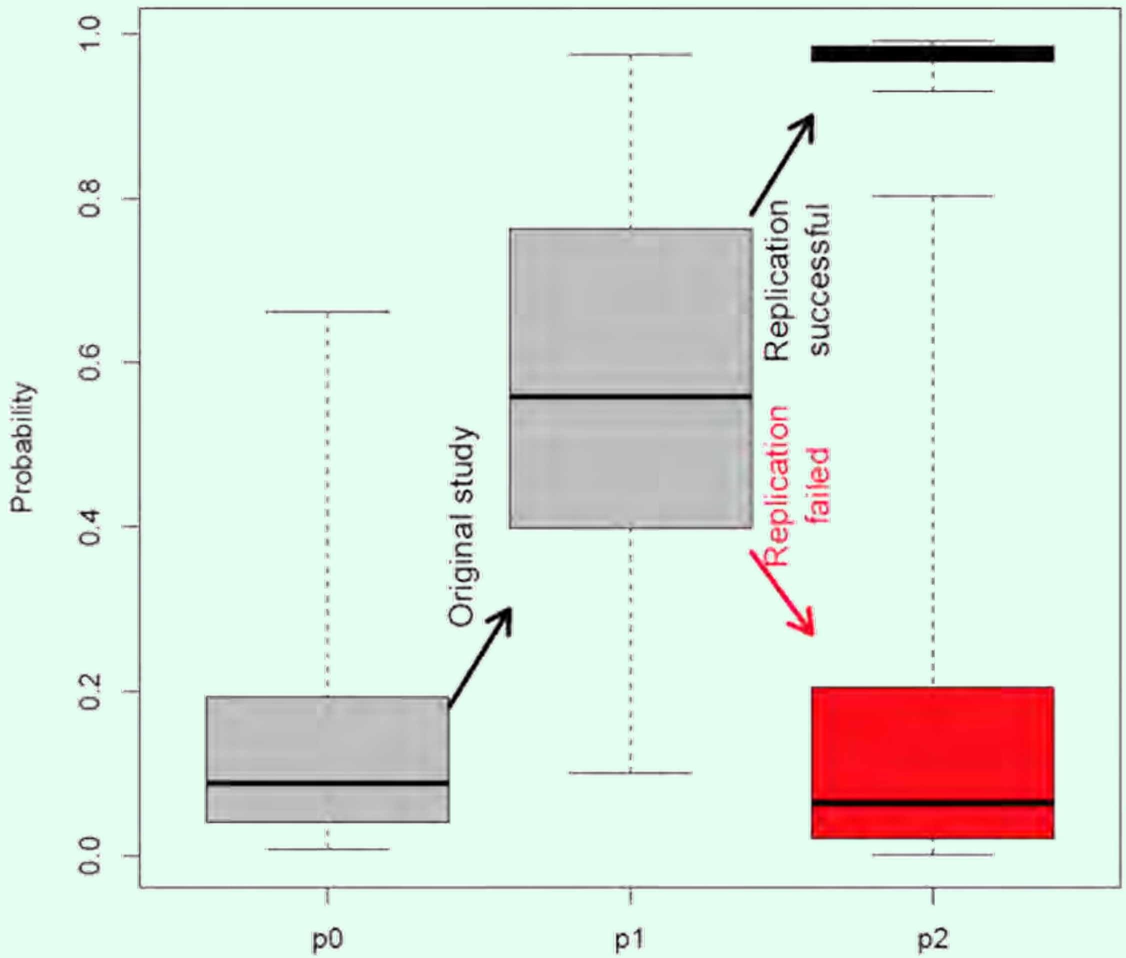
Huber et al. 2023 *PNAS*

Design heterogeneity

- Substantial design heterogeneity: 1.6 times as large as the average standard error of effect size estimates



Learning from replications



Whiskers: range
Boxes: 1st to 3rd quartiles
Thick lines: medians

Dreber et al. 2015 *PNAS*

Discussion

- In my case – what would have happened without the WAF grant?
- Short term vs long term funding – when is what ideal?

Discussion

- Team science for generalizability

- High risk high reward?
- Learning from ”failures”
  - Need for more data?

- Experimenting

- How to encourage and evaluate team science?

Thanks!

anna.dreber@hhs.se