

Scan to read a summary!

<https://osf.io/preprints/metaarxiv/3hb6g>



Advancing Openness and Transparency in Scientific Contributions

Rima-Maria Rahal^{1,2} & Susann Fiedler^{2,3}

¹ Tilburg University

² Max Planck Institute for Research on Collective Goods

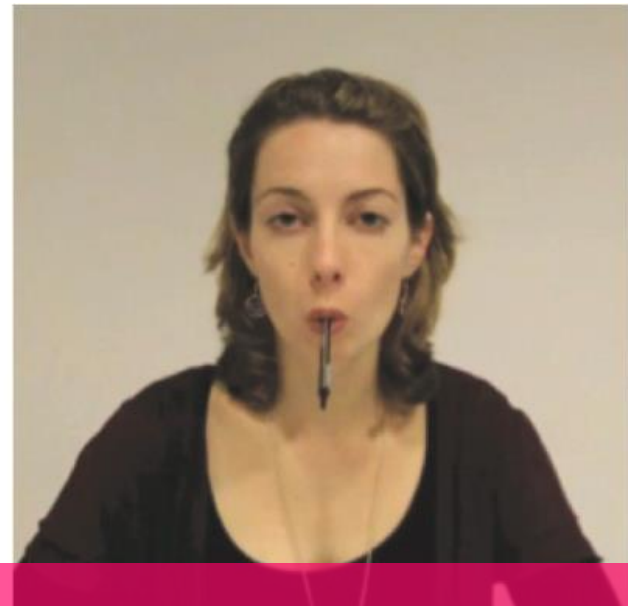
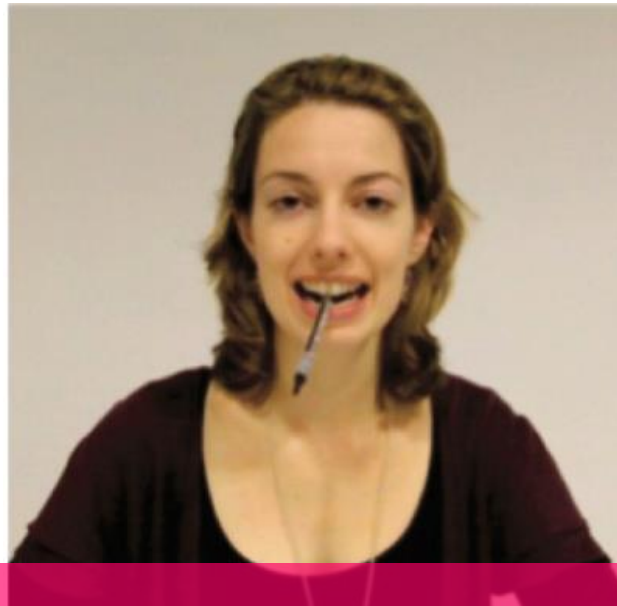
³ Open Science Framework

Pretty different...



Do we have anything in common?



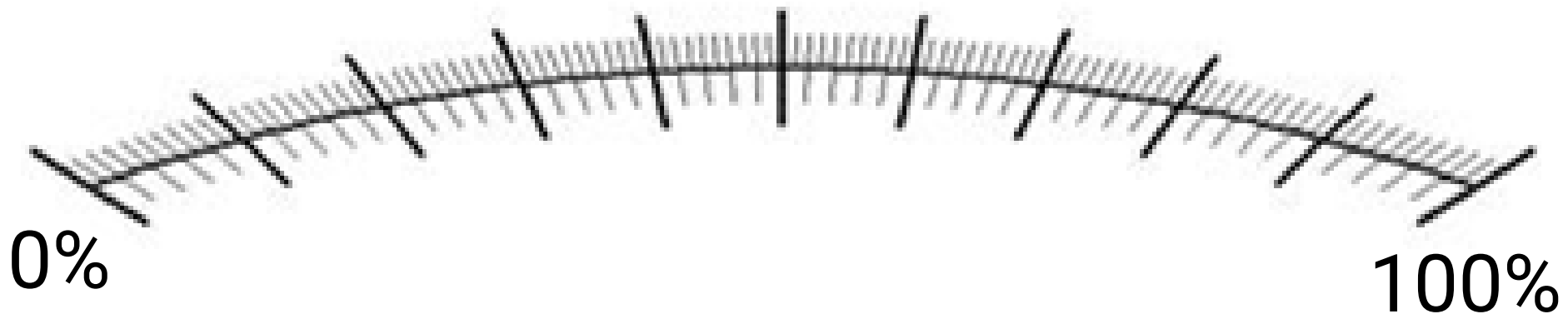


Tension between Mission and Practice.



Seizing up the problem.

Estimating the replicability of scientific research.



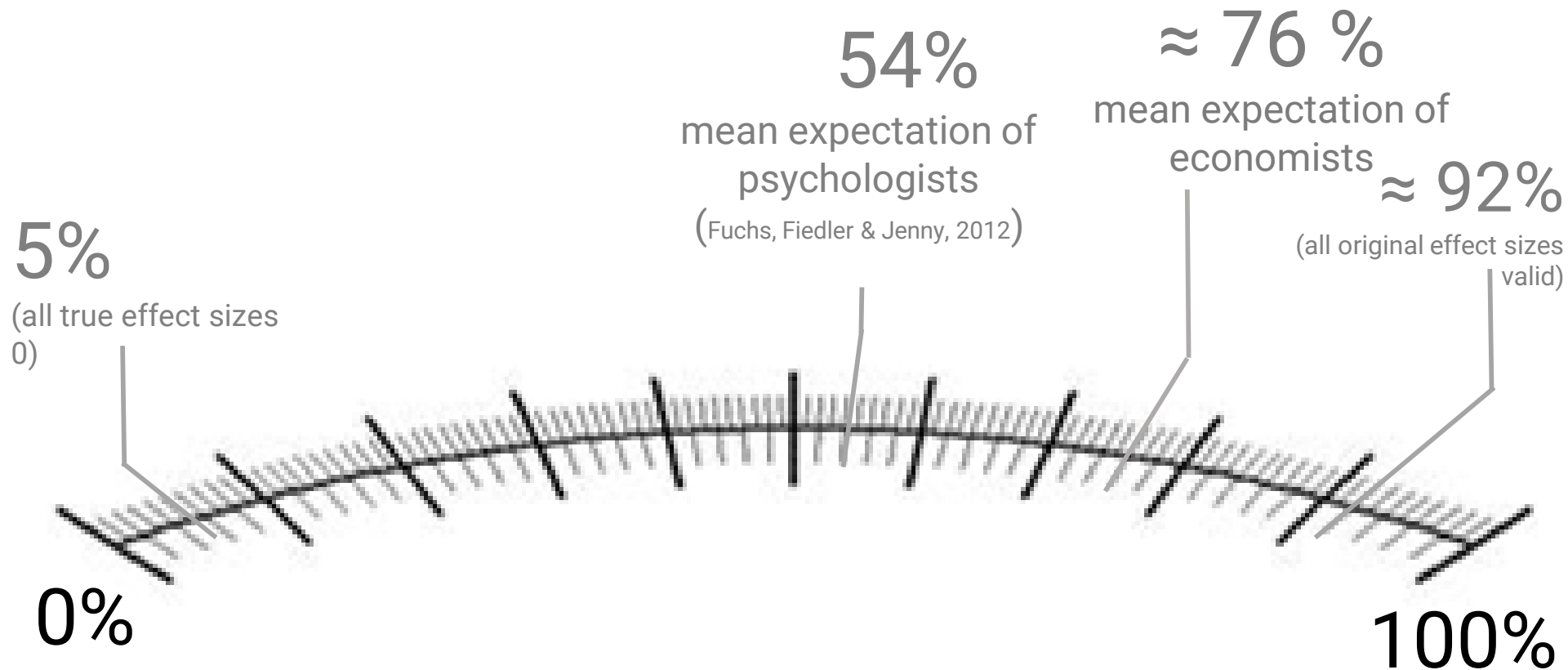
Predictions of Replication.



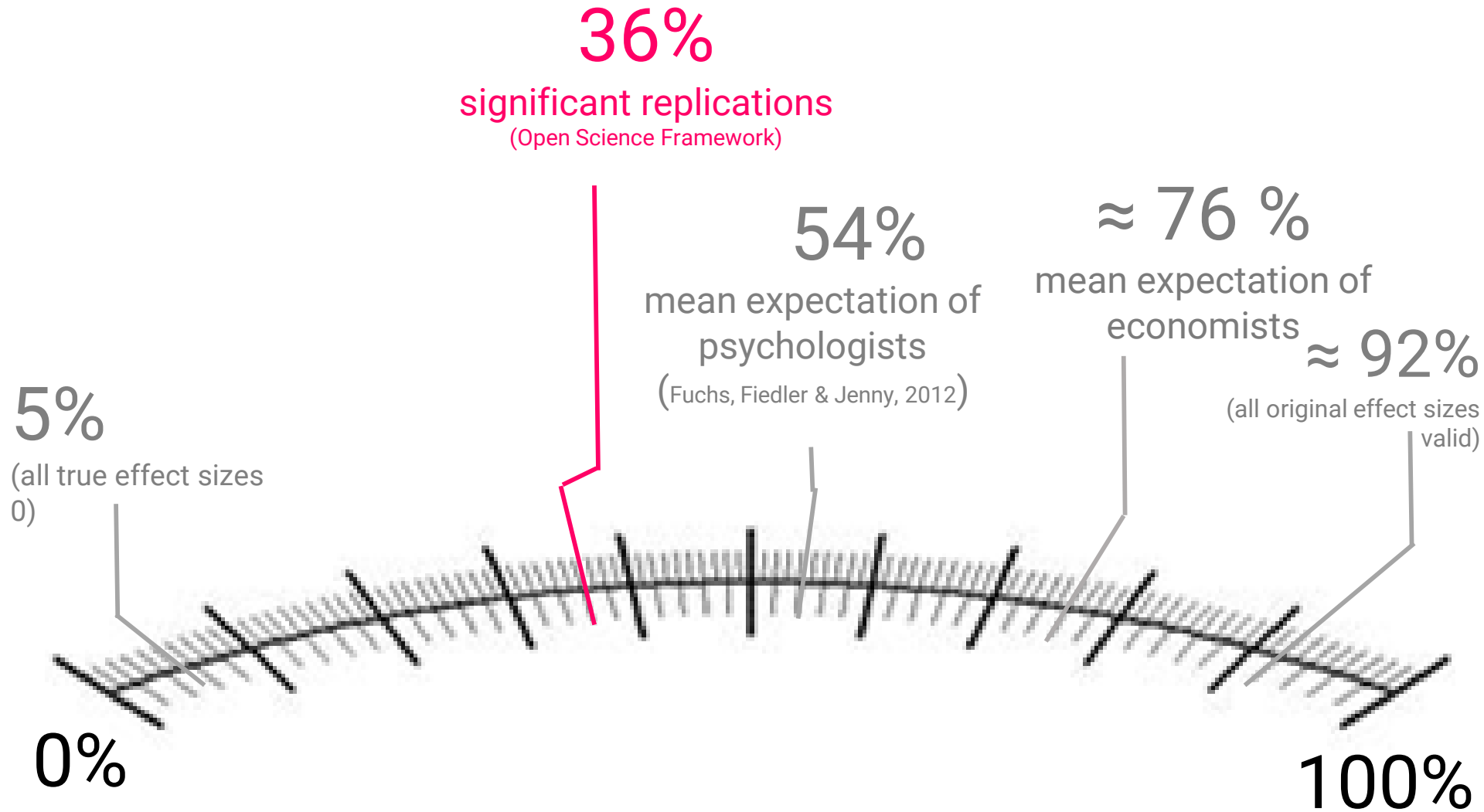
Predictions of Replication.



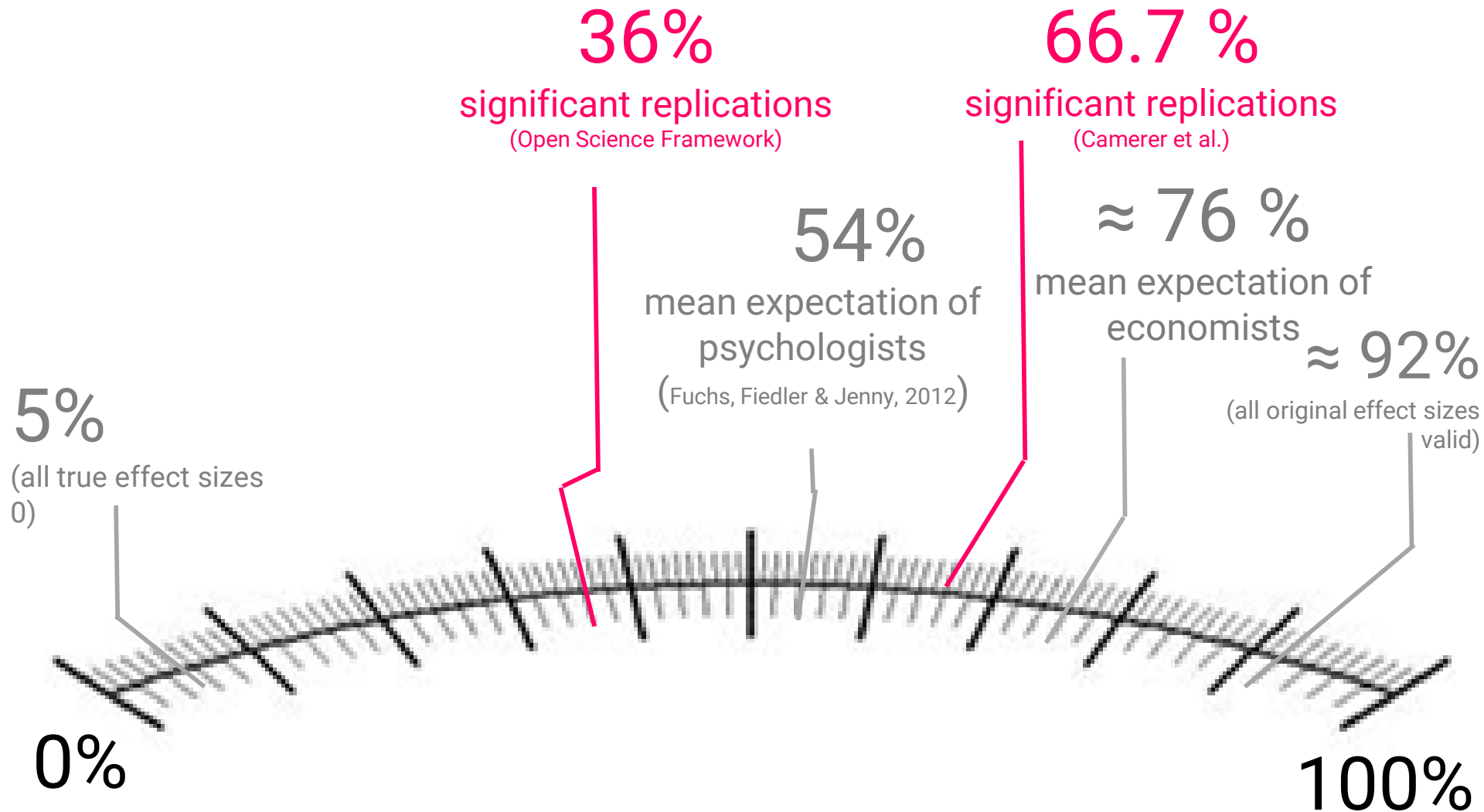
Predictions of Replication.



Predictions of Replication.



Predictions of Replication.



Is reproducibility a concern only
for psych & behavioral econ?



Cancer Biology Reproducibility Project Sees Mixed Results

By Courtney Humphries on Wed, 18 Jan 2017

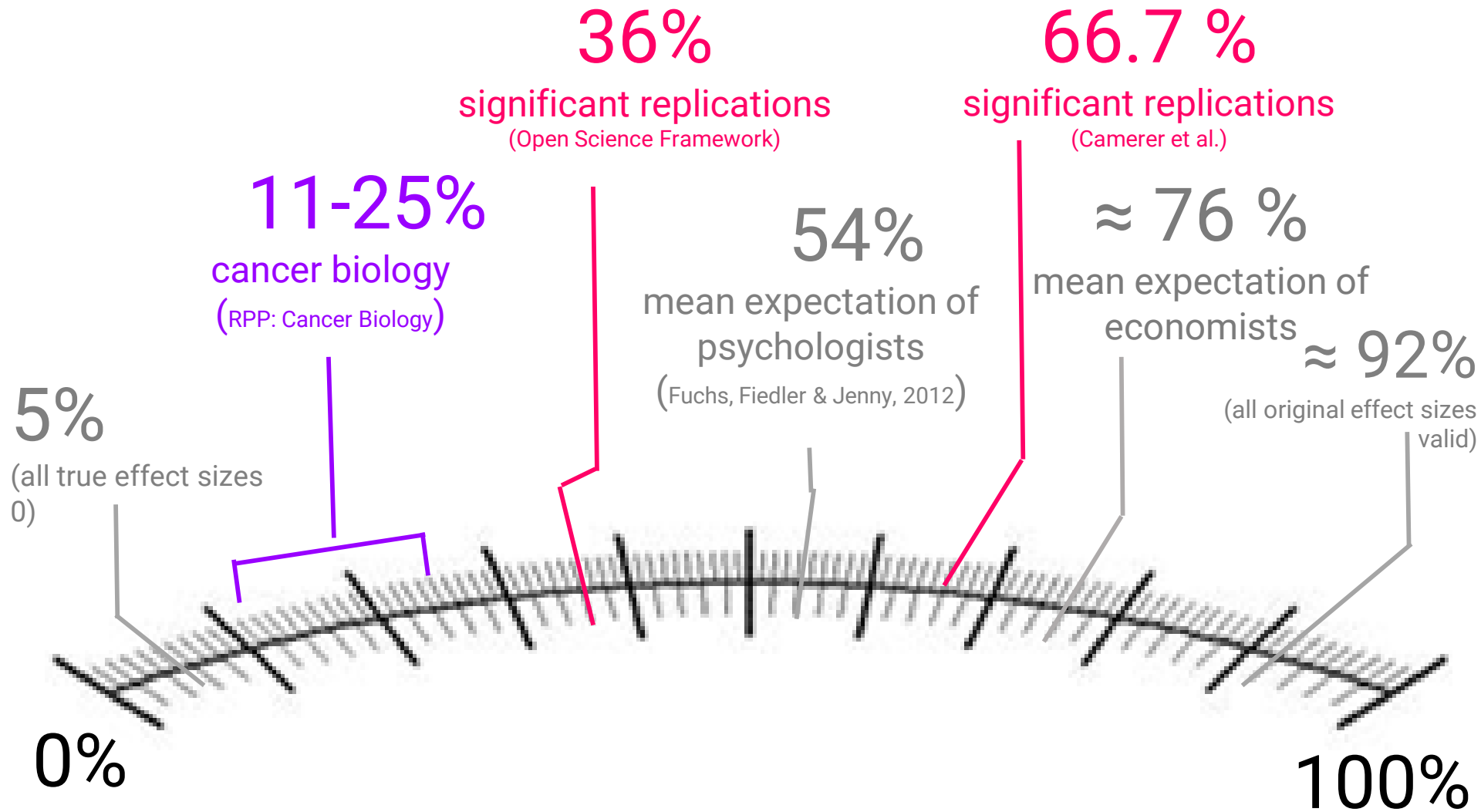
[Read Later](#)

[Like](#)

[Tweet](#)

How trustworthy are the findings from scientific studies?

Predictions of Replication.



SCIENCE —

“Mindless Eating,” or how to send an entire life of research into question

Now questioning: 3,700 citations in 25 different journals—and eight books—over 20+ years.

CATHLEEN O'GRADY - 4/24/2017, 1:30 PM



52 publications from Brian Wansink which (are alleged to) contain minor to very serious issues (which are summarized in this post), which have been cited **over 4000** times, are published in **over 25** different journals, and in **8** books, spanning **over 20 years** of research, **7** articles have been retracted, and **15** articles have been corrected.

The Generalizability of Survey Experiments*

Kevin J. Mullinix^(a1), Thomas J. Leeper^(a2), James N. Druckman^(a3) and Jeremy Freese^(a4) 

DOI: <https://doi.org/10.1017/XPS.2015.19> Published online: 12 January 2016

Abstract

Survey experiments have become a central methodology across the social sciences. Researchers can combine experiments' causal power with the generalizability of population-based samples. Yet, due to the expense of population-based samples, much research relies on convenience samples (e.g. students, online opt-in samples). The emergence of affordable, but non-representative online samples has reinvigorated debates about the external validity of experiments. We conduct two studies of how experimental treatment effects obtained from convenience samples compare to effects produced by population samples. In Study 1, we compare effect estimates from four different types of convenience samples and a population-based sample. In Study 2, we analyze treatment effects obtained from 20 experiments implemented on a population-based sample and Amazon's Mechanical Turk (MTurk). The results reveal considerable similarity between many treatment effects obtained from convenience and nationally representative population-based samples. While the results thus bolster confidence in the utility of convenience samples, we conclude with guidance for the use of a multitude of samples for advancing scientific knowledge.

Is Economics Research Replicable? Sixty Published Papers from
Thirteen Journals Say "Usually Not"

Replication and contradiction of highly cited research papers in psychiatry: 10-year follow-up

Aran Tajika, Yusuke Ogawa, Nozomi Takeshima, Yu Hayasaka and Toshi A. Furukawa

Background

Contradictions and initial overestimates are not unusual among highly cited studies. However, this issue has not been researched in psychiatry.

Aims

To assess how highly cited studies in psychiatry are replicated by subsequent studies.

Method

We selected highly cited studies claiming effective psychiatric treatments in the years 2000 through 2002. For each of these studies we searched for subsequent studies with a better-controlled design, or with a similar design but a larger sample.

Results

Among 83 articles recommending effective interventions,

40 had not been subject to any attempt at replication, 16 were contradicted, 11 were found to have substantially smaller effects and only 16 were replicated. The standardised mean differences of the initial studies were overestimated by 132%. Studies with a total sample size of 100 or more tended to produce replicable results.

Conclusions

Caution is needed when a study with a small sample size reports a large effect.

Declaration of interest

None.

Copyright and usage

© The Royal College of Psychiatrists 2015.

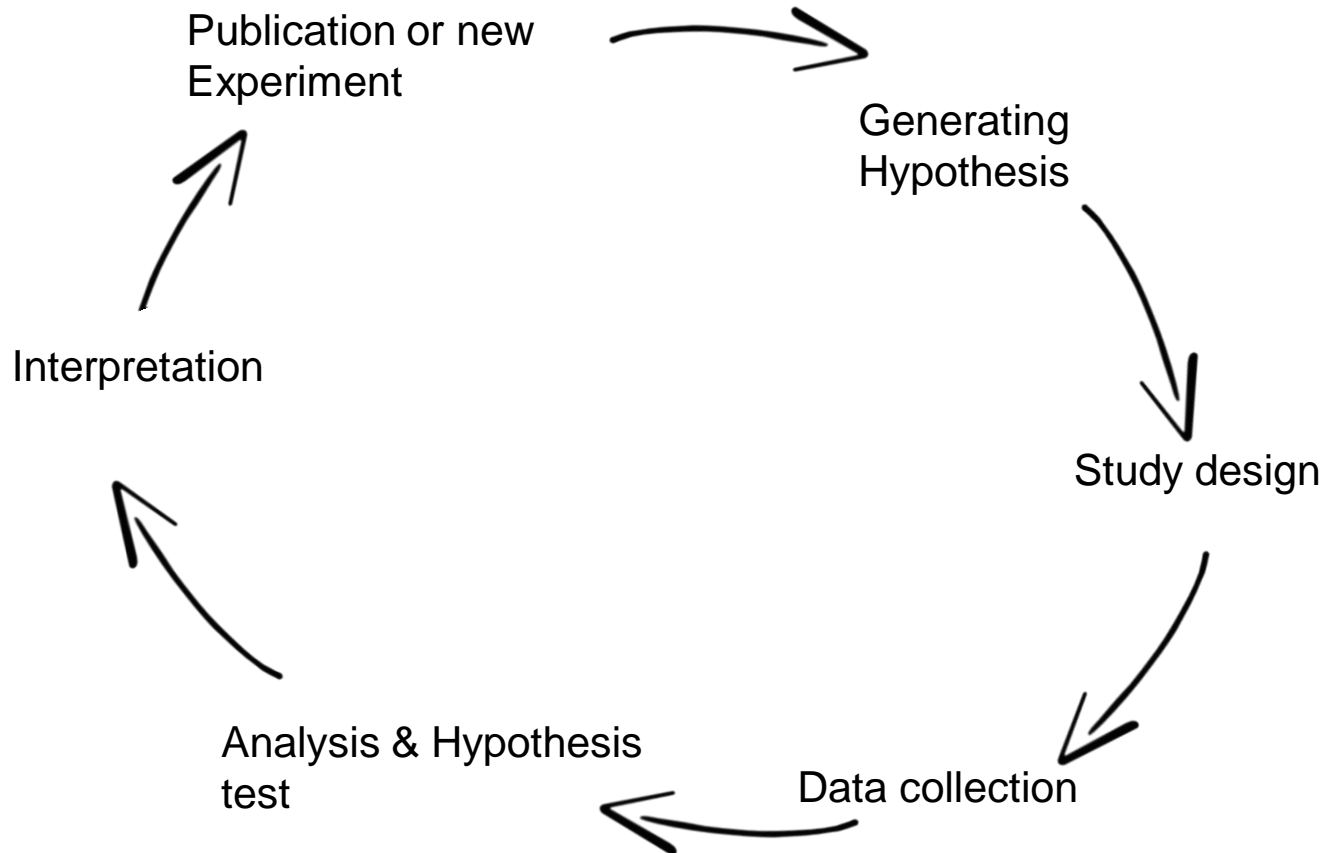
Causes and Effects

The incentive system

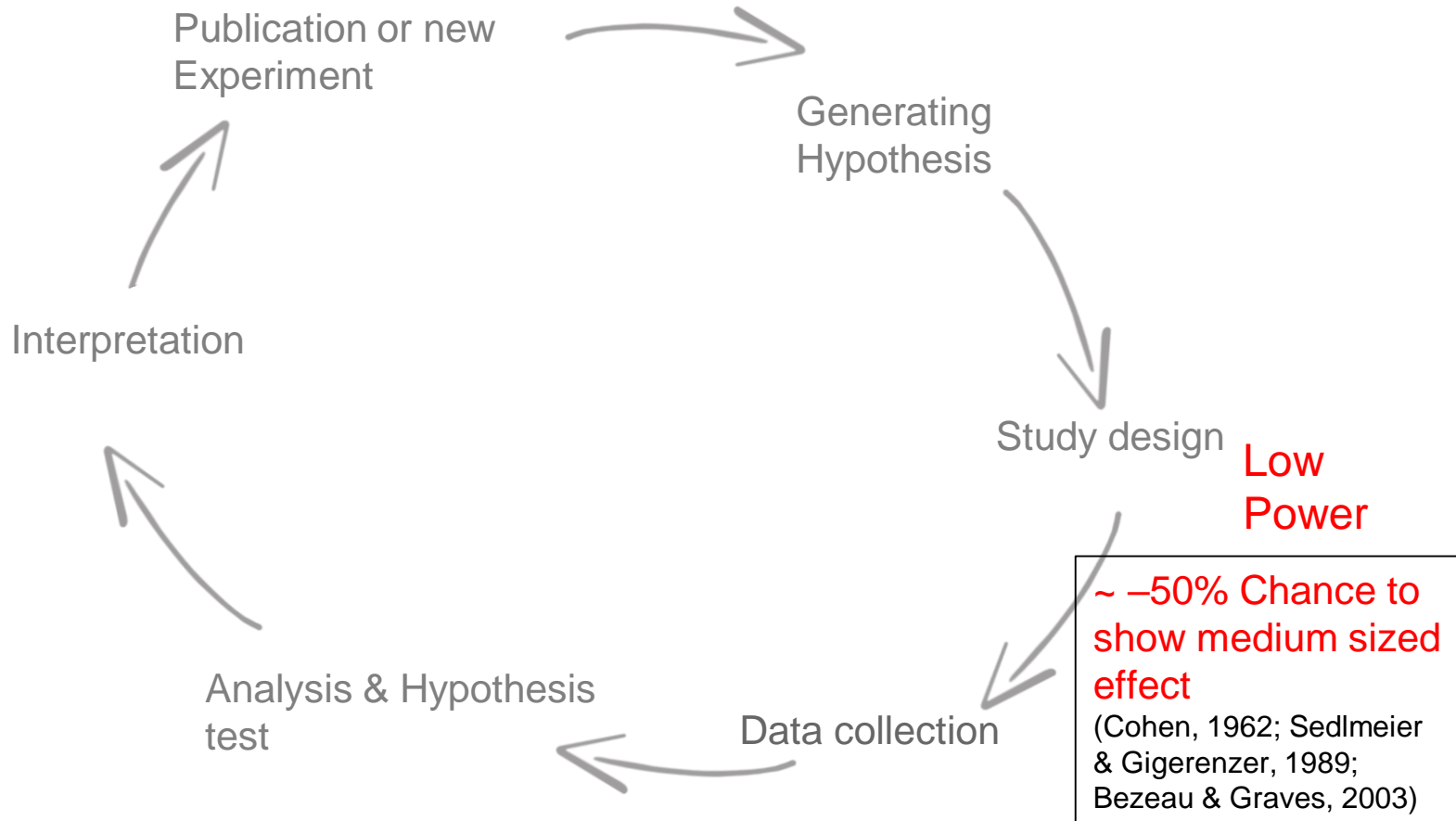
Culture of „Publish or Perish!“

- Publishing as much as possible (count of IF)
 - Multiple underpowered studies („efficient resource allocation“)
- Preferential publishing of statistically significant studies
- Preferential publishing of surprising (counter-intuitive) and catchy results
- Data are rather rhetorical devices

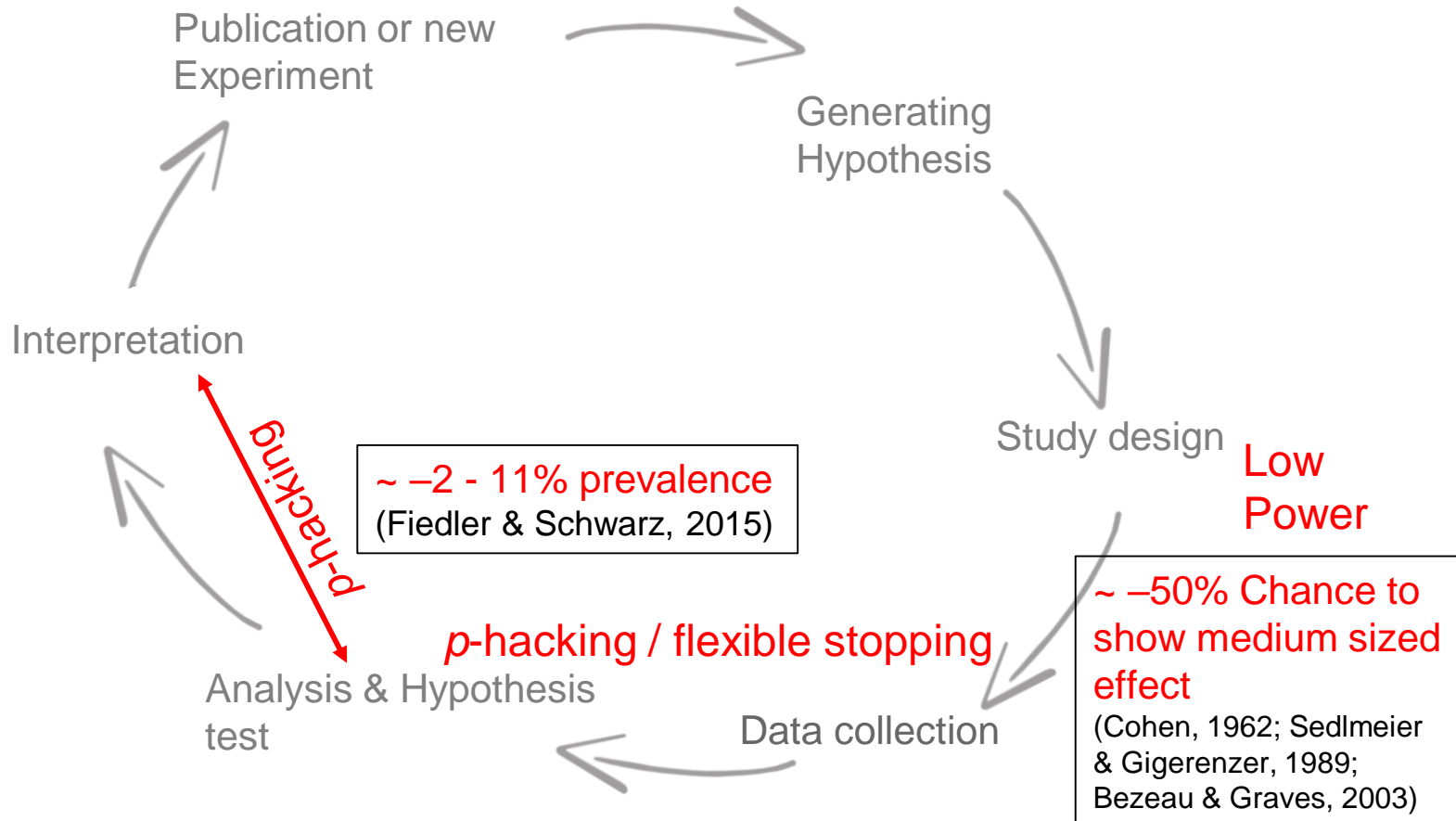
Research Cycle



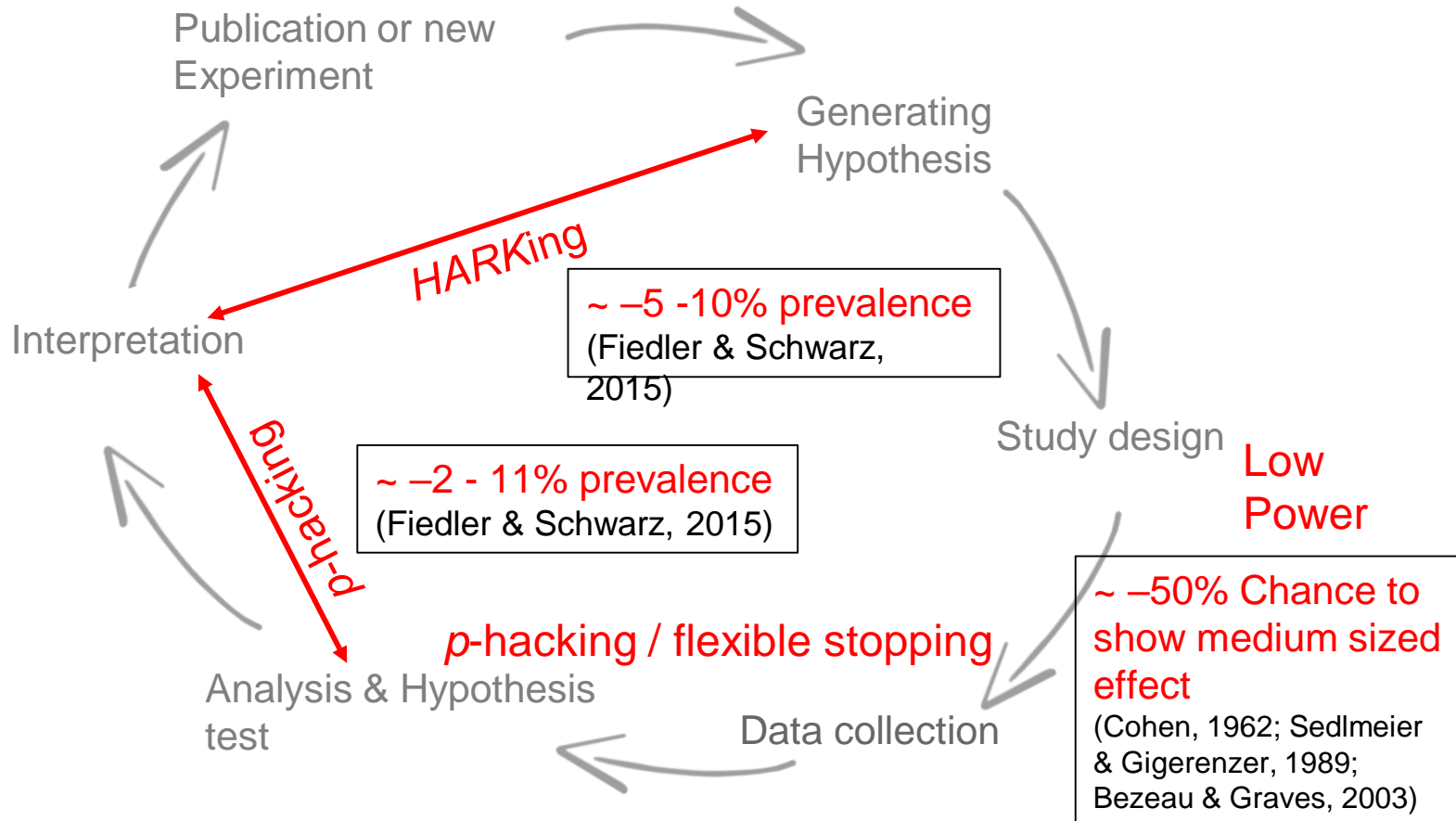
Research Cycle: Goes Wrong



Research Cycle: Goes Wrong



Research Cycle: Goes Wrong



Research Cycle: Goes Wrong

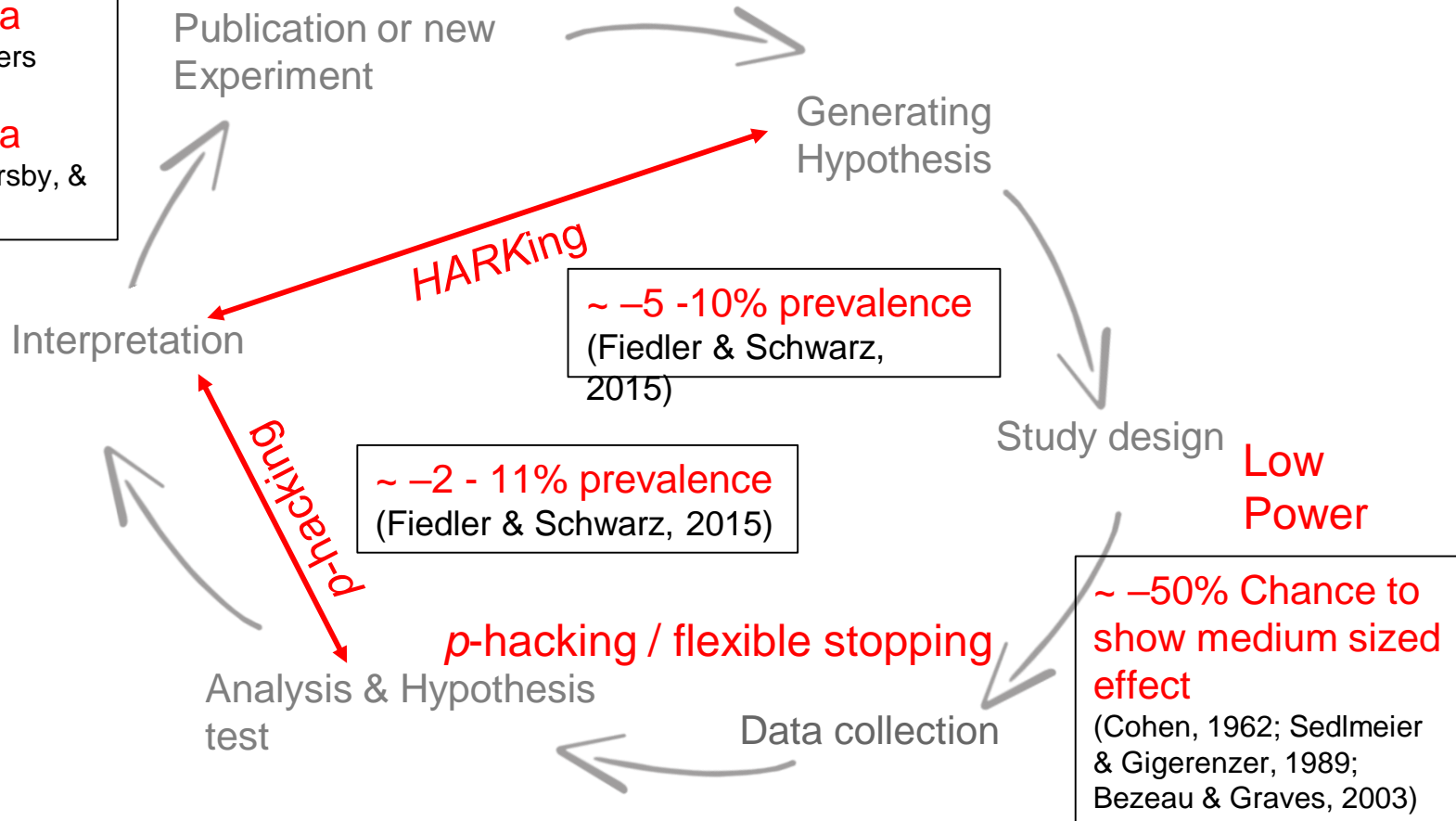
Publication bias & lack in transparency

~ -70% no data

(Psychology, Wichters et al. 2006)

~ -66% no data

(AER, Dewald, Thursby, & Anderson, 1986)



Research Cycle: Goes Wrong

Publication bias & lack in transparency

~ -70% no data

(Psychology, Wichters et al. 2006)

~ -66% no data

(AER, Dewald, Thursby, & Anderson, 1986)

Publication or new Experiment

Generating Hypothesis

Lack of replications

1 in 1000 papers
(Makel et al. 2012)

27% should be replications
(Fuchs et al. 2011)

HARKing

~ -5 -10% prevalence
(Fiedler & Schwarz, 2015)

Interpretation

p-hacking

~ -2 - 11% prevalence
(Fiedler & Schwarz, 2015)

Analysis & Hypothesis test

p-hacking / flexible stopping

Data collection

Study design

Low Power

~ -50% Chance to show medium sized effect
(Cohen, 1962; Sedlmeier & Gigerenzer, 1989; Bezeau & Graves, 2003)

Open Science Ambassadors

Research Cycle: Goes Wrong

October 2015 - January 2016: top 5 medical journals
([NEJM](#), [JAMA](#), [The Lancet](#), [Annals of Internal Medicine](#), [BMJ](#))



Here's what we found.

67

TRIALS CHECKED

9

TRIALS WERE
PERFECT

354

OUTCOMES NOT
REPORTED

357

NEW OUTCOMES
SILENTLY ADDED

On average, each trial reported just 58.2% of its specified outcomes. And on average, each trial silently added 5.3 new outcomes.

<http://compare-trials.org>

False Discovery Rate

- The **smaller the studies conducted** in a scientific field...
- The **smaller the effect sizes** in a scientific field...
- The **greater the number** and the lesser the selection **of tested relationships** in a scientific field..
... the less likely the research findings are to be true.

(Ioannidis, 2005)

Norms

Problematic

Secrecy

Closed

Self-interestedness

Treat science as a competition

Organized dogmatism

Invest career promoting one's own theories, findings

Quantity

Norms

Lack of Open Data and Materials

P-hacking, HARKing, publication bias

Lack of replications, publication bias

Low statistical power

Problematic

Secrecy

Closed

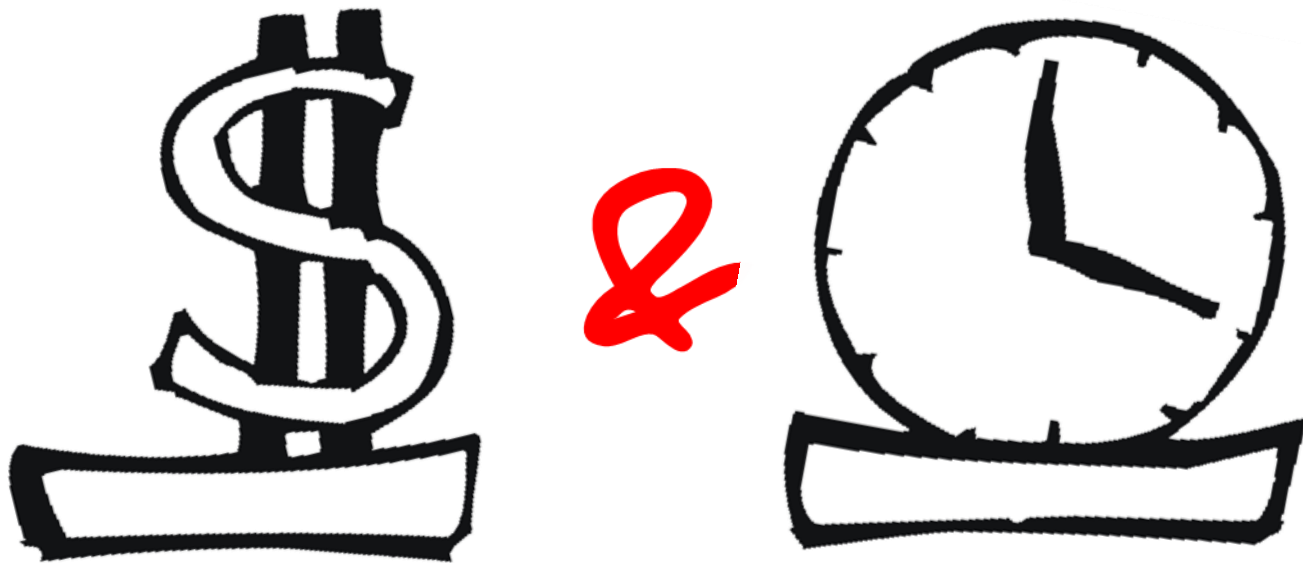
Self-interestedness

Treat science as a competition

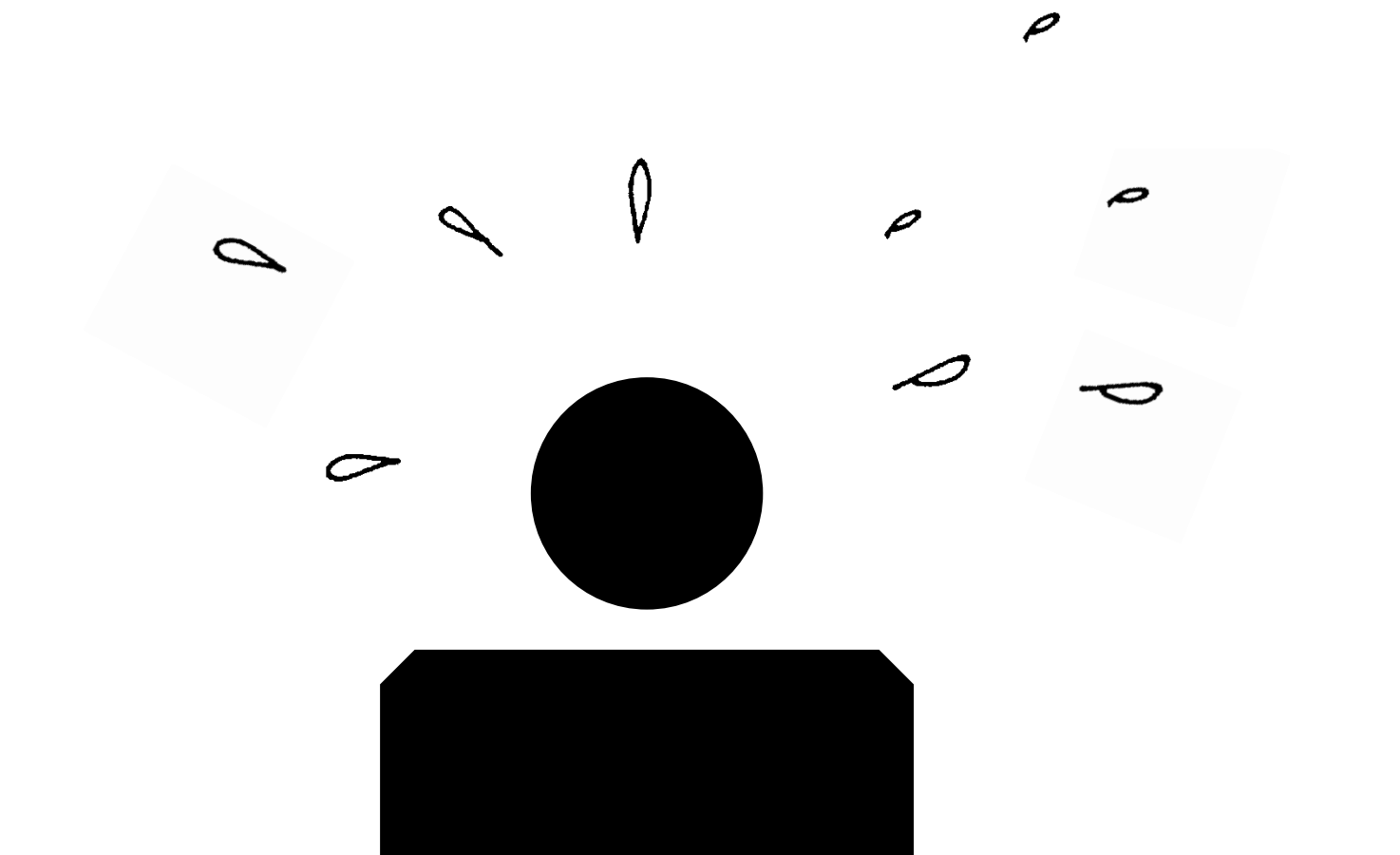
Organized dogmatism

Invest career promoting one's own theories, findings

Quantity



consuming

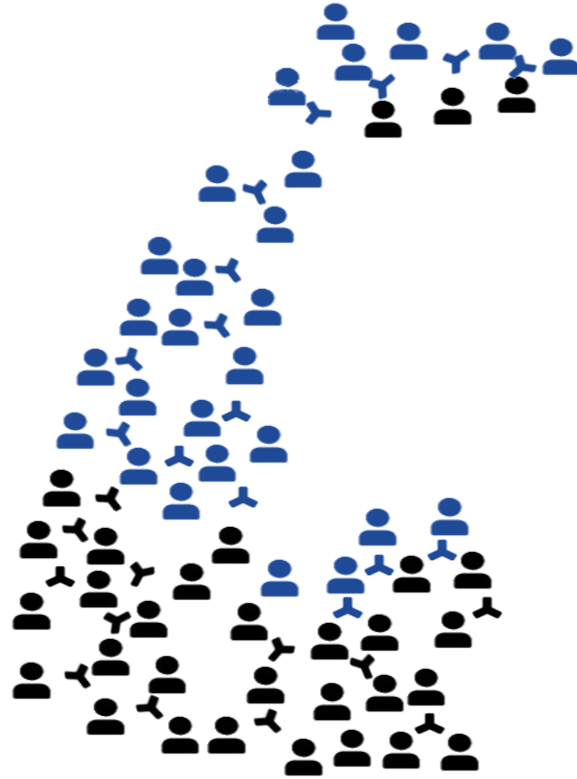


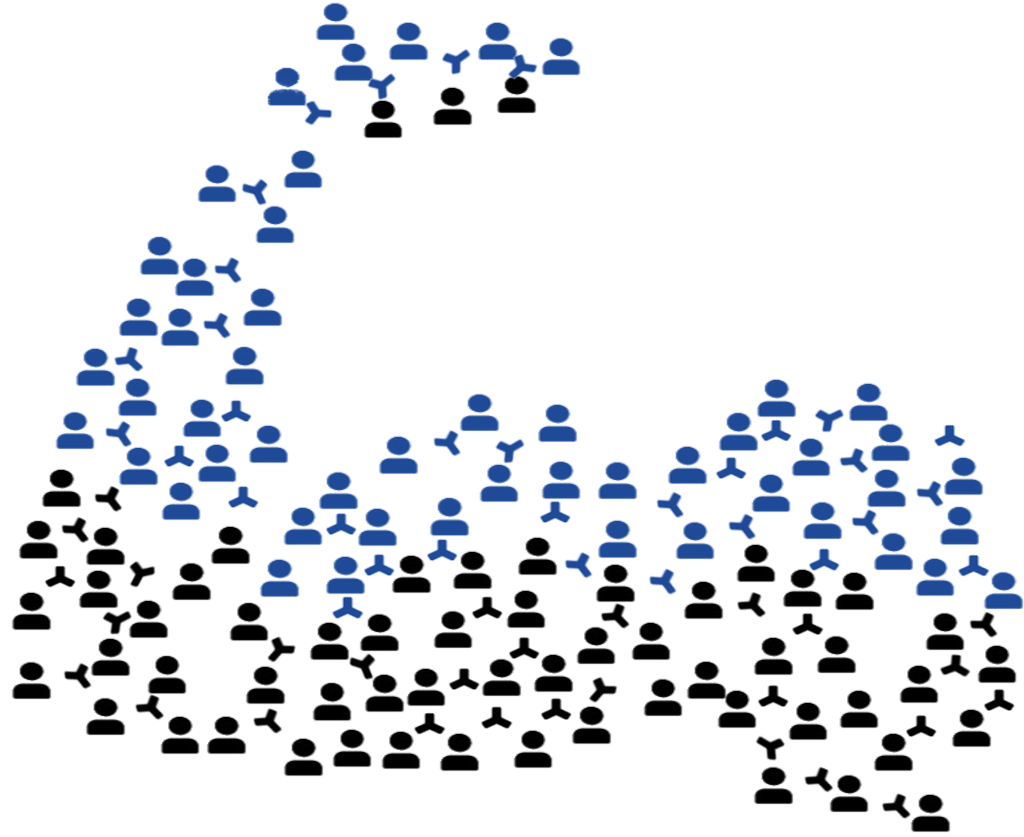
CHANGE

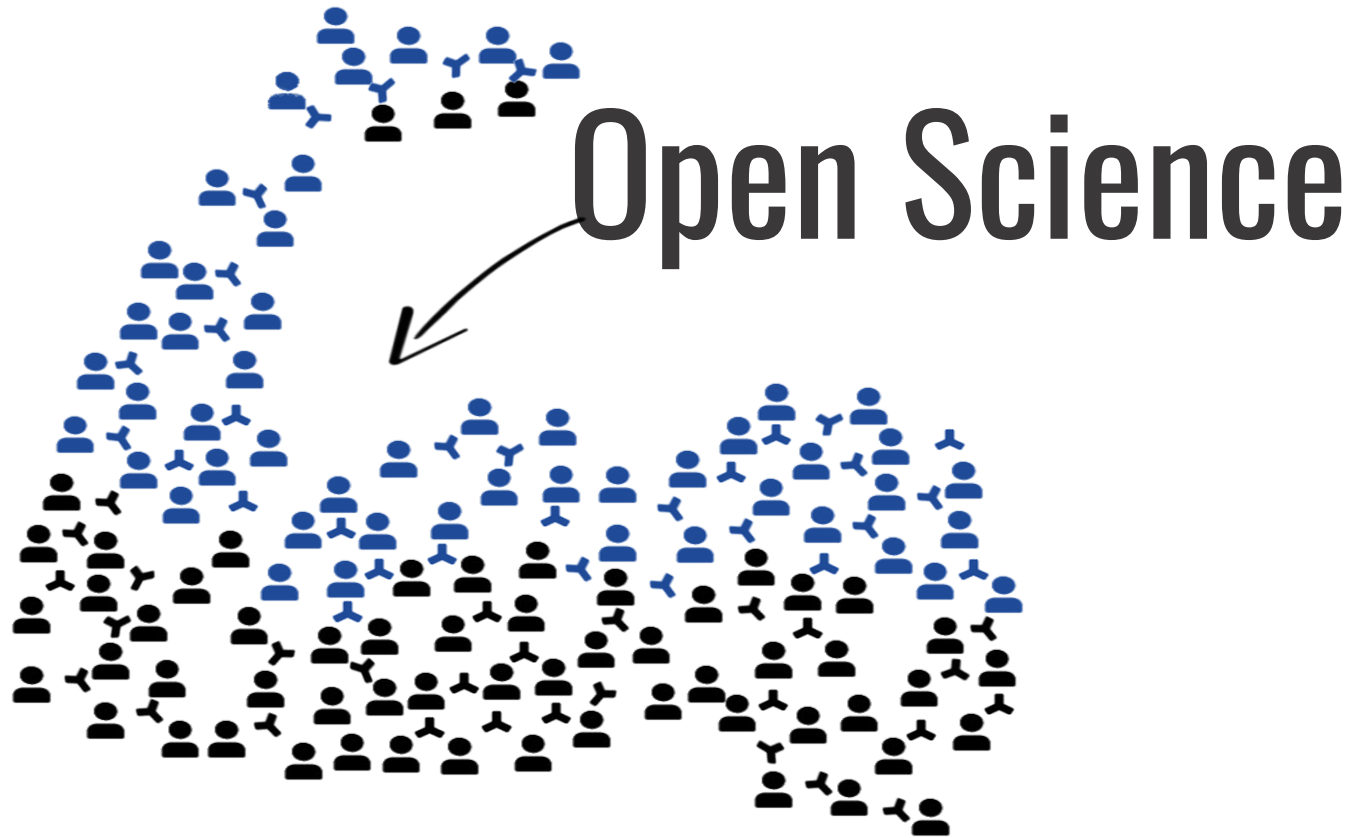












Who makes the rules?



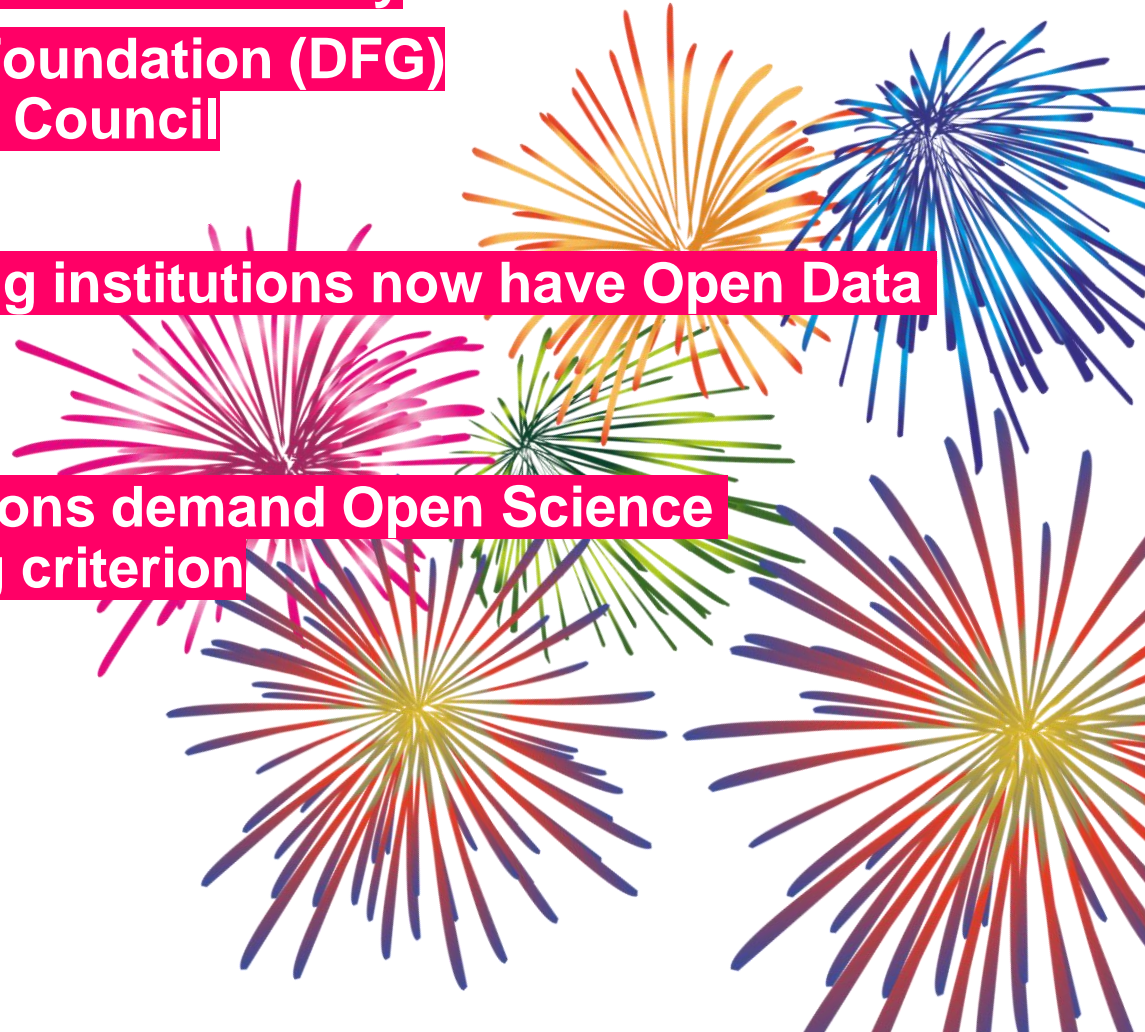
„There is no single solution for such a complex problem“ (Pashler & Wagenmakers, 2012)

The rules ARE changing.

**Open Science principles demanded by
German Research Foundation (DFG)
European Research Council**

**40% of scientific funding institutions now have Open Data
guidelines.**

**More and more institutions demand Open Science
commitment as a hiring criterion**





2019 Research

2019 Researcher



2019 Researcher



actually make a plan ...

Pre-registration

theguardian

[News](#) [Sport](#) [Comment](#) [Culture](#) [Business](#) [Money](#) [Life & style](#)

[News](#) [Science](#) [Peer review and scientific publishing](#)

Trust in science would be improved by study pre-registration

Open letter: We must encourage scientific journals to accept studies before the results are in

Chris Chambers, Marcus Munafo and more than 80 signatories
theguardian.com, Wednesday 5 June 2013 12.45 BST

[Jump to comments \(43\)](#)



The quest: a better understanding of nature. Photograph: Sebastian Kaulitzki/Alamy

In an ideal world, scientific discoveries would be independent of what scientists wanted to discover. A good researcher would begin with an

THE WORLD
UNIVERSITY
RANKINGS

[PROFESSIONAL](#) [JOBS](#) [SUMMITS](#) [RANKINGS](#) [STU](#)

Pre-registration would put science in chains

The pre-registration of study designs must be resisted, says Sophie Scott

July 25, 2013



Science is not well served by people deciding that their methodology is the only legitimate one

What Problems Does Pre-registration fix?

- Pre-registration makes the distinction between confirmatory and exploratory research more clear.

Context of confirmation

- Traditional hypothesis testing
- Results held to the highest standards of rigor
- Goal is to minimize false positives – don't rely on anything false!

p-values interpretable

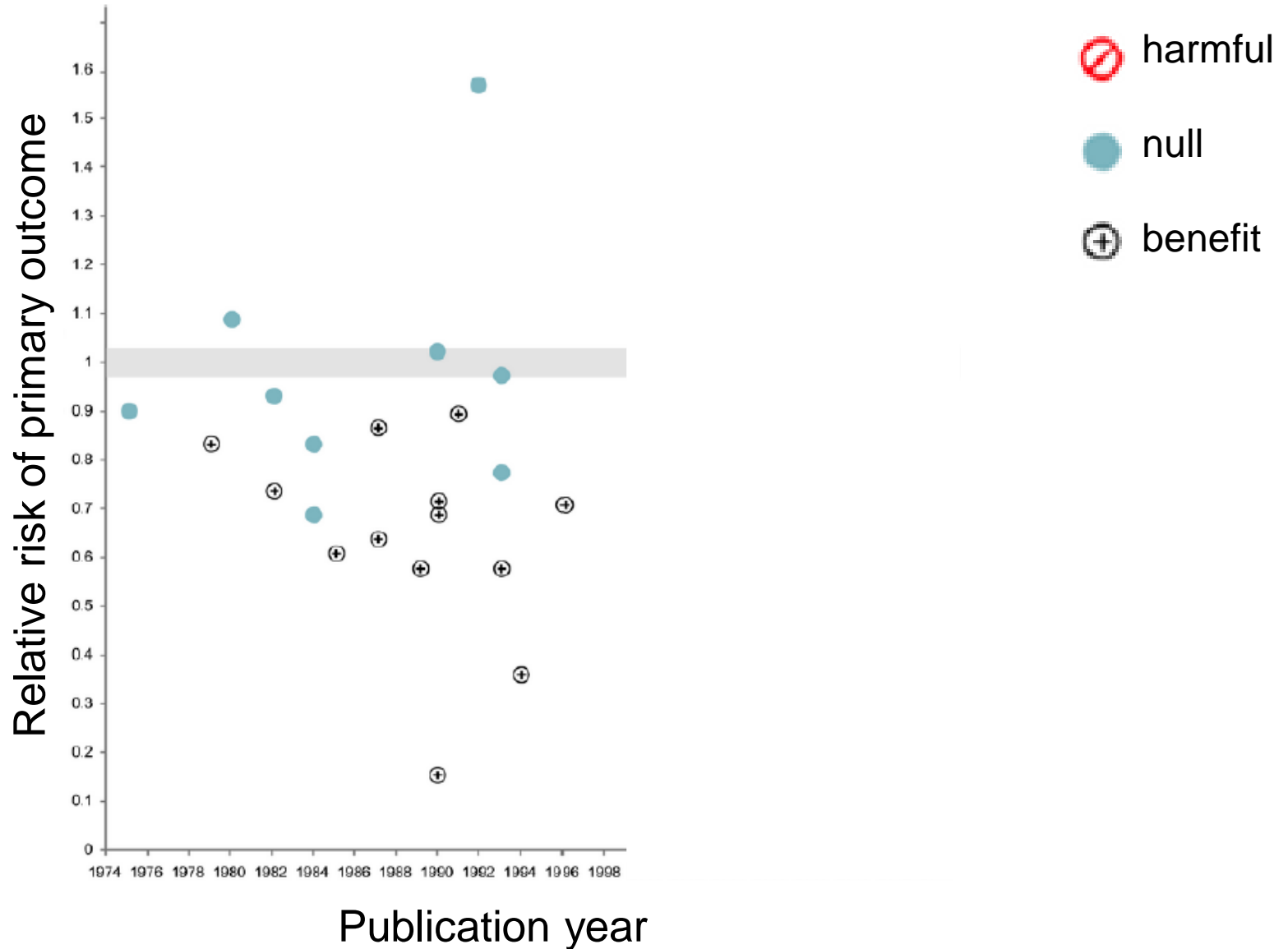
Context of discovery

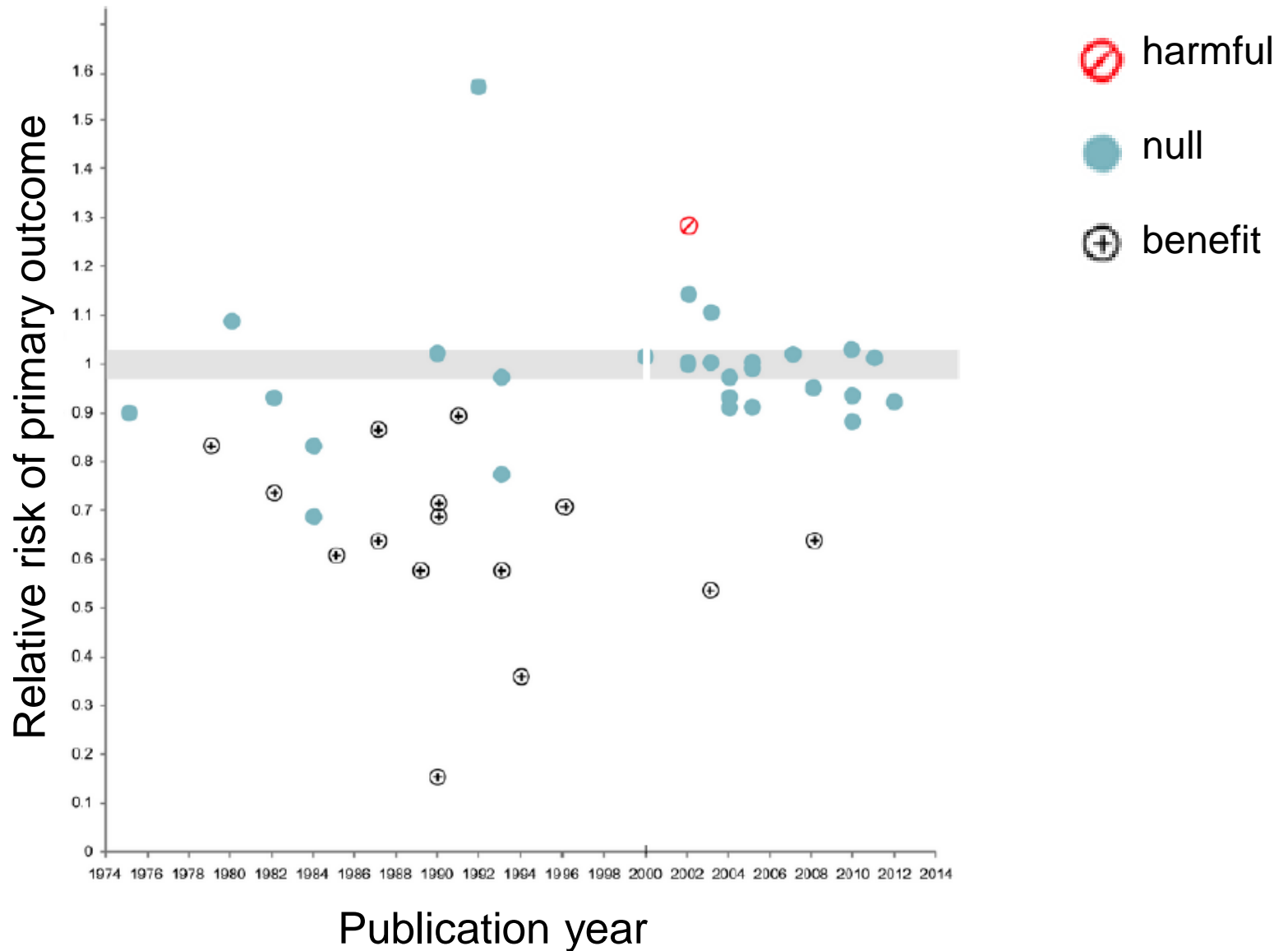
- Pushes knowledge into new areas/ data-led discovery
- Finds unexpected relationships
- Goal is to minimize false negatives – don't miss anything new!

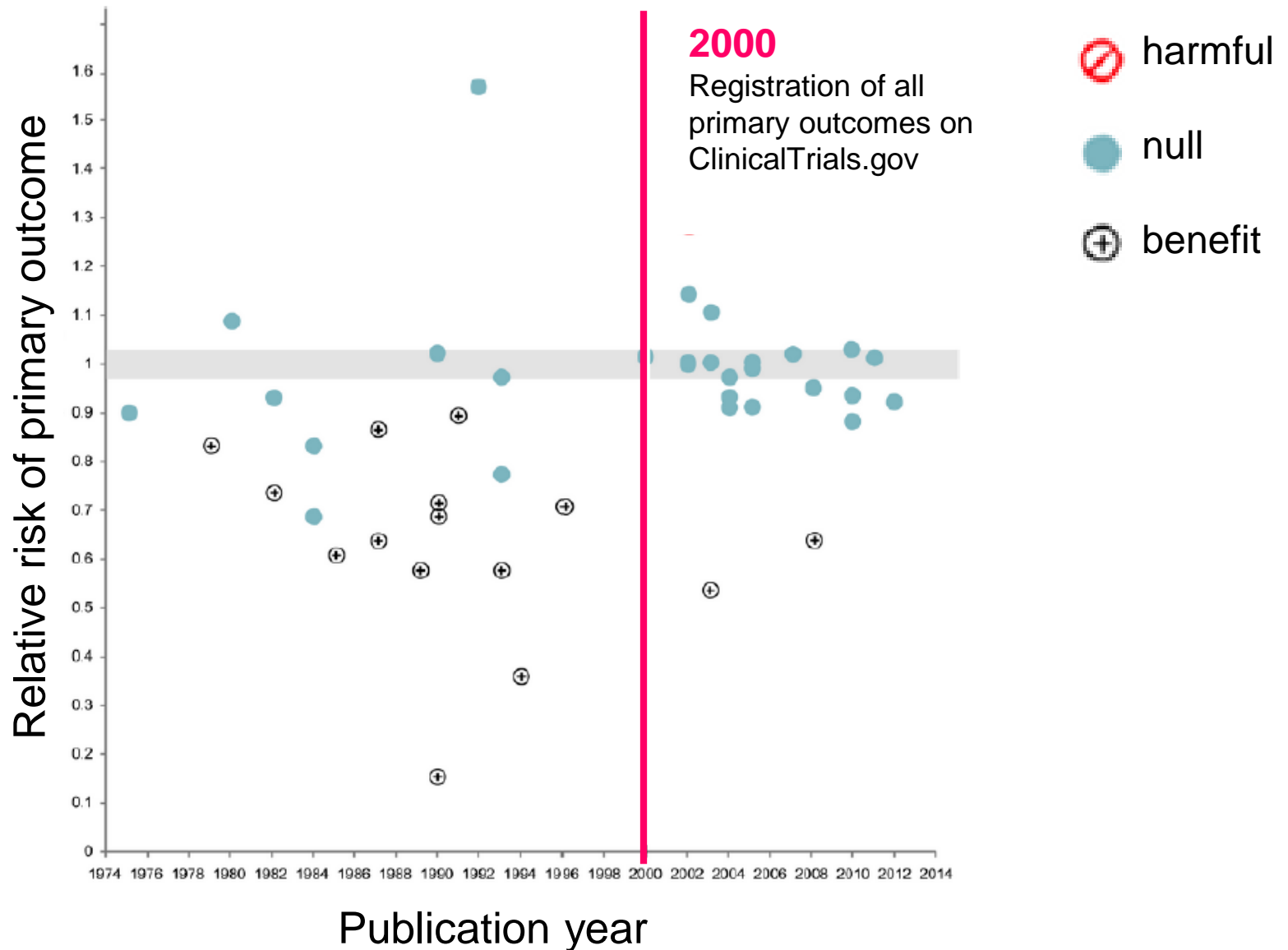
p-values meaningless

- As a field we want to produce both new discoveries AND reliable evidence (or even well tested theories)

How will this probably look like?







Create an Analysis Plan

```
*****
*       areage example
*****

. all
. c mem 600m
use "C:\Users\User\Desktop\trade_test.dta"

** Step1: sorting your data for panel format
sort id year

** Step2: data prep for 5 year areage given the data coverage 1990-2010
gen year2 = .
replace year2 = 1 if year>=1990 & year<1995
replace year2 = 2 if year>=1995 & year<2000
replace year2 = 3 if year>=2000 & year<2005
replace year2 = 4 if year>=2005

** Step3: Calculating 5 year average using a variable you just generated
** This will serve your new id
collapse(mean) trade, by(id countryname year2)

*****
*       are
*****
```

- Extensive degrees of freedom in the analysis
 - 241 fMRI studies reported 223 unique combinations of data cleaning and analysis (Carp, 2012)
 - When using all these combinations 90% of all voxels show a significant difference in at least one analysis (Carp, 2012)
- Decide before the experiment: What are the critical conditions, measures, exclusion criteria, potential moderators and covariates
 - This is the essence of confirmatory testing

It's becoming the norm

2016



Sunday

Total

122309

Files

87056

Users

27296

Components

3709

Projects

2917

Registrations

1323



Search registrations...

294,249 searchable registrations as of August 25, 2019



292,926 new registrations

Changing Incentives



If you have a project that is entering the planning or data collection phase, we'd like you to try out a preregistration. Through our **\$1 Million Preregistration Challenge**, we're giving away \$1,000 to 1,000 researchers who preregister their projects before they publish them. It's straightforward to complete and will really enhance your research output.

[Get Started Now](#)




2019 Researcher



you need the power

Power in Observations



A close-up photograph of a person's legs from the knees down, standing on a silver, circular digital scale. The person is wearing a white skirt. The background is a plain, light-colored wall and floor. A semi-transparent white banner is overlaid across the middle of the image, containing text.

Gender effect on weight ($d = .59$) → around 40 per cell

Adequate power -- joint data collection



Many Babies Project

manybabies.github.io

ManyBabies website

The ManyBabies Project

ManyBabies is a collaborative project for replication and best practices in developmental psychology research. Our goal is to bring labs together to address difficult outstanding theoretical and methodological questions about the nature of early development and how it is studied.

Psychological Science Accelerator



548 Labs in 72 countries



Power in Solving Problems

ATLAS collaboration

Testing the Standard Model

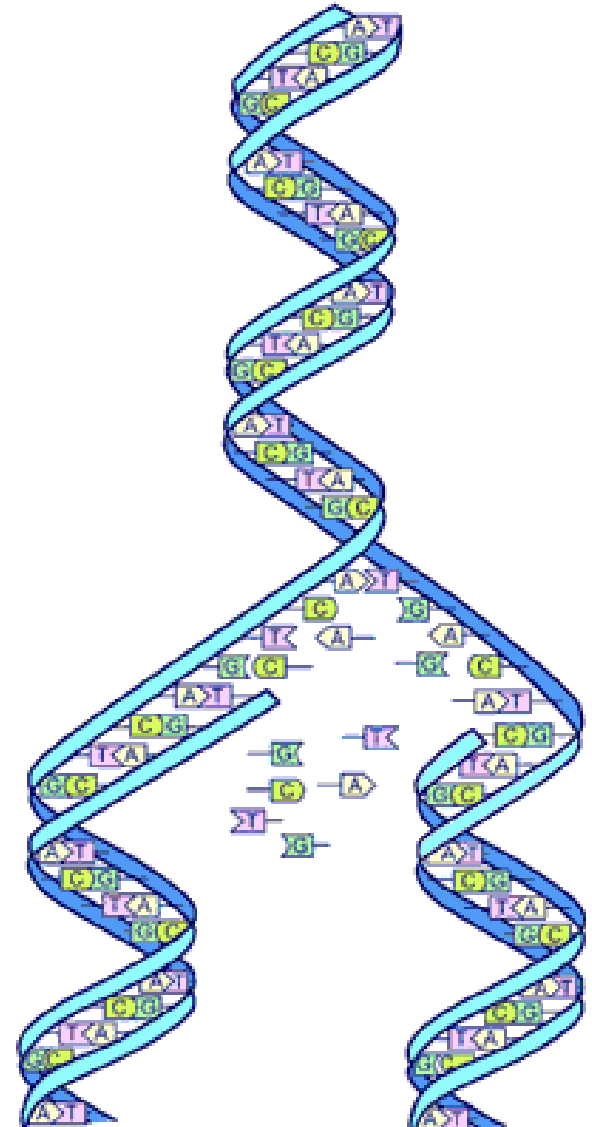
3000 authors from 183 institutions



Human Genome Project

13-year joint effort to sequence the human genome

data now publicly available



Open Hardware

Gathering for Open Science Hardware

[ABOUT GOSH](#) ⌵[JOIN GOSH FORUM](#)[CODE OF CONDUCT](#)[MANIFESTO](#)[ROADMAP](#)[THE GATHERINGS](#) ⌵[NEWS](#)

 [Read the GOSH Roadmap!](#)

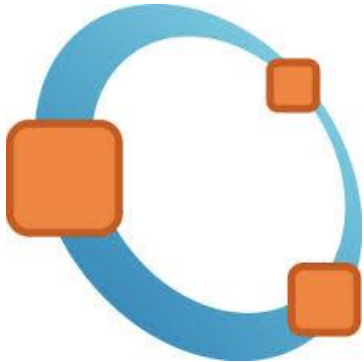


 [GOSH 2018 – 10-13 Oct in Shenzhen](#)



 [Join the GOSH Forum](#)

Open Software



open source
initiative

Open Software



statcheck on the web

[Behavior Research Methods](#)

December 2016, Volume 48, [Issue 4](#), pp 1205–1226 | [Cite as](#)

The prevalence of statistical reporting errors in psychology (1985–2013)

Authors

[Authors and affiliations](#)

Michèle B. Nuijten , Chris H. J. Hartgerink, Marcel A. L. M. van Assen, Sacha Epskamp, Jelte M. Wicherts

2019 Researcher



get it out there ...

Share your work!

Share your Materials, Data & Code

- release your work under a license and indicate explicitly in the paper or in the metadata how you want others to give you credit



Document & Share your Procedure



openlabnotebooks.org

A growing team of groundbreaking scientists around the world are now sharing their lab notebooks online

STRUCTURE SEARCH

HOME

Browse notebooks by
LABORATORIES

Browse notebooks by
PEOPLE

Browse notebooks by
DISEASES

Browse notebooks by
PROJECTS

THE TEAM

ABOUT



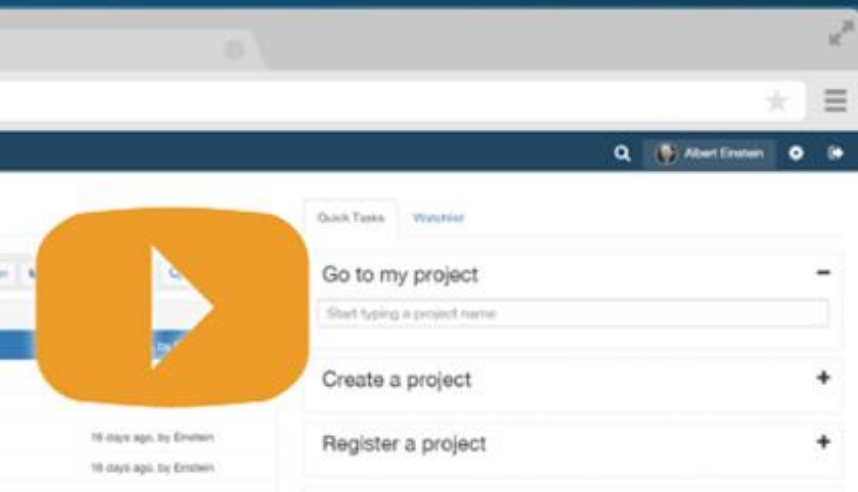
Open Science Framework

A scholarly commons to connect the entire research cycle



FREE AND OPEN SOURCE. ST

By clicking "Sign up free", you agree to our [Terms](#) and that you have read our [Privacy](#)



Facilitate Meta-Analysis

[REPLICATIONS](#)[ABOUT](#)[FEATURES](#)[SANDBOX](#)[FAQ](#)

Social Priming

Schnall, Benton, & Harvey (2008a) -- Replications (7) [Tweet](#) [Link](#)

With a Clean Conscience: Cleanliness Reduces the Severity of Moral Judgments

DOI:[10.1111/j.1467-9280.2008.02227.x](https://doi.org/10.1111/j.1467-9280.2008.02227.x)

[Original Abstract]

Original Studies & Replications

Schnall et al. (2008a) Study 1



Johnson et al. (2014a) Study 1



Johnson et al. (2014b)

Lee et al. (2013)



Arbesfeld et al. (2014)



Besman et al. (2013)



Huang (2014) Study 1

Data/Syntax

Study_1.sav



Exp1_Data.sav



Online_Rep.sav



lee_data.csv



study1.sav

Materials/Pre-reg



OSF folder



OSF folder

N

40

208

736

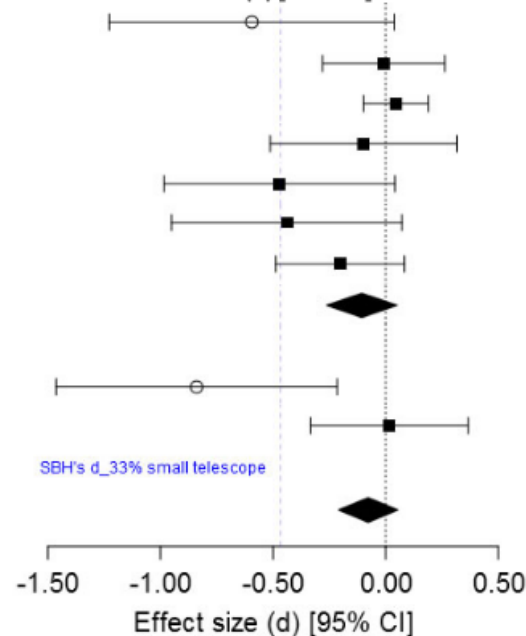
90

60

60

189

Effect size (d) [95% CI]



SBH's d_33% small telescope

Schnall et al. (2008a) Study 2



Johnson et al. (2014a) Study 2



Study_2.sav



Exp2_Data.sav



OSF folder

43

126

Current meta-analytic estimate of all replications (random-effects):

[Underlying data (CSV)] [R-code]

Summary: The main finding that cleanliness priming reduces the severity of moral judgments does not (yet) appear to be replicable (overall meta-

Share your work!

Share your Paper!

- <http://www.sherpa.ac.uk/romeo/index.php>
- Check **which** version of your paper you are allowed to share



Share your paper!



Home

Search - Publisher copyright policies & self-archiving

One journal found when searched for: **journal of personality and social psychology**

Journal:	Journal of Personality and Social Psychology (ISSN: 0022-3514, ESSN: 1939-1315 [1])
RoMEO:	This is a RoMEO green journal
Paid OA:	A paid open access option is available for this journal.
Author's Pre-print:	✓ author can archive pre-print (ie pre-refereeing)
Author's Post-print:	✓ author can archive post-print (ie final draft post-refereeing)
Publisher's Version/PDF:	✗ author cannot archive publisher's version/PDF
General Conditions:	<ul style="list-style-type: none"> • Authors' pre-print on a web-site • Authors' pre-print must be labeled with date and accompanied with statement that paper has not (yet) been published • Copy of authors final peer-reviewed manuscript as accepted for publication • Authors' post-print on author's web-site, employers server or institutional repository, after acceptance • Publisher copyright and source must be acknowledged • Must link to publisher version with DOI • Article must include the following statement: 'This article may not exactly replicate the final version published in the APA journal. It is not the copy of record' • Publisher's version/PDF cannot be used • APA will submit NIH author articles to PubMed Central, after author completion of form
Mandated OA:	(Awaiting information)
Paid Open Access:	Article Sponsorship
Copyright:	Self-archiving Policy - NIH Authors and PubMed
Updated:	20-May-2015 - Suggest an update for this record
Link to this page:	http://www.sherpa.ac.uk/romeo/issn/0022-3514/
Published by:	American Psychological Association - Green Policies in RoMEO
Footnotes:	1. Extra ISSN(s) found - 1939-1293.

Pre-prints



A free preprint service for the psychological sciences
Maintained by [The Society for the Improvement of Psychological Science](#)
[Powered by OSF Preprints](#)

or

[Add a preprint](#)

[See an example](#)

Pre-print servers and general repositories accepting pre-prints.

Preprint server or repository*	Subject areas	Repository open source?	Public API?	Can leave feedback?†	Third party persistent ID?
arXiv arxiv.org	physics, mathematics, computer science, quantitative biology, quantitative finance, statistics	No	Yes	No	No‡
bioRxiv biorxiv.org	biology, life sciences	No	No	Yes	Yes (DOI)
CERN document server cds.cern.ch	high-energy physics	Yes (GPL)	Yes	No	No
Cogprints cogprints.org	psychology, neuroscience, linguistics, computer science, philosophy, biology	No	Yes	No	No
EconStor econstor.eu	economics	No	Yes	No	Yes (Handle)
e-LiS eprints.rclis.org	library and information sciences	No§	Yes	No	Yes (Handle)
figshare figshare.com	general repository for all disciplines	No	Yes	Yes	Yes (DOI)
Munich Personal RePEc Archive mpra.ub.uni-muenchen.de	economics	No¶	Yes	No	No
Open Science Framework osf.io	general repository for all disciplines	Yes (Apache 2)	Yes	Yes	Yes (DOI/ARK)
PeerJ Preprints peerj.com/archive-s-preprints	biological, life, medical, and computer sciences	No	Yes	Yes	Yes (DOI)
PhilSci Archive philsci-archive.pitt.edu	philosophy of science	No**	Yes	No	No
Self-Journal of Science www.sjscience.org	general repository for all disciplines	No	No	Yes	No
Social Science Research Network ssrn.com	social sciences and humanities	No	No	Yes	Yes (DOI)
The Winnower thewinnower.com	general repository for all disciplines	No	No	Yes	Yes (DOI)††
Zenodo zenodo.org	general repository for all disciplines	Yes (GPLv2)	Yes	No	Yes (DOI)

2019 Researcher



change the game ...

2019 Researcher



Planning

Doing

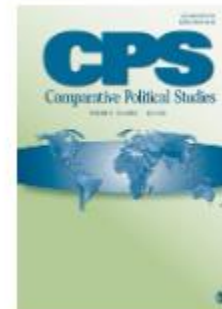
Reporting

change the game ...

Who Publishes Registered Reports?



Neuroscience



eLIFE

204 journals currently accept registered reports

See the full list: <http://cos.io/rr>

Research Cycle: Goes Wrong

Publication bias & lack in transparency

~ -70% no data

(Psychology, Wichers et al. 2006)

~ -66% no data

(AER, Dewald, Thursby, & Anderson, 1986)

Publication or new Experiment

Generating Hypothesis

Lack of replications

1 in 1000 papers
(Makel et al. 2012)

27% should be replications
(Fuchs et al. 2011)

Interpretation

HARKing

~ -5 -10% prevalence
(Fiedler & Schwarz, 2015)

Study design

~ -2 - 11% prevalence
(Fiedler & Schwarz, 2015)

Low

Power

~ -50% Chance to show medium sized effect
(Cohen, 1962; Sedlmeier & Gigerenzer, 1989; Bezeau & Graves, 2003)

Analysis & Hypothesis test

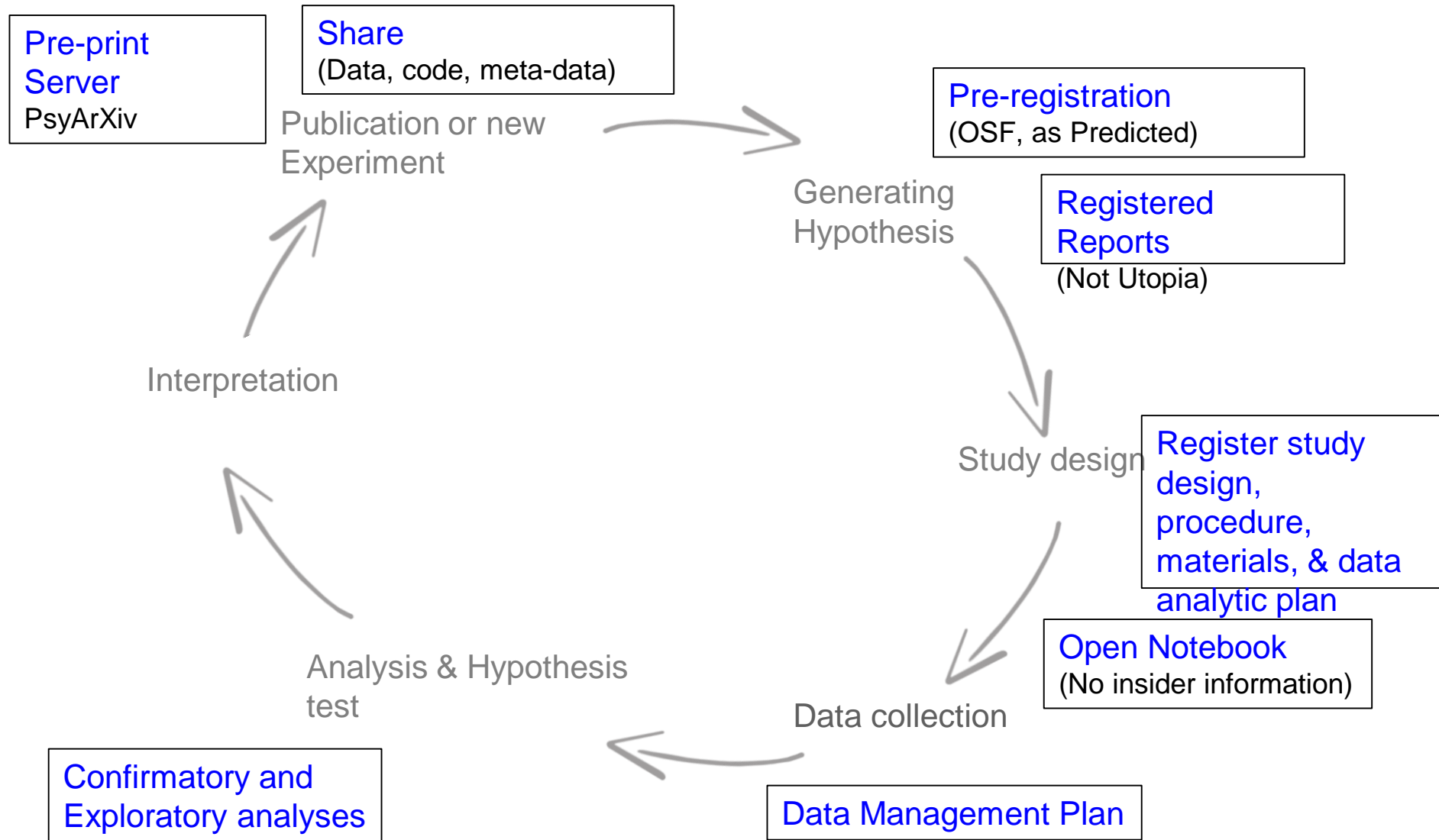
p-hacking / flexible stopping

Data collection

Open Science Ambassadors

p-hacking

Research Cycle: Goes GRRRRREAT!



Norms

Open

Communality

Open sharing

Disinterestedness

Motivated by knowledge and discovery

Organized skepticism

Consider all new evidence, even against one's prior work

Quality

Problematic

Secrecy

Closed

Self-interestedness

Treat science as a competition

Organized dogmatism

Invest career promoting one's own theories, findings

Quantity

Norms

Communality

Open sharing

Disinterestedness

Motivated by knowledge and discovery

Organized skepticism

Consider all new evidence, even against one's prior work

Quality

Open

Open data & Material

**Pre-registration,
Transparency**

**Large scale replication
efforts, Registered
Reports, Curate Science**

**Increased statistical
power**

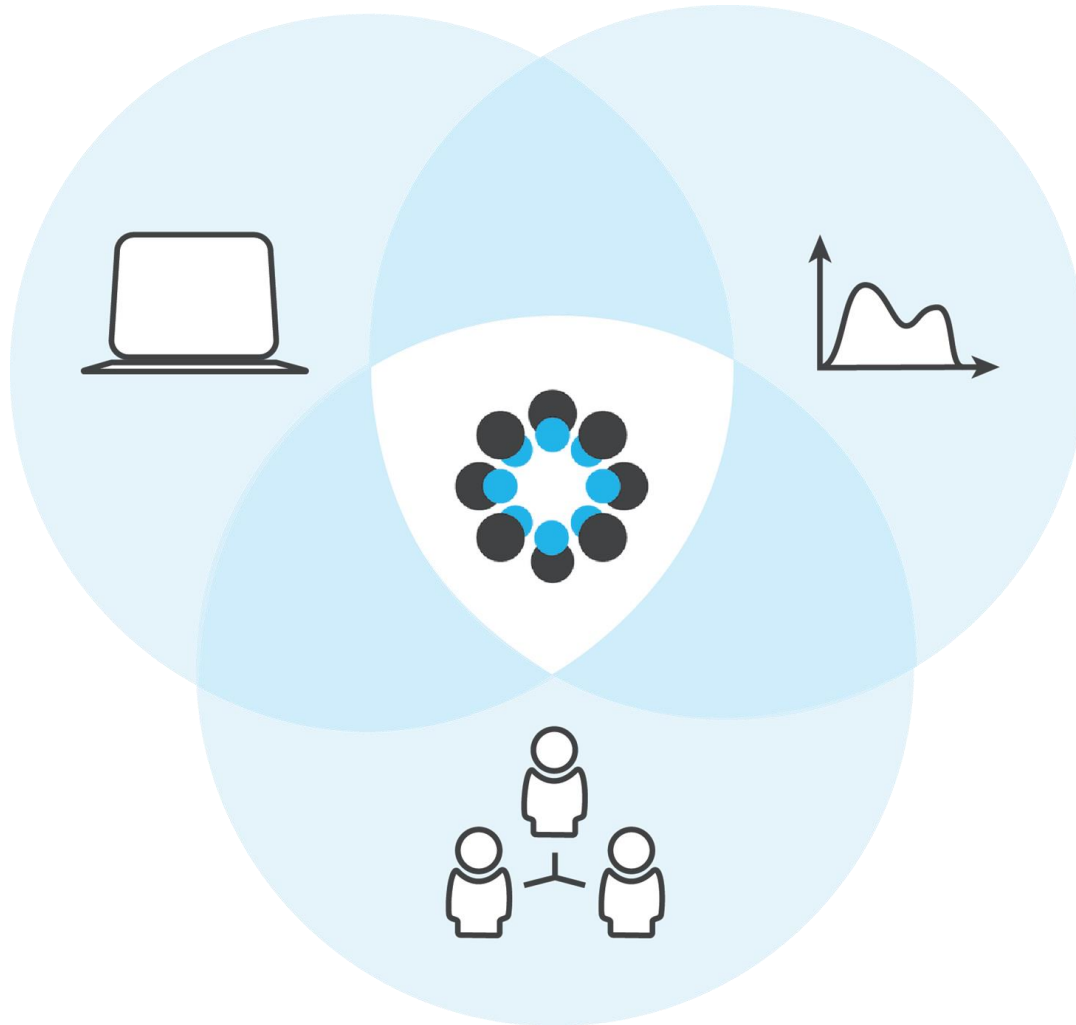
“The first principle is that you must not fool yourself -and you are the easiest person to fool”

Feynman, 1974

Put reality back into the published literature!



Join Your Research Community!



Expand your Research Community



datacolada.org
the100.ci

andrewgelman.com
thehardestscience.com



@openscience
#openscience

Ask Open Science

<https://ask-open-science.org/>



Open Science Radio
The Black Goat



PsychMAP
Psychological Methods
Discussion Group

Learn new and refresh old stuff!



The image is a video thumbnail. On the right side, there is a portrait of a woman with long, wavy brown hair, wearing a dark blue button-down shirt. In the top right corner of the video frame, the Tilburg University logo is visible, consisting of a crest and the text 'TILBURG UNIVERSITY'. On the left side, there is a large pink rectangular box with the text 'Open for Insight' in white, bold, sans-serif font. Below this pink box is a white rectangular box with the text 'Chapter 1: What is true?' in a dark grey, sans-serif font. At the bottom right of the video frame, there is a grey rectangular box containing the text 'Dr. Rima-Maria Rahal' in a bold, sans-serif font, and below it, in a smaller font, 'Tilburg School of Social and Behavioral Sciences'.

Open for Insight

Chapter 1: What is true?

Dr. Rima-Maria Rahal
Tilburg School of Social and Behavioral Sciences

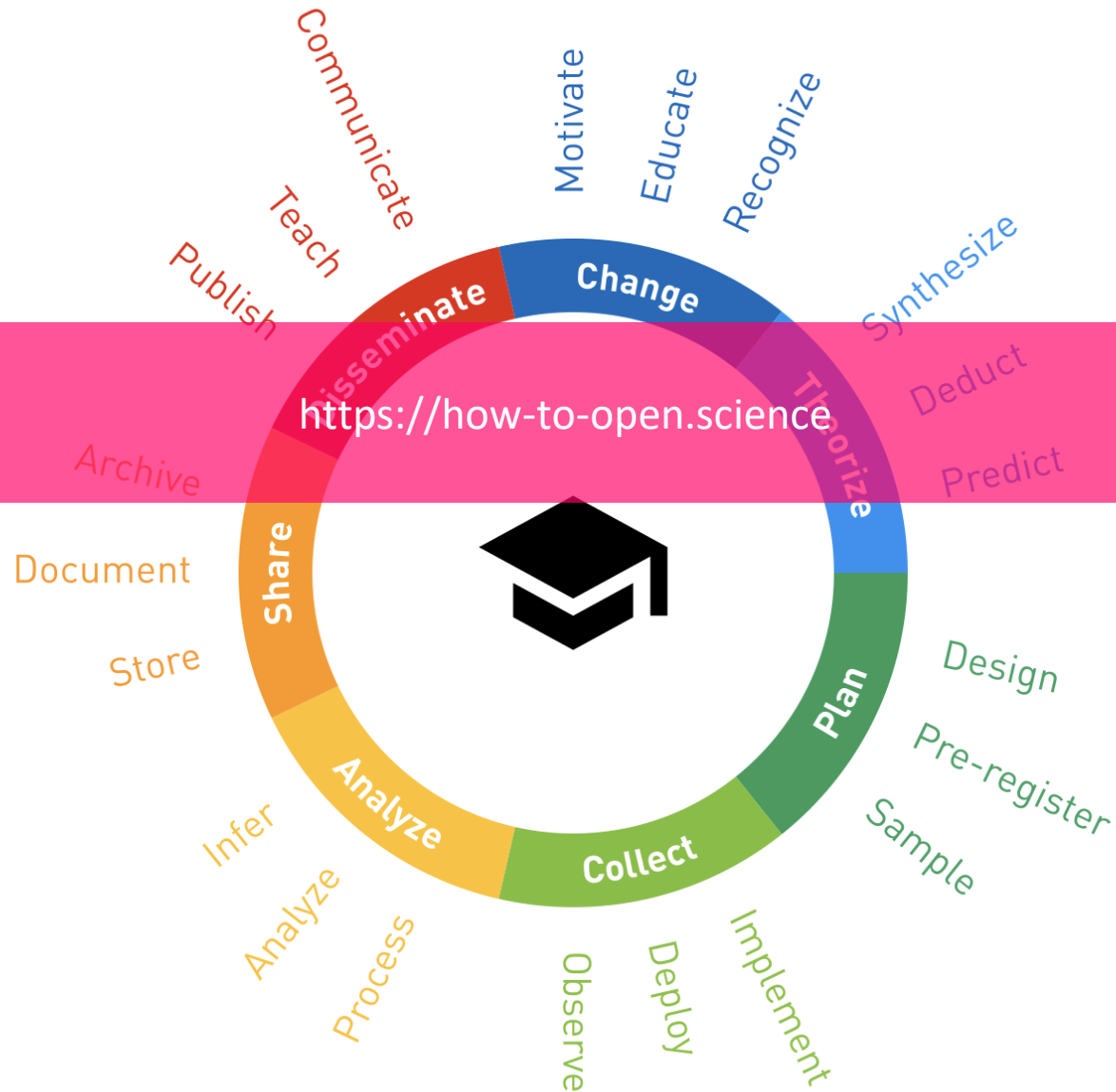
<https://rimamrahal.wordpress.com>

Learn new and refresh old stuff!

The screenshot displays the Coursera website interface. At the top, the Coursera logo is on the left, followed by an 'Explore' dropdown menu and a search bar containing the text 'What do you want to learn?'. On the right, a user profile icon for 'Rima' is visible. A teal banner across the top contains the text 'You can pick up where you left off. Just join a new session and we'll reset your deadlines.' and a 'Join a session' button. Below the banner, the course title 'This Course: Improving your statistical Inferences' is shown, along with navigation links for 'Back to Week 1', 'Lessons', 'Prev', and 'Next'. The left sidebar lists the course structure: 'Course Introduction', 'Week 1: Overview', 'Lecture 1.1', 'Lecture 1.2: What is a p-value?', 'Lecture 1.3: Type 1 and Type 2 errors', 'Assignment 1: Which p-values can you expect?', and 'Exam Week 1'. The 'Type 1 and Type 2 errors' lecture is highlighted with a green play button icon and '18 min' duration. The main video player shows a man in a patterned polo shirt standing next to a white text box that reads: 'Error control: Goal: not to make a fool out of yourself too often in the long run.' The video player includes a progress bar at 0:47 / 18:22 and standard playback controls. Below the video, the title 'Type 1 and Type 2 errors' is displayed.

<https://www.coursera.org/learn/statistical-inferences>

Keep asking questions!



Jump right in!

Planning

Doing

Reporting

Scan to read a summary!

<https://osf.io/preprints/metaarxiv/3hb6g>



change the game ...