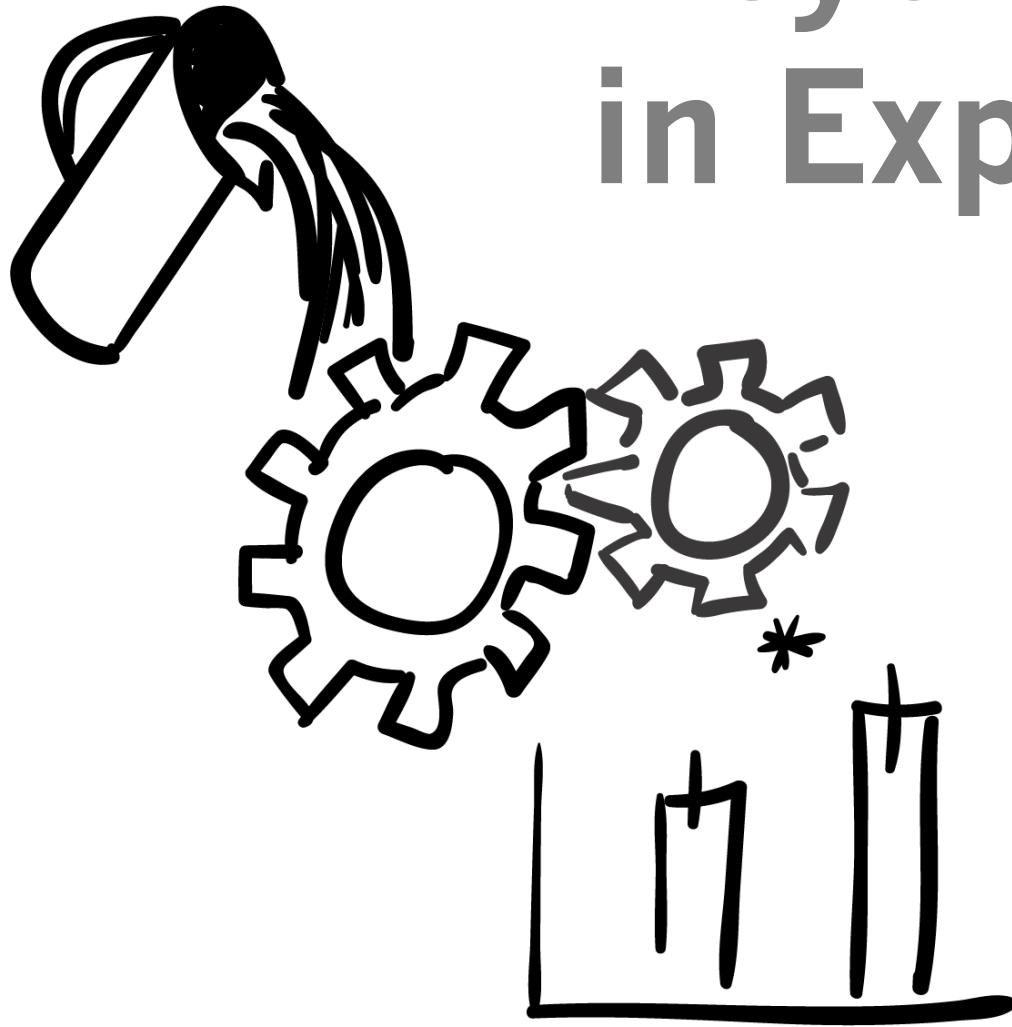


# Psych 251: Lab in Experimental Methods



# Introductions

- Mike (mcfrank@)
  - Developmental and cognitive groups
  - Word learning, social cognition, how we learn from other people, how language works
  - Computational and experimental methods, eye-tracking, web-based studies, “big data” (really, medium-sized data)

# Introductions

- Ian Eisenberg (ieisenbe@)
  - Fourth year in PoldrackLab
  - Understanding the large scale structure of cognitive systems and their relationship to neural entities. Individual differences in self-regulation.
  - Loves animals and has two cats and a snake named Amy. Has killed many fish.
- Catherine Thomas (ccthomas@)
  - Second year in Social Area
  - Leveraging culturally-relevant motivations in messaging around cash-based welfare both to increase policy support among the general public and to improve outcomes among low-income recipients in Kenya and in the US.
  - Visited about 40 countries and lived (6+ months) in 4.

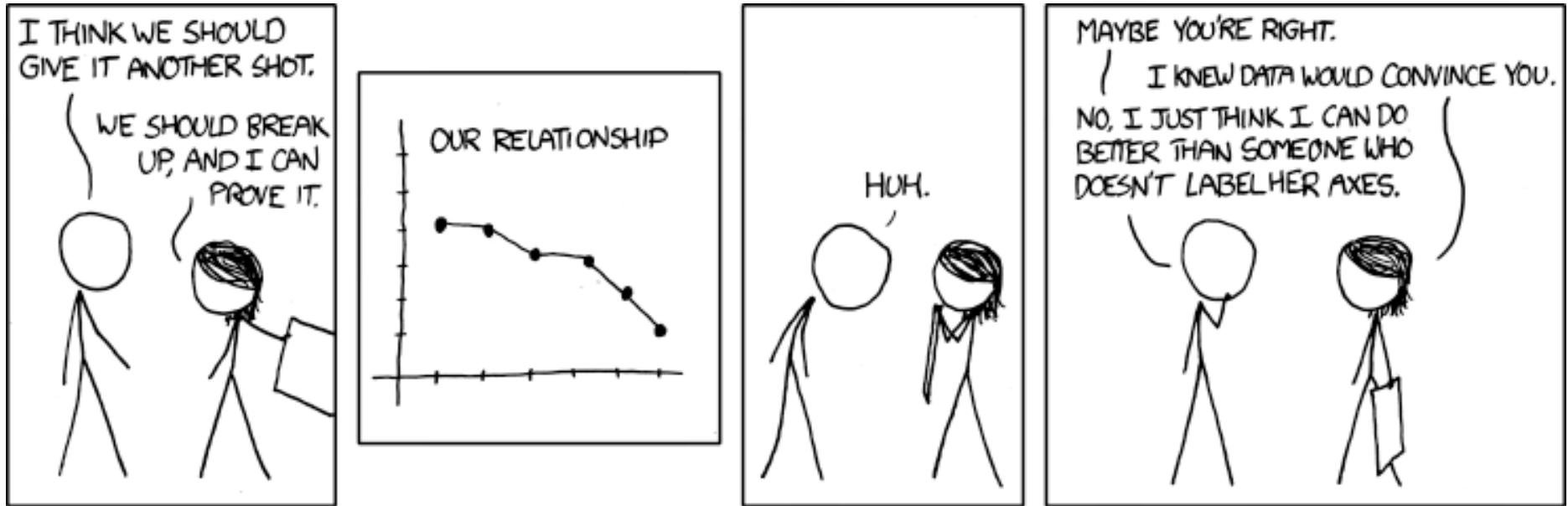


# What this course is about



Why we should care about quantitative standards?

# What this course is about



Basically, all the stuff that you are supposed to know but aren't taught in courses on psychological “content”

# Brief history

- 2012 – 2017 this course was psych 254
  - Winter time slot
  - Grad statistics as a prerequisite
  - Substantial R prerequisites
- This AY:
  - Fall timeslot
  - First course for incoming Psych PhDs
  - No stats prereq (psych 10 useful)
  - Basic R or other statistical programming useful

# Motivation of the course

- I am sloppy and error-prone
  - Need to think about how to do things efficiently and correctly
  - Or else I will just do them wrong
  - More on this later in the course
- So I think a lot about workflow...
- But I've come to recognize that sloppiness and error are field-wide risks for bad science
- The story of my “methods radicalization”



**Don't make it  
broken by  
design!**

**Reproducible** = Can get the same results from the same dataset

**Replicable** = Can get the same\* findings in a new dataset

Both are critical prerequisites of a cumulative science

Initial rumblings that  
something isn't right...

# Questionable research practices

General Article

---

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

**Joseph P. Simmons<sup>1</sup>, Leif D. Nelson<sup>2</sup>, and Uri Simonsohn<sup>1</sup>**

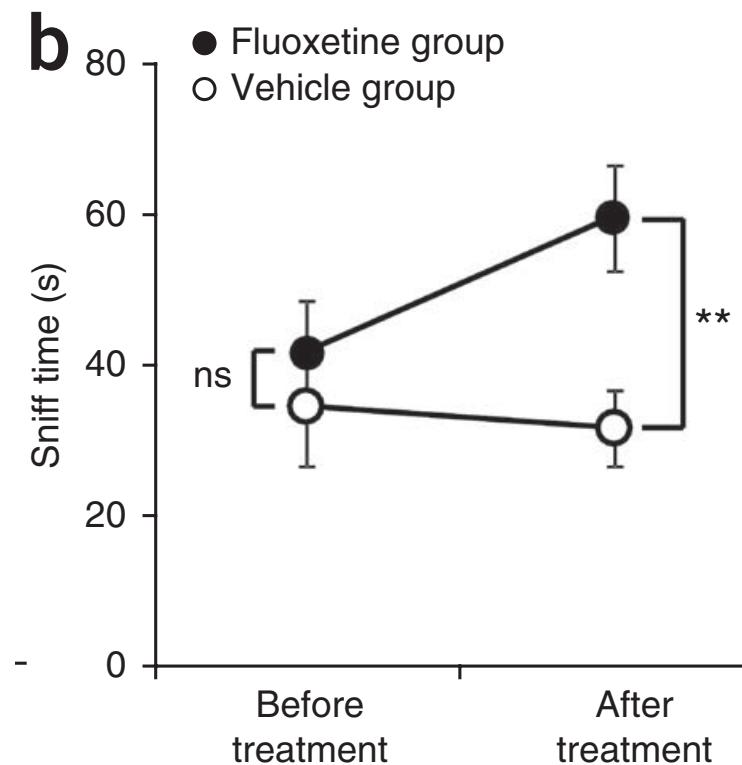
<sup>1</sup>The Wharton School, University of Pennsylvania, and <sup>2</sup>Haas School of Business, University of California, Berkeley

**Table 1.** Likelihood of Obtaining a False-Positive Result

Researcher degrees of freedom	Significance level		
	p < .1	p < .05	p < .01
Situation A: two dependent variables ( $r = .50$ )	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

# Erroneous analyses of interactions in neuroscience: a problem of significance

Sander Nieuwenhuis<sup>1,2</sup>, Birte U Forstmann<sup>3</sup> & Eric-Jan Wagenmakers<sup>3</sup>



Gelman et al.:  
The difference between  
significant and not  
significant is not itself  
statistically significant.

**Table 1** Outcome of the main literature analysis

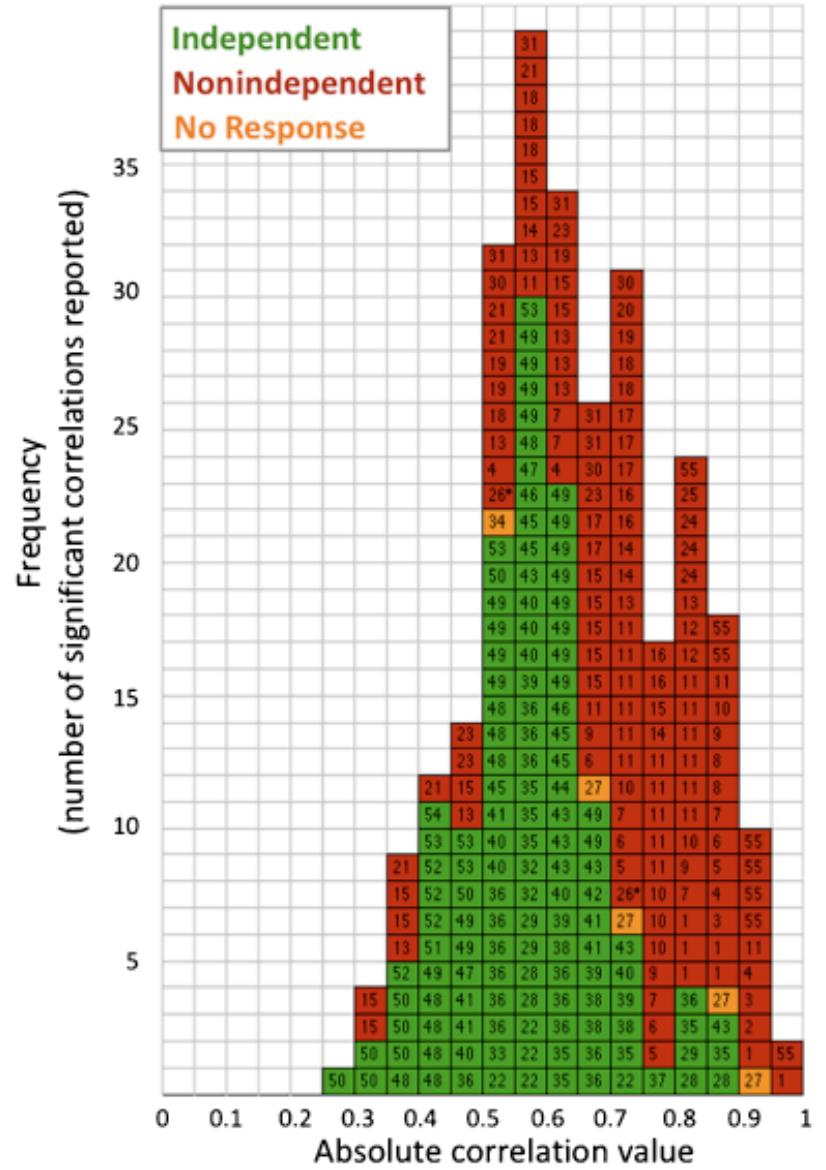
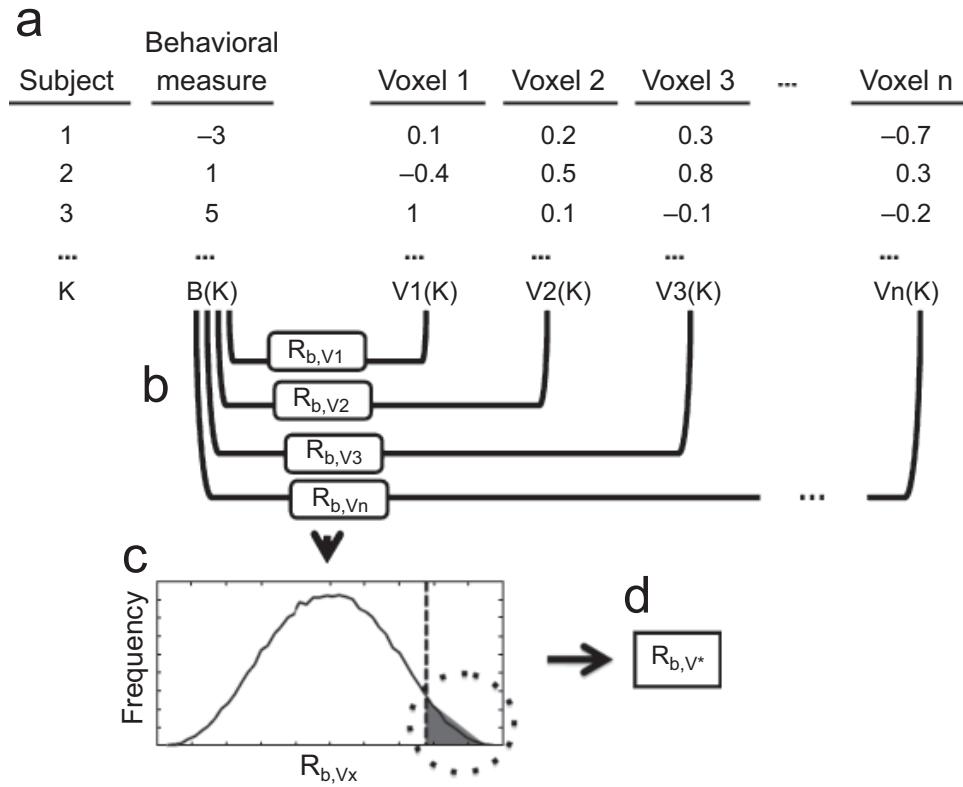
	<i>Nature</i>	<i>Science</i>	<i>Nature Neuroscience</i>	<i>Neuron</i>	<i>Journal of Neuroscience</i>	Summed
Total reviewed	34	45	117	106	211	513
Correct count	3	9	17	13	36	78
Error count	7	11	16	15	30	79

For this analysis, we included every article of which the abstract referred to behavior, cognitive function or brain imaging.

# Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition<sup>1</sup>

Edward Vul,<sup>1</sup> Christine Harris,<sup>2</sup> Piotr Winkielman,<sup>2</sup> & Harold Pashler<sup>2</sup>

<sup>1</sup>Massachusetts Institute of Technology and <sup>2</sup>University of California, San Diego





## Harvard Dean Confirms Misconduct in Hauser Investigation

by [Greg Miller](#) on 20 August 2010, 3:11 PM | [2 Comments](#)

[Email](#)  [Print](#) |  [f](#)  [t](#)  [+1](#)  [0](#)

[r](#)  [SU](#)  [+ More](#)

[PREVIOUS ARTICLE](#)

[NEXT ARTICLE](#)

In an e-mail sent earlier today to Harvard University faculty members, Michael Smith, dean of the Faculty of Arts and Sciences (FAS), confirmed that he had directed the university's Office of the Provost to conduct an investigation into the matter.

November 13, 2011

## Fraud Scandal Fuels Debate Over Practices of Social Psychology

Even legitimate researchers cut corners, some admit



By Christopher Shea

The discovery that the Dutch researcher Diederik A. Stapel made up the data for dozens of research papers has shaken up the field of social psychology, fueling a discussion not just about outright fraud, but also about subtler ways of misusing research data. Such misuse

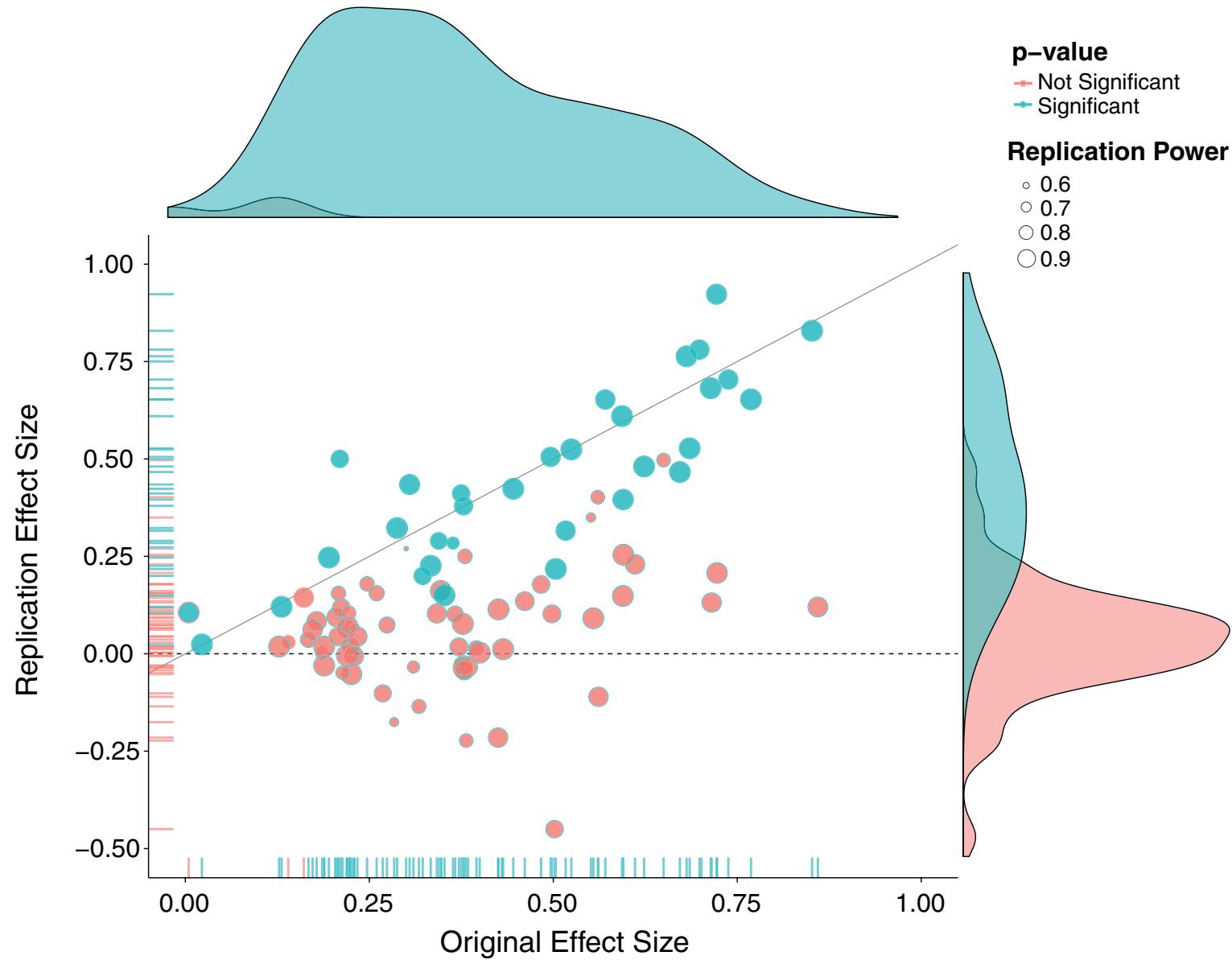
# Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem  
Cornell University

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of psi are *precognition* (conscious cognitive awareness) and *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or nonconscious, cognitive or affective. This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by "time-reversing" well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur. Data are presented for 4 time-reversed effects: precognitive approach to erotic stimuli and precognitive avoidance of negative stimuli; retroactive priming; retroactive habituation; and retroactive facilitation of recall. The mean effect size ( $d$ ) in psi performance across all 9 experiments was 0.22, and all but one of the experiments yielded statistically significant results. The individual-difference variable of stimulus seeking, a component of extraversion, was significantly correlated with psi performance in 5 of the experiments, with participants who scored above the midpoint on a scale of stimulus seeking achieving a mean effect size of 0.43. Skepticism about psi, issues of replication, and theories of psi are also discussed.

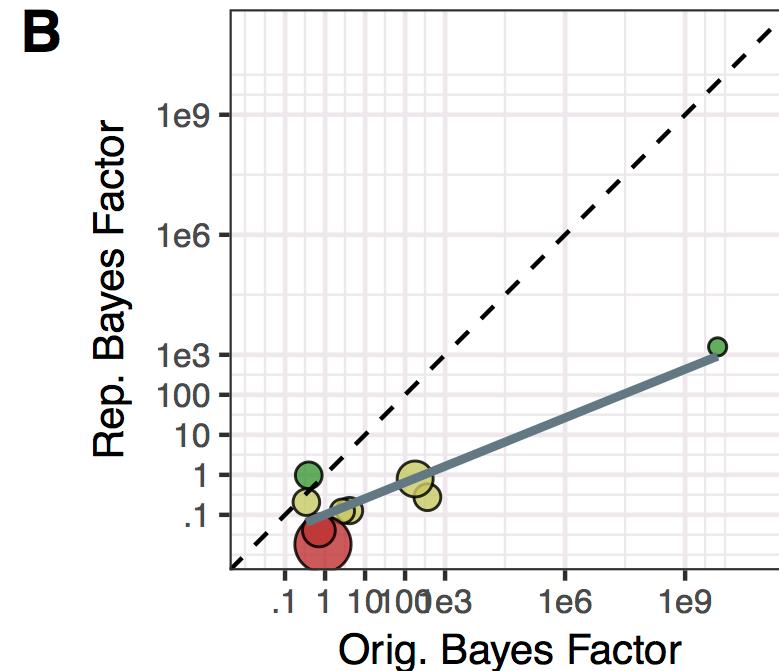
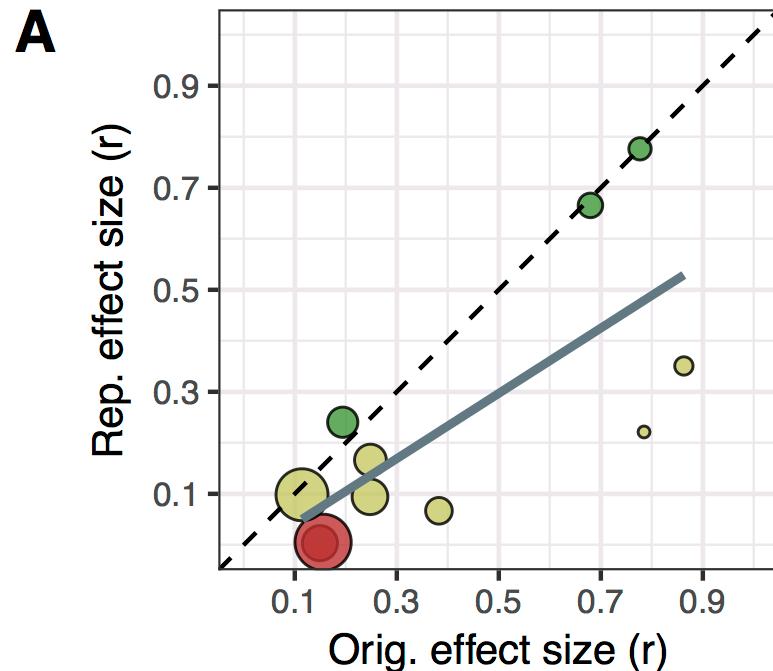
*Keywords:* psi, parapsychology, ESP, precognition, retrocausation

Ok, but *everything* isn't that  
bad, right?



Open Science Framework (2015), inc. 4 psych254 projects

# 11 studies replicated, sampled from Psych Science 2015



Rep. N ○ 100 ○ 200 ○ 300

Rep. df/Orig. df ○ 0.5 ○ 2.0 ○ 5.0

Subjective replication rating ● No ● Partial ● Yes

# The (mis)reporting of statistical results in psychology journals

Marjan Bakker · Jelte M. Wicherts

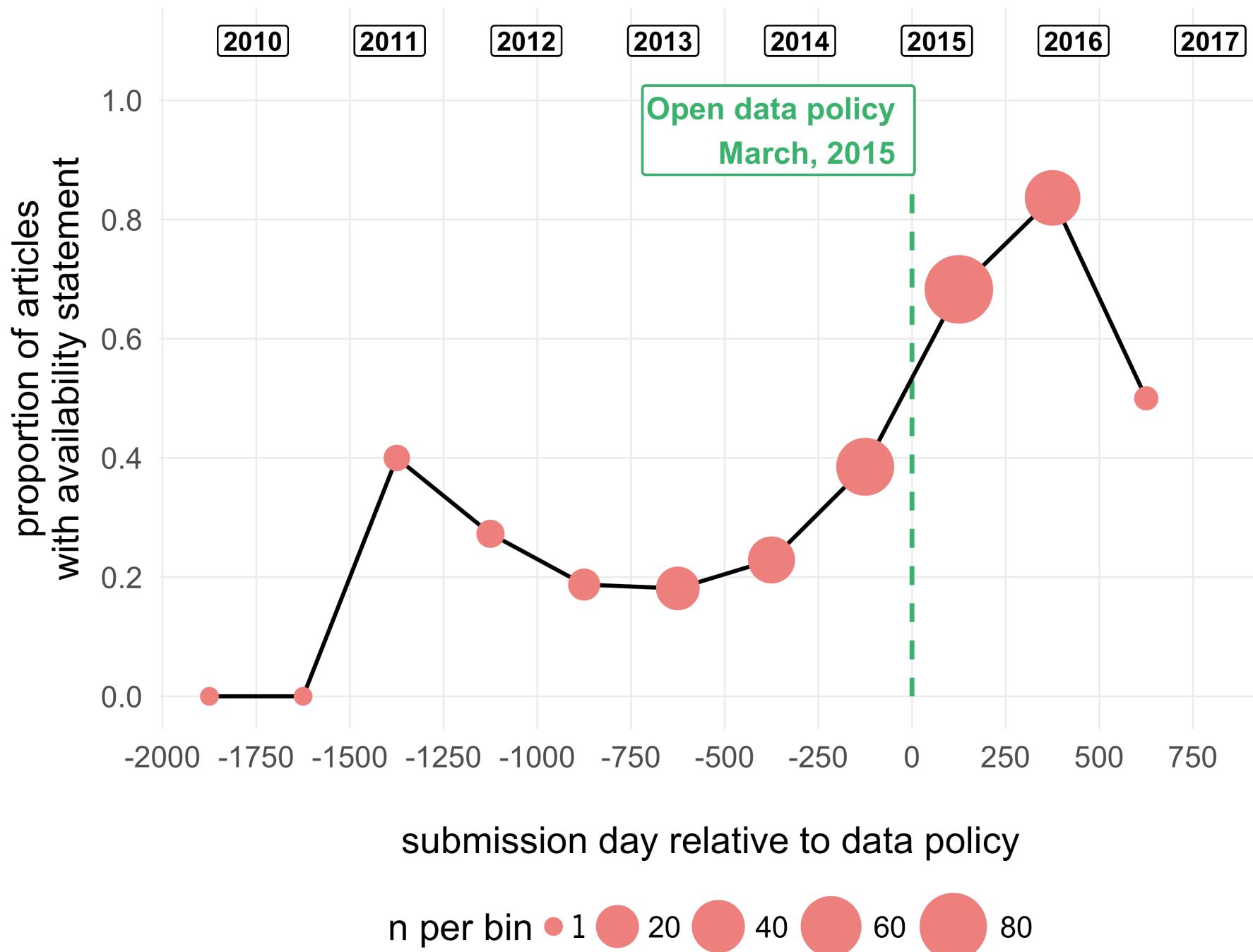
**Table 5** Number of statistics, errors, and gross errors per significance category in high- and low-impact journals for exactly reported statistical results

		No. Statistics	No. Errors	No. Gross Errors
High	$p \leq .05$	533	89 (16.7%)	10 (1.9%)
	$p > .05$	428	55 (12.9%)	1 (0.2%)
Low	$p \leq .05$	113	33 (29.2%)	5 (4.4%)
	$p > .05$	94	23 (24.5%)	2 (2.1%)
Total	$p \leq .05$	646	122 (18.9%)	15 (2.3%)
	$p > .05$	522	78 (14.9%)	3 (0.6%)

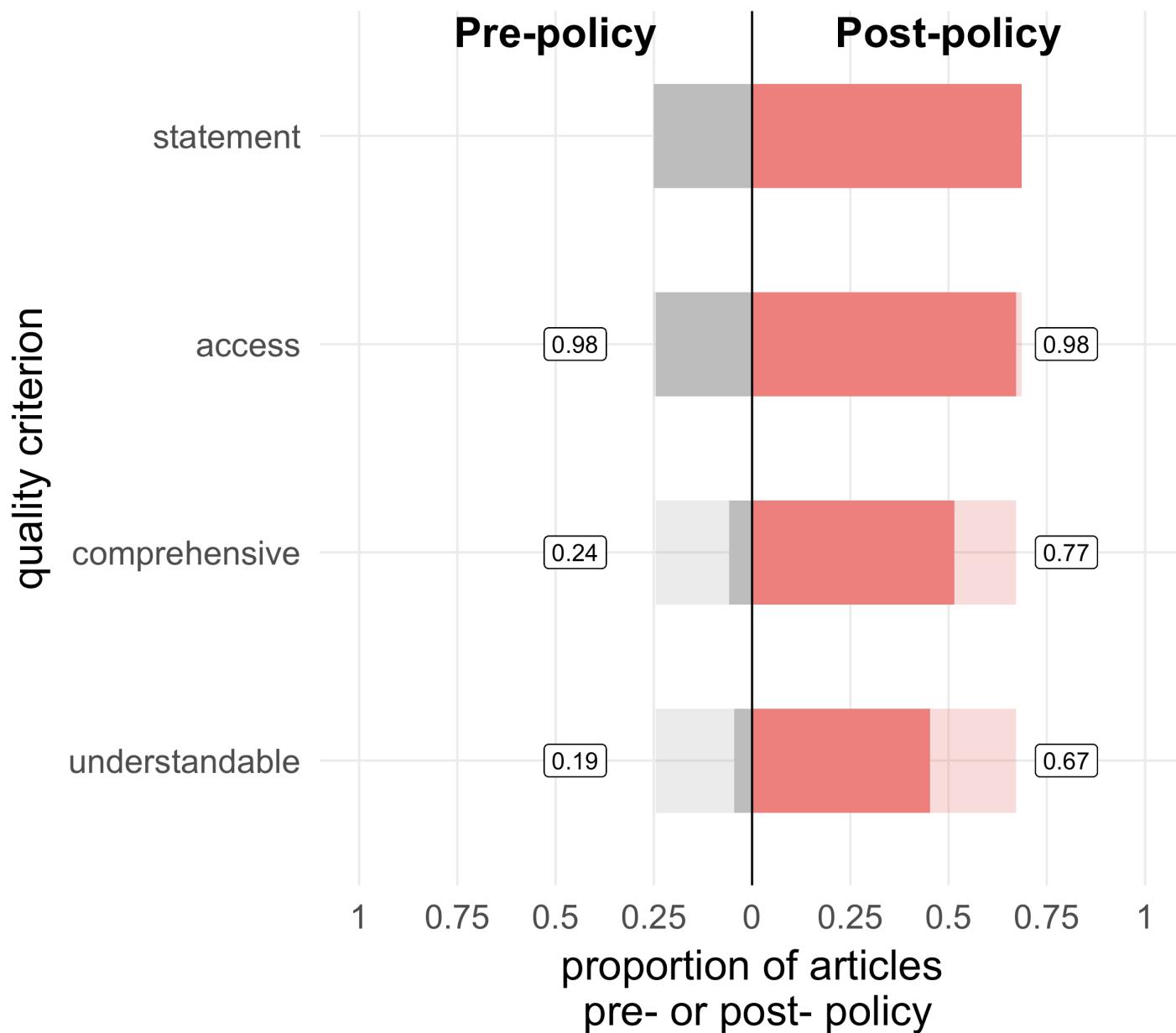
# A study of reproducibility



- Mandatory open data policy was introduced at the journal Cognition on 1<sup>st</sup> March, 2015.
- **Aim 1:** Quantify data availability pre- and post- policy
- **Aim 2:** Assess whether shared data are actually high quality (accessible, comprehensive, understandable).
- **Aim 3:** Establish whether data sharing enables computational reproducibility by re-running the reported analyses and trying to obtain the reported outcomes.
  - **Note: this was a problem set last year! We'll do a similar one this year!**



# Data availability



# Reproducibility

In progress: 12

Success without author assistance (10) – 43%	Success with author assistance (7) – 30%
Failure without author assistance (0)	Failure despite author assistance (6) – 26%

- >> Mandatory data policies are mandatory in name only without consistent editorial oversight.
- >> Open data alone is unlikely to be sufficient to enable data re-use.

So what do we do?

# Ways forward (this course)

1. Transparent and reproducible workflow
2. Better understanding of statistical inference and role of preregistration
3. Decreasing the costs of replication
4. Increasing sample sizes

# Course goals

- Master best practices for experimental data management, storage, and analysis, and
- Be able to analyze a study in terms of reliability, validity, and statistical inference choices,
- Be able to perform a psychology experiment on the web,
- Have a reproducible workflow for experimental data analysis and visualization, including comfort with a variety of R packages in the “tidyverse.”

<http://psych251.github.io>

and

[https://mailman.stanford.edu/mailman/listinfo/  
/psych251-aut1718](https://mailman.stanford.edu/mailman/listinfo/psych251-aut1718)

ROB!  
YOU USE UNIX!

COME QUICK!



TO DISARM THE BOMB,  
SIMPLY ENTER A VALID  
tar COMMAND ON YOUR  
FIRST TRY. NO GOGLING.  
YOU HAVE **TEN** SECONDS.

~# -



...ROB?

I'M SO SORRY.



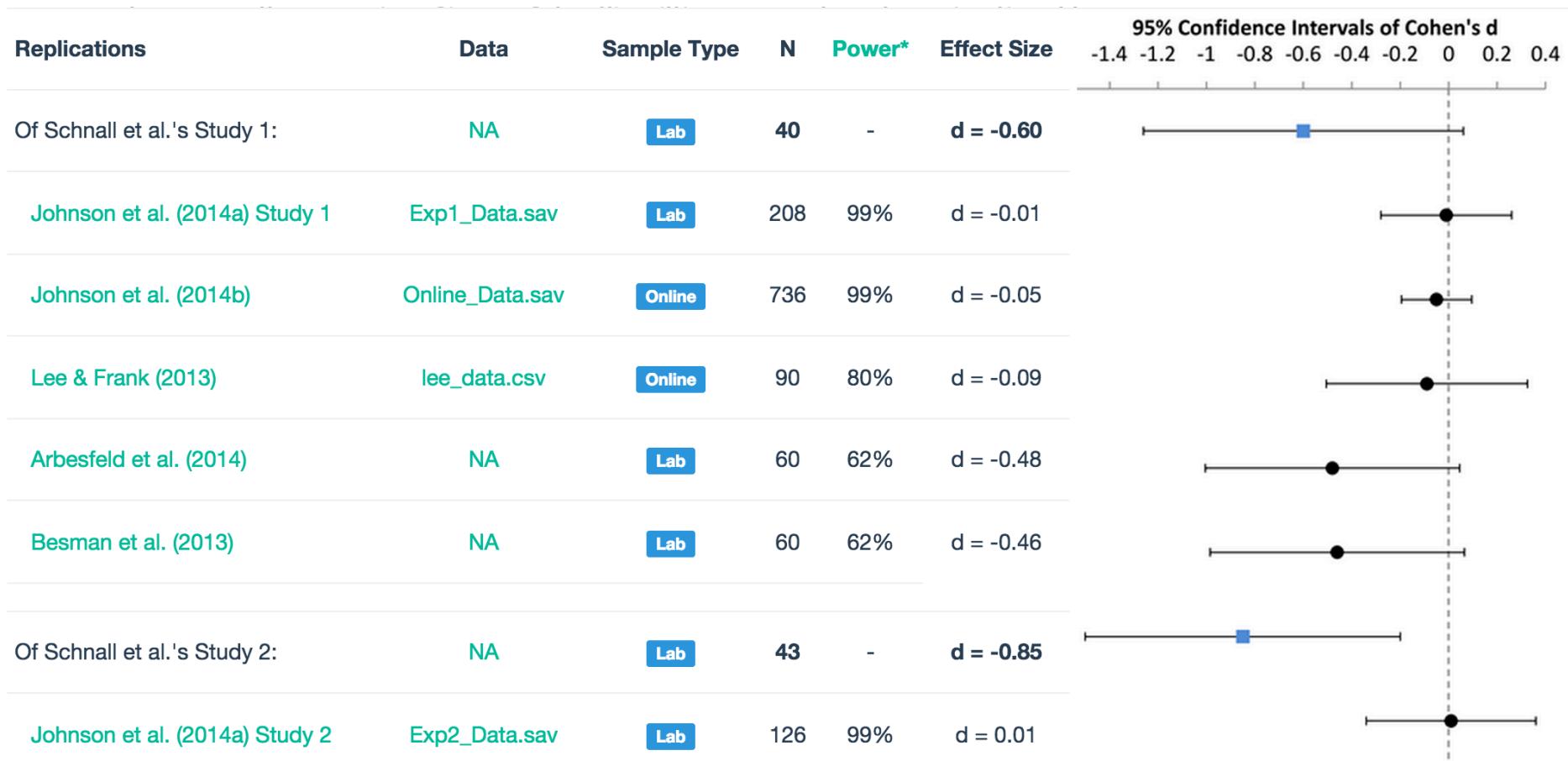
Learn to learn to do stuff with the computer...

# Your project

- Pick a study
- Replicate it (often but not always on Mechanical Turk)
- Write up a description
  - Via the Open Science Collaboration framework format
- Consider what to do next
  - Class meta-science project?
  - Submit to a registry?
  - Follow up for fun and profit?

# Another replication of Schnall, Benton, & Harvey (2008)

Simone Schnall, in [her recent blogpost](#), notes that she has received many requests for materials and data to investigate her work on cleanliness priming. One of those requests came from [Fiona Lee](#), a student in my replication-based graduate research methods course (info and syllabus [here](#)). Fiona did a project conducting a replication of Study 1 from Schnall, Benton, & Harvey (2008) using Amazon Mechanical Turk. We'd like to say at the



# A Second Look at Automatic Theory of Mind: Reconsidering Kovács, Téglás, and Endress (2010)



**Jonathan Phillips<sup>1,2</sup>, Desmond C. Ong<sup>3</sup>, Andrew D. R. Surtees<sup>4</sup>, Yijing Xin<sup>5</sup>, Samantha Williams<sup>5</sup>, Rebecca Saxe<sup>5</sup>, and Michael C. Frank<sup>3</sup>**

<sup>1</sup>Department of Psychology, Yale University; <sup>2</sup>Department of Philosophy, Yale University; <sup>3</sup>Department of Psychology, Stanford University; <sup>4</sup>Department of Psychology, University of Birmingham; and <sup>5</sup>Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

## Abstract

In recent work, Kovács, Téglás, and Endress (2010) argued that human adults automatically represented other agents' beliefs even when those beliefs were completely irrelevant to the task being performed. In a series of 13 experiments, we replicated these previous findings but demonstrated that the effects found arose from artifacts in the experimental paradigm. In particular, the critical findings demonstrating automatic belief computation were driven by inconsistencies in the timing of an attention check, and thus do not provide evidence for automatic theory of mind in adults.

2015, Psych Science

Psychological Science  
1–15  
© The Author(s) 2015  
Reprints and permissions:  
[sagepub.com/journalsPermissions.nav](http://sagepub.com/journalsPermissions.nav)  
DOI: 10.1177/0956797614558717  
[pss.sagepub.com](http://pss.sagepub.com)  
SAGE

# Wednesday - RMarkdown

Reproducible analysis workflow: how to create and share code, data, and presentation easily

## Berkowitz et al. (2015) reanalysis

*Michael C. Frank*

2016-01-04

- 1 Introduction
- 2 Exclusions and Data Preparation
- 3 Technical Comment Analyses
- 4 Acknowledgements

## 1 Introduction

This document reports code for my Technical Comment on Berkowitz et al. (2015). For a fuller analysis see this repository (<https://github.com/mcfrank/berkowitz>).

```
d <- read.csv("aac7427-Accessory-Data-File-S1.csv") %>%
  mutate(condition = ifelse(cond.dum == 1, "Math App", "Reading App"))
```

## 2 Exclusions and Data Preparation

One of the tricky things in this dataset is figuring out how to make the exclusions line up with the paper.

We need to exclude:

- twins (twin)
- dropouts during the first year (year1dropout)

We also note that we need to include 31 children due to experimenter errors (0.1 fall, 1.6 math and 0

# Friday – version control!

