

OPEN SCIENCE: ADVANTAGES & CHALLENGES

**Simine Vazire
UC Davis**

THE CREDIBILITY REVOLUTION

The open science movement is not just about openness
The replicability crisis is not just about replicability
The common goal is credibility

OVERVIEW

PART I: Values and norms in science

PART II: Stats review

PART III: Crisis

PART IV: Solutions

PART I: VALUES AND NORMS

PART I: Values and norms in science

PART II: Stats review

PART III: Crisis

PART IV: Solutions

VALUES AND NORMS IN SCIENCE

What makes science different from other ways of knowing?

MERTON'S NORMS

■ Universalism

- The validity of a scientific claim does not depend on who is making it.
- No hierarchy. Status should not matter.

■ Communalism

- The findings of science belong to everyone, they are not private property.
- No secrecy. Open communication is key.

■ Disinterestedness

- Scientists should be focused on finding the truth, not on their own success.
- No self-interest. Report whatever you find, even if it makes you look bad.

■ Organized skepticism

- Do not take things at face value. Verify others' claims.
- Nothing is sacred.

DOES SCIENCE FOLLOW MERTON'S NORMS?

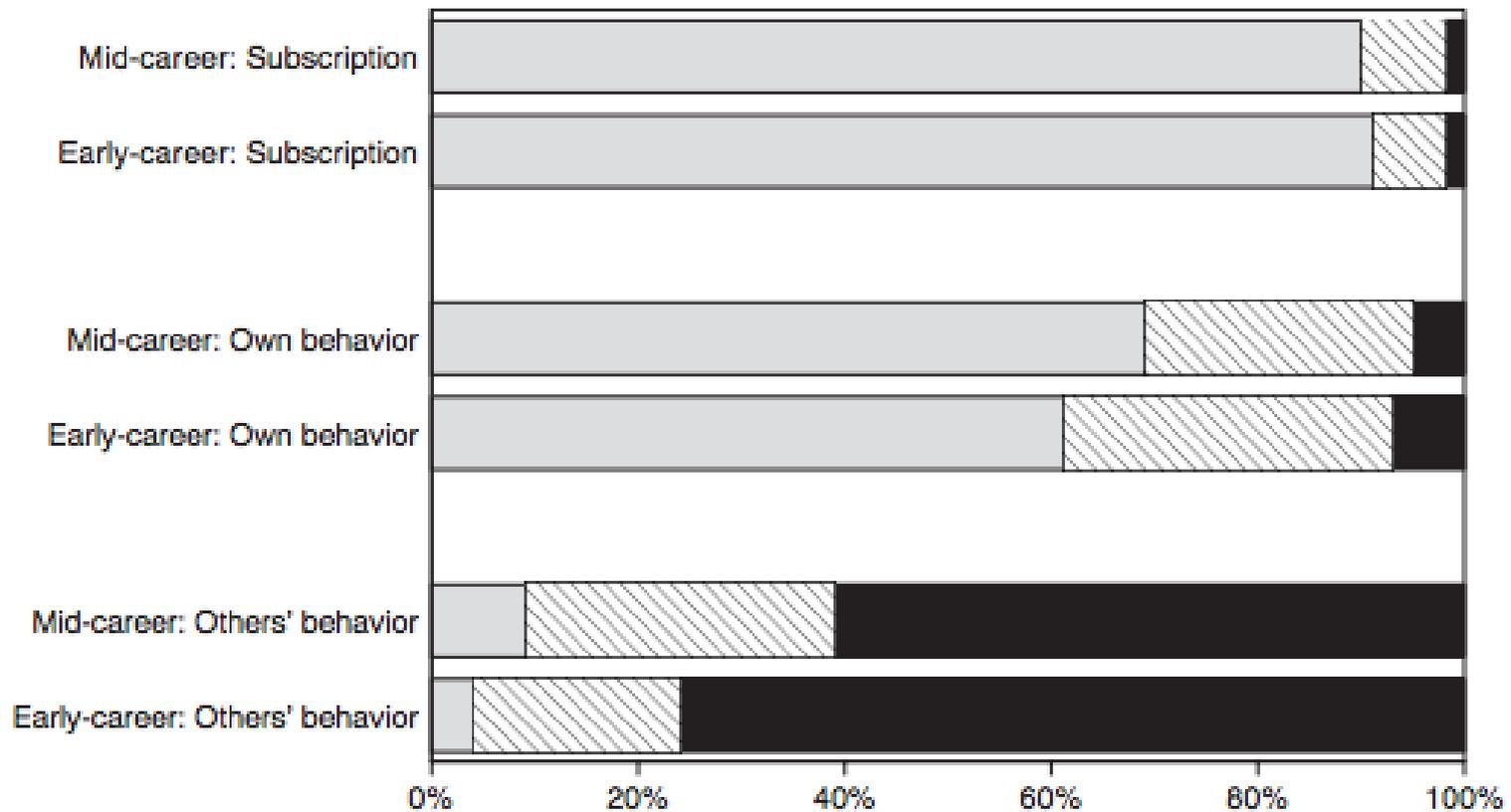


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).

WHY TRANSPARENCY?

“The credibility of scientific claims comes, in part, from the fact that their meaning is, at a minimum, available for other scholars to rigorously evaluate [...]

Such open access to the origins of others' claims is the hallmark of scientific ways of knowing.

Accordingly, when social scientists fail to document their assumptions, decisions, and actions [...] it limits others' ability to understand the meaning of scientists' claims.”

-Lupia & Elman, 2014, p. 20

TRANSPARENCY IS NECESSARY FOR CREDIBILITY



**“Nullius in Verba”
Take no one’s word**

**Science means
never having to say
“trust me”**

PART I: VALUES AND NORMS

**Open Science Advantage #1:
It's what makes science science**

TRANSPARENCY IS NECESSARY FOR CREDIBILITY



UNIVERSITY
of CALIFORNIA
PRESS

Collabra: Psychology



Start Submission

Reading: Quality Uncertainty Erodes Trust in Science

Share: [f](#) [t](#) [g+](#) [in](#)

Perspective/Opinion

Quality Uncertainty Erodes Trust in Science

Author: [Simine Vazire](#) 



*The Market for
“Lemons”: Quality
Uncertainty and the
Market Mechanism*

G. AKERLOF

PART I: VALUES AND NORMS

Open Science Advantage #1:

It's what makes science science

Open Science Advantage #2:

If we don't do it, others will stop trusting us

IS SCIENCE SELF-CORRECTING?



James Heathers

@jamesheathers

Following



"Science is self-correcting" - sure, *when we correct it*, not because of Magical Progress (tm).

RETWEETS

28

LIKES

76



10:57 AM - 25 Mar 2017

IS SCIENCE SELF-CORRECTING?

- Self-correction is an active process, and requires (Merton's norms):
 - Universalism: No hierarchy – anyone can challenge anyone else's finding
 - Communalism: Transparency – everything is out in the open
 - Disinterestedness: No conflicts of interest – truth is more important than personal success
 - Skepticism: Tearing things down is as valuable as finding new things

PART I: VALUES AND NORMS

Open Science Advantage #1:

It's what makes science science

Open Science Advantage #2:

If we don't do it, others will stop trusting us

Open Science Advantage #3:

It's necessary for self-correction

PART II: STATS REVIEW

PART I: Values and norms in science

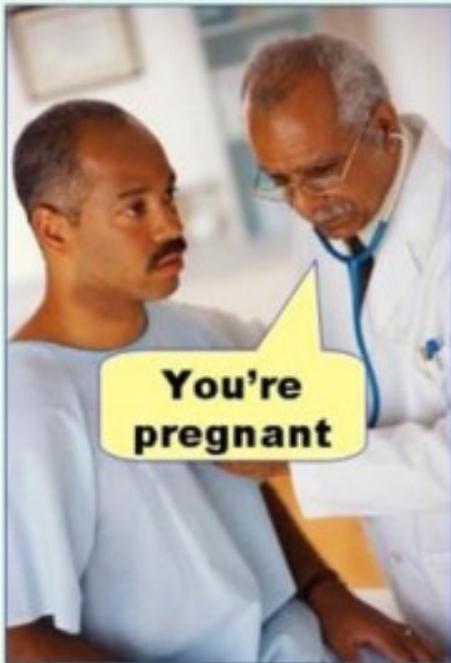
PART II: Stats review

PART III: Crisis

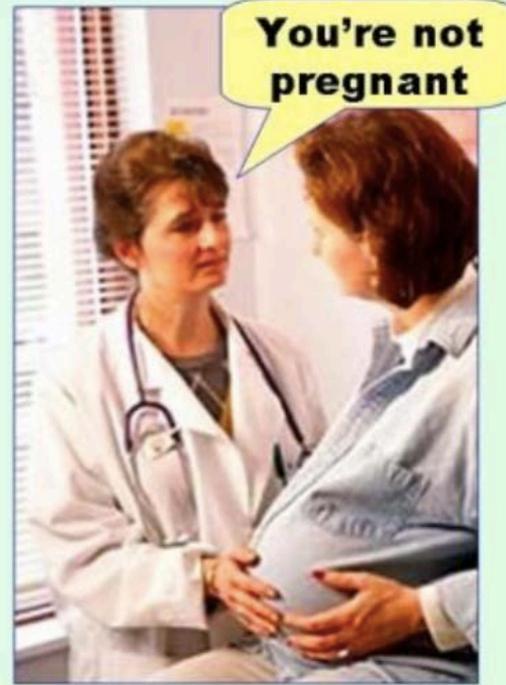
PART IV: Solutions

ERRORS IN SCIENCE

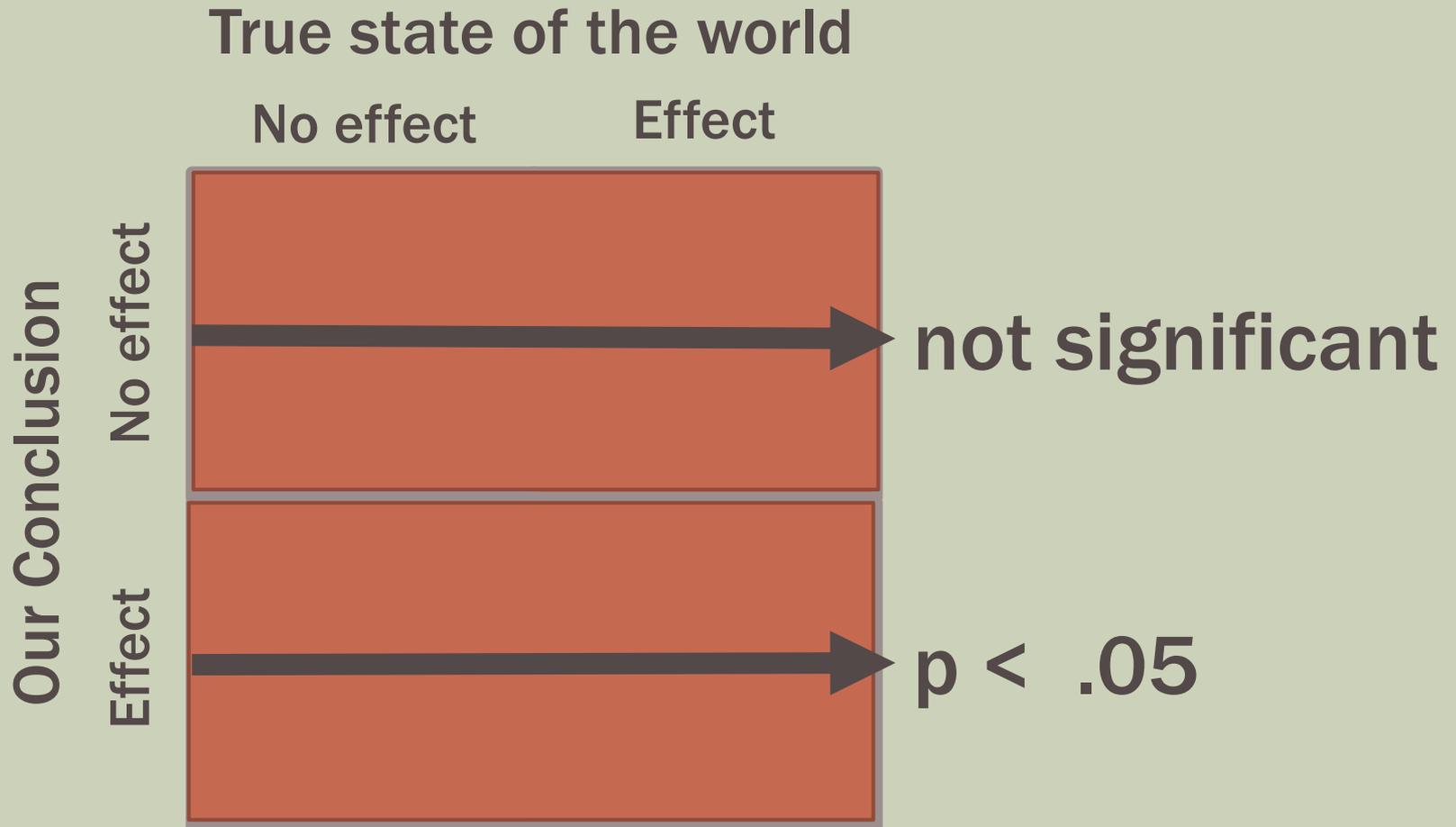
Type I error
(false positive)



Type II error
(false negative)



IS THERE AN EFFECT?



IS THERE AN EFFECT?

		True state of the world		
		No effect	Effect	
Our Conclusion	No effect	Correct Rejection		not significant
	Effect		Hit	$p < .05$

TWO KINDS OF ERRORS

		True state of the world		
		No effect	Effect	
Our Conclusion	No effect	Correct Rejection	False Negative	not significant
	Effect		Hit	$p < .05$

TWO KINDS OF ERRORS

		True state of the world		
		No effect	Effect	
Our Conclusion	No effect	Correct Rejection	False Negative	not significant
	Effect	False Positive	Hit	$p < .05$

FALSE POSITIVE RATE

		True state of the world	
		No effect	Effect
Our Conclusion	No effect	Correct Rejection 95%	False Negative
	Effect	False Positive 5%	Hit

WARNING!!

The Type I error rate is only 5%
if we follow the rules of NHST

The #1 rule of NHST is:
Make your prediction, then test it

or

“Hypotheses cannot be tested using the same data that
were used to generate the hypotheses in the first place”
-Wagenmakers/De Groot/Pierce

Open Science Advantage #4:

**Others can tell if you followed the rules of NHST
(and therefore, if your p-values are interpretable)**

PART III: CRISIS

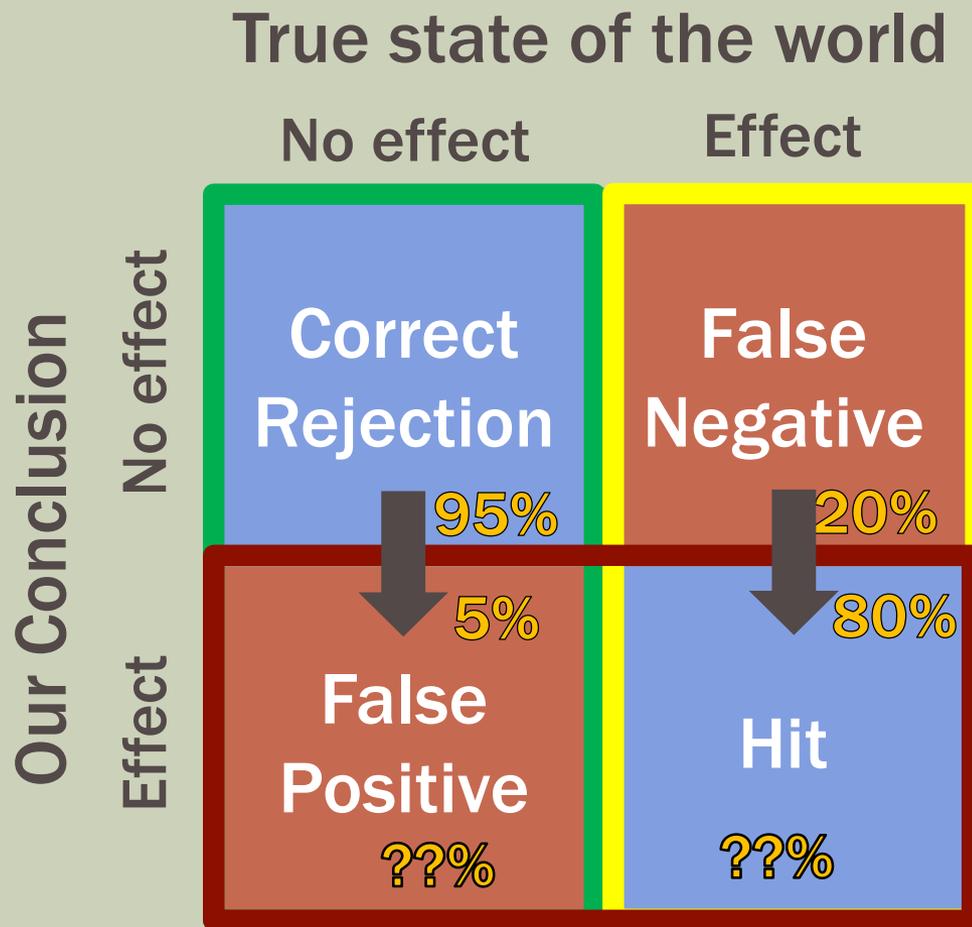
PART I: Values and norms in science

PART II: Stats review

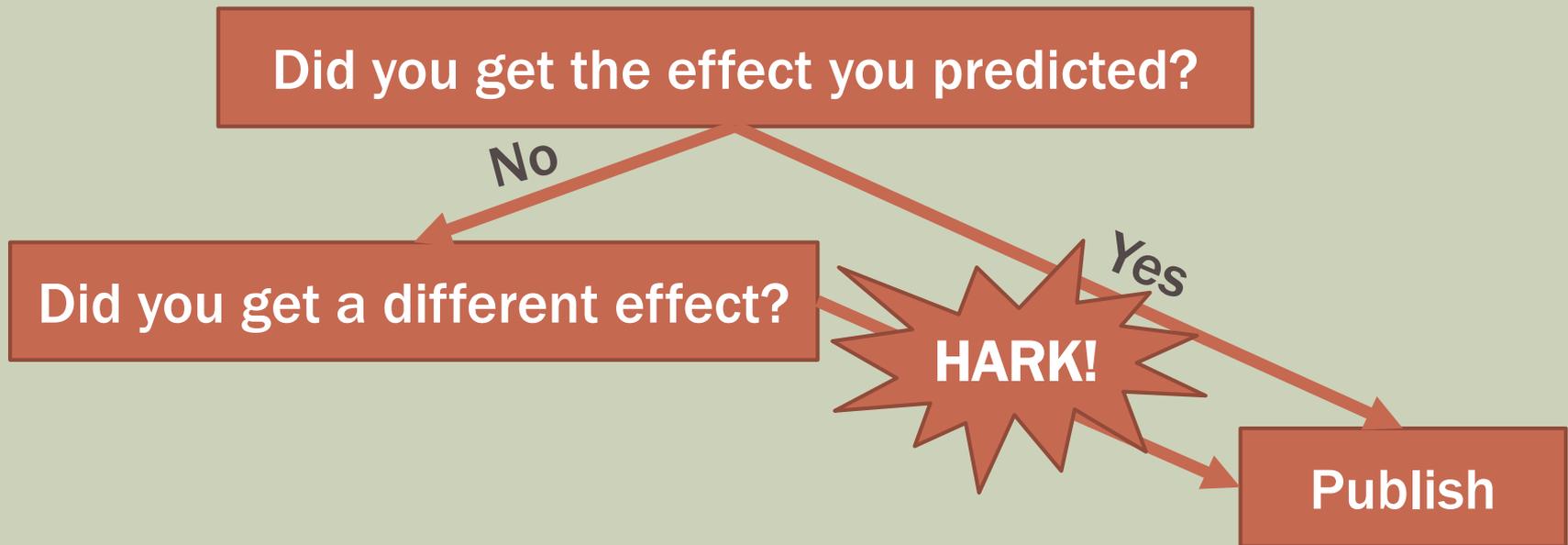
PART III: Crisis

PART IV: Solutions

WHAT HAPPENS WHEN WE DON'T FOLLOW THE RULES OF NHST?

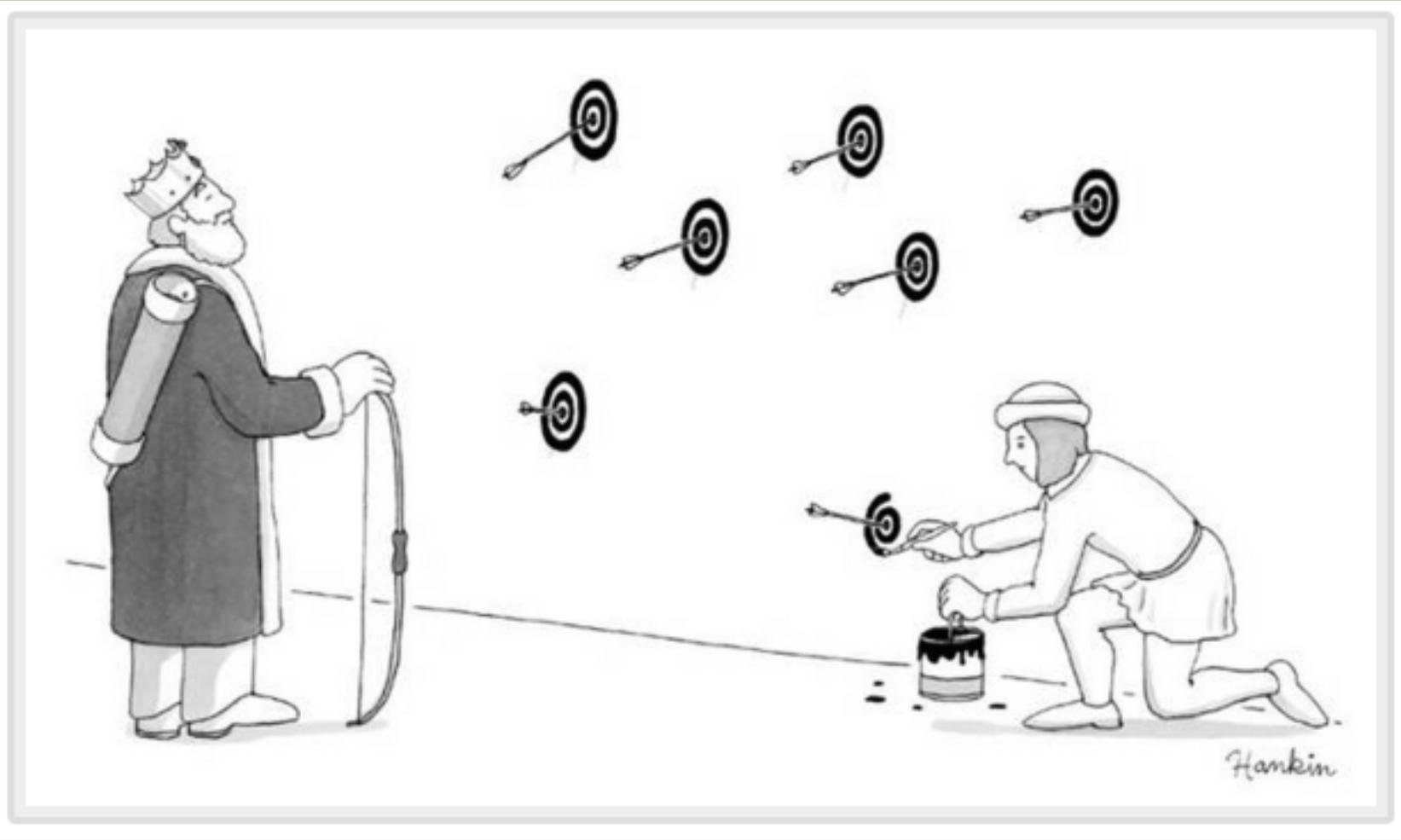


HOW TO TURN ANYTHING INTO A SIGNIFICANT RESULT



HARKING





HOW TO TURN ANYTHING INTO A SIGNIFICANT RESULT

Did you get the effect you predicted?

No

Did you get a different effect?

No

Can you make an effect significant?

Yes

HARK!

p-hack!

Publish

Six Ways to p-Hack

1. Stop collecting data once $p < .05$
2. Analyze many measures, but report only those with $p < .05$.
3. Collect and analyze many conditions, but only report those with $p < .05$.
4. Use covariates to get $p < .05$.
5. Exclude participants to get $p < .05$.
6. Transform the data to get $p < .05$.

H1: The U.S. economy is affected by whether Republicans or Democrats are in office

1 CHOOSE A
POLITICAL PARTY

Republicans

Democrats

H1: The U.S. economy is affected by whether Republicans or Democrats are in office

1 CHOOSE A POLITICAL PARTY

Republicans

Democrats

2 DEFINE TERMS

Which politicians do you want to include?

- Presidents
- Governors
- Senators
- Representatives

How do you want to measure economic performance?

- Employment
- Inflation
- GDP
- Stock prices

Other options

- Factor in power
Weight more powerful positions more heavily
- Exclude recessions
Don't include economic recessions

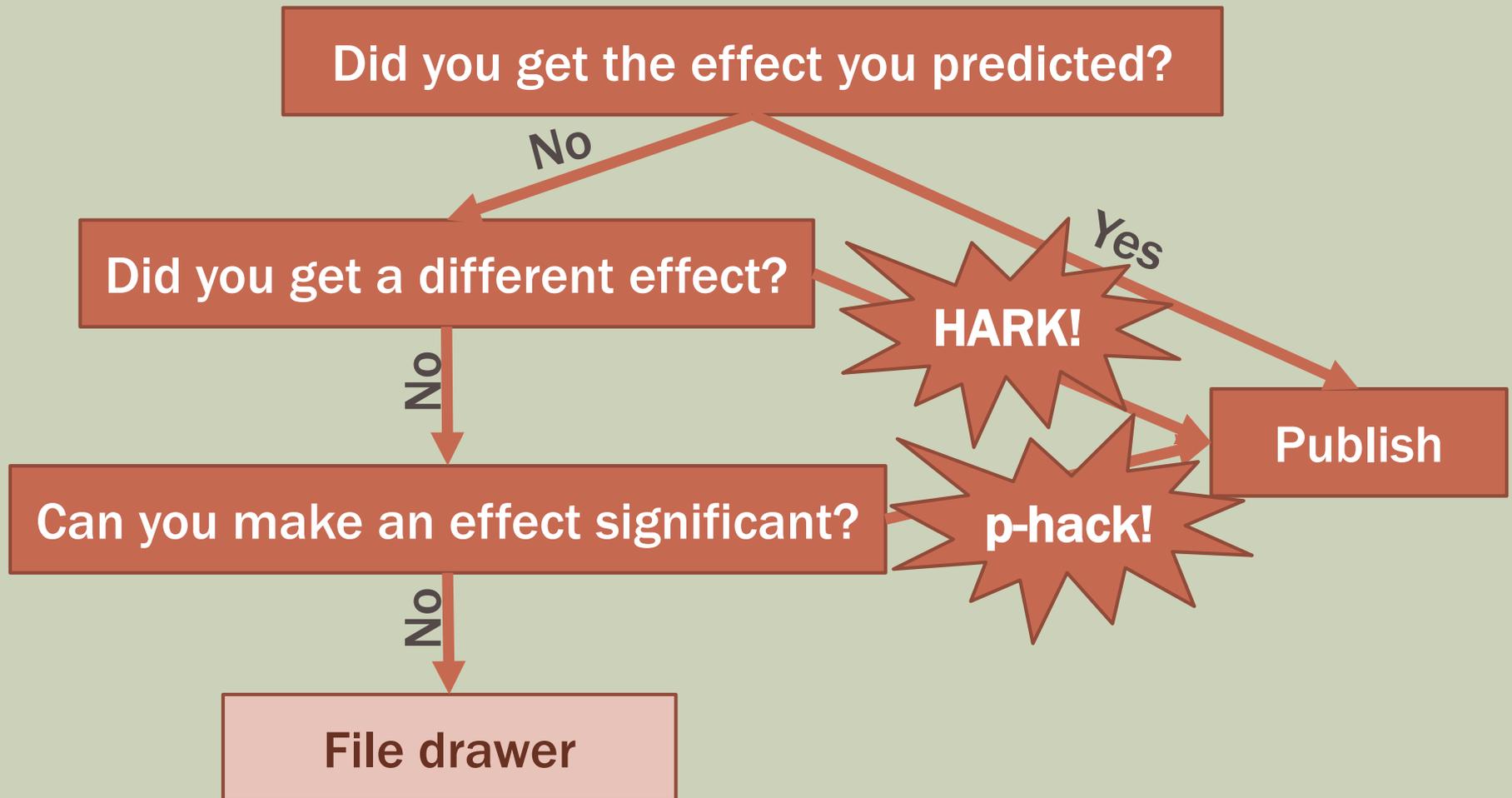
False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons¹, Leif D. Nelson², and Uri Simonsohn¹

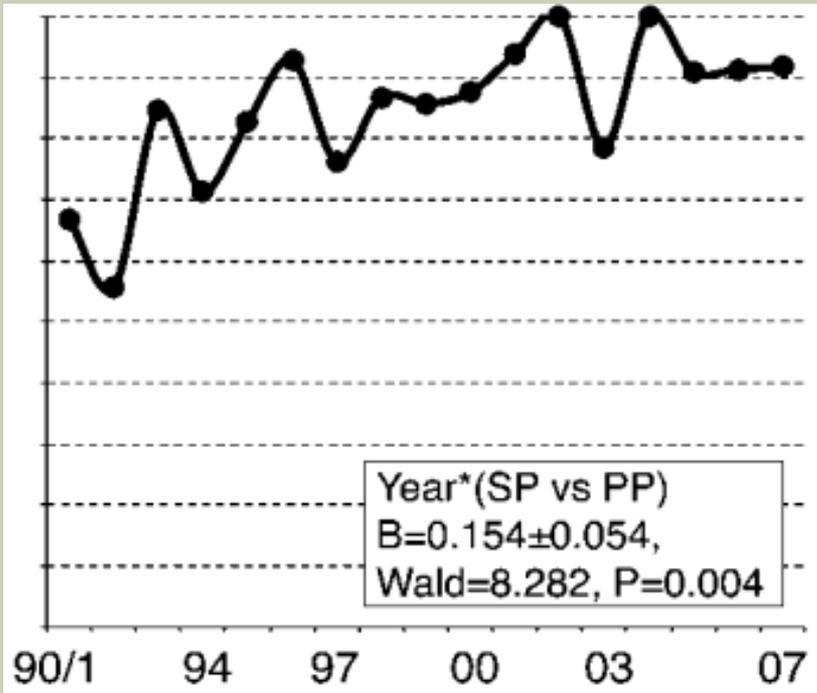
¹The Wharton School, University of Pennsylvania, and ²Haas School of Business, University of California, Berkeley

An ANCOVA revealed the predicted effect: According to their birth dates, people were nearly a year-and-a-half younger after listening to “When I’m Sixty-Four” (adjusted $M = 20.1$ years) rather than to “Kalimba” (adjusted $M = 21.5$ years), $F(1, 17) = 4.92, p = .040$.

HOW TO TURN ANYTHING INTO A SIGNIFICANT RESULT



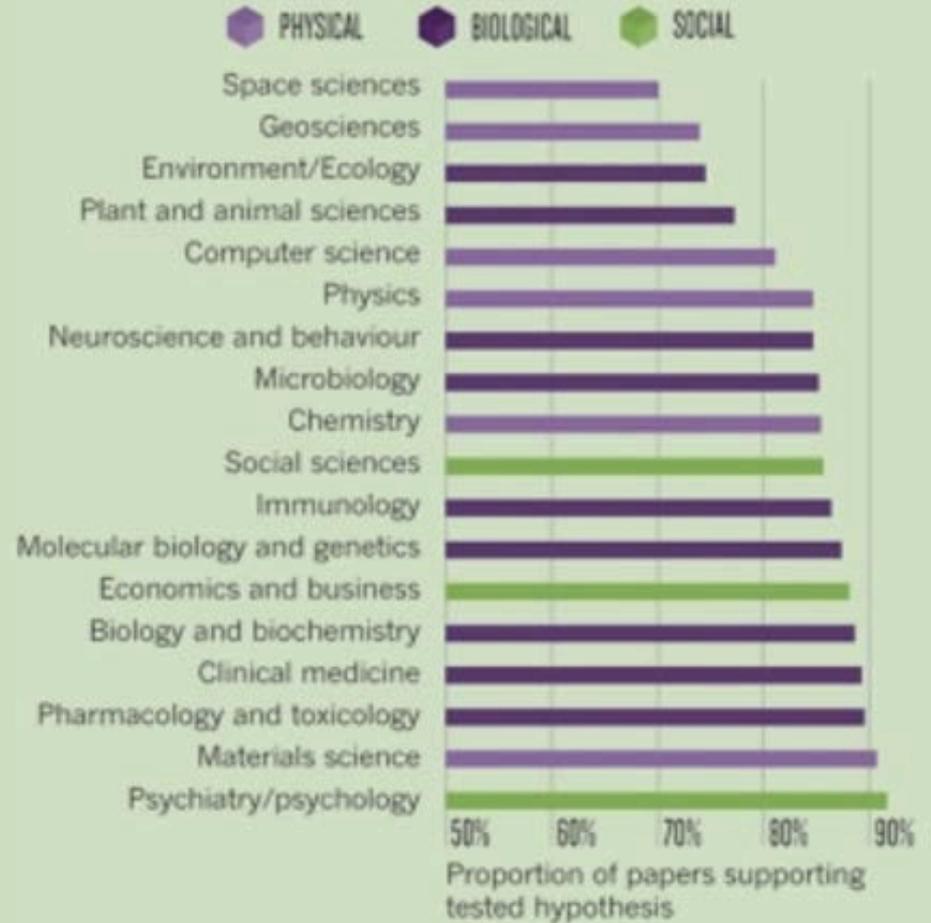
% significant (Psychology)



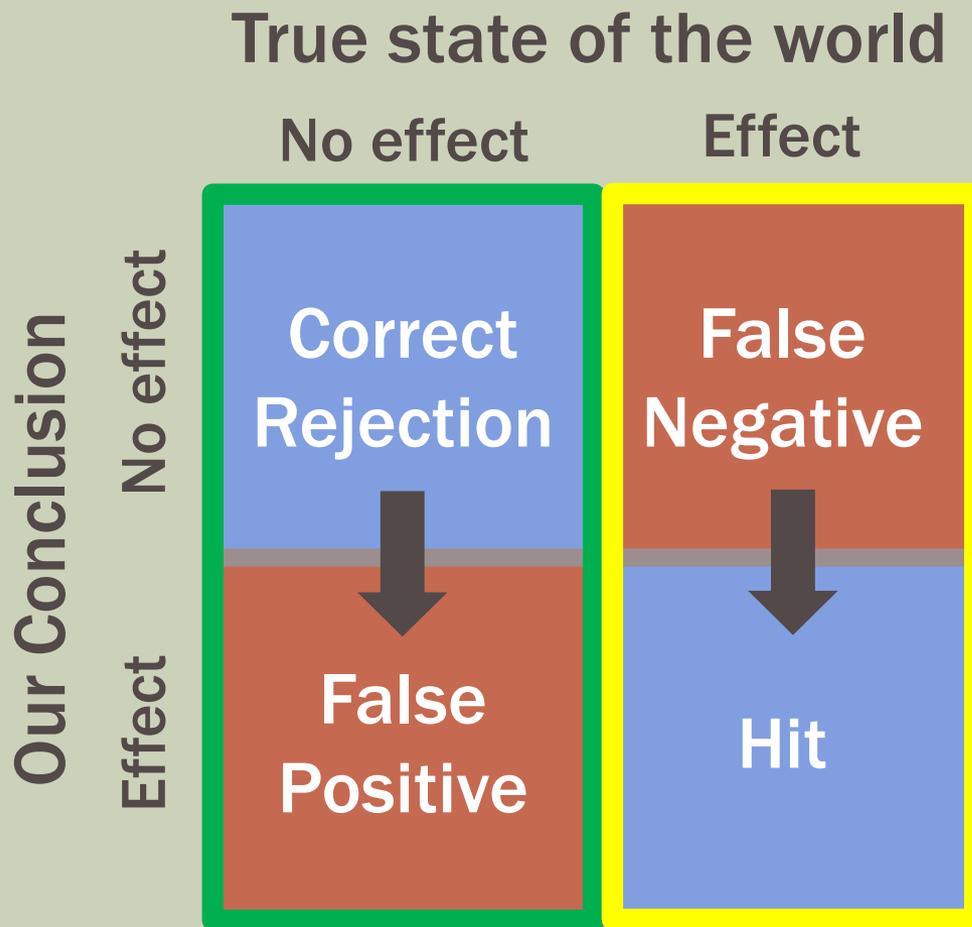
Fanelli (2012)

ACCENTUATE THE POSITIVE

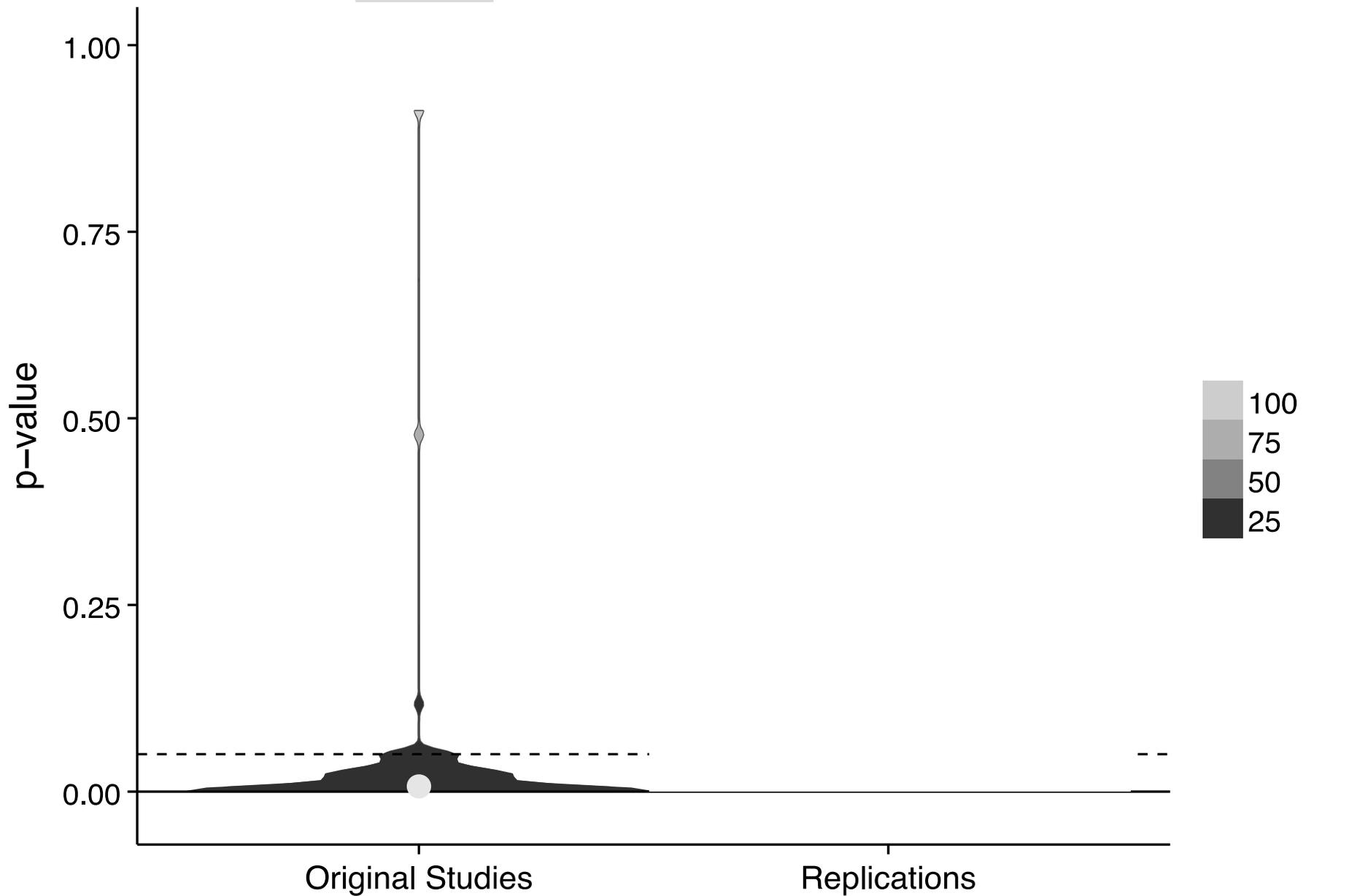
A literature analysis across disciplines reveals a tendency to publish only 'positive' studies — those that support the tested hypothesis. Psychiatry and psychology are the worst offenders.



WHAT HAPPENS WHEN WE DON'T FOLLOW THE RULES OF NHST?



97%



OPINION EXTRA

NATURE | NEWS



Let's

Cancer reproducibility project releases first results

An open-science effort to replicate dozens of cancer-biology studies is off to a confusing start.

Monya Baker & Elie Dolgin

18 January 2017



PDF



Rights & Permissions



cancer papers

By Jocelyn Kaiser | Jan. 18, 2017, 1:00 PM

Branden Camp



ROBE
Wash

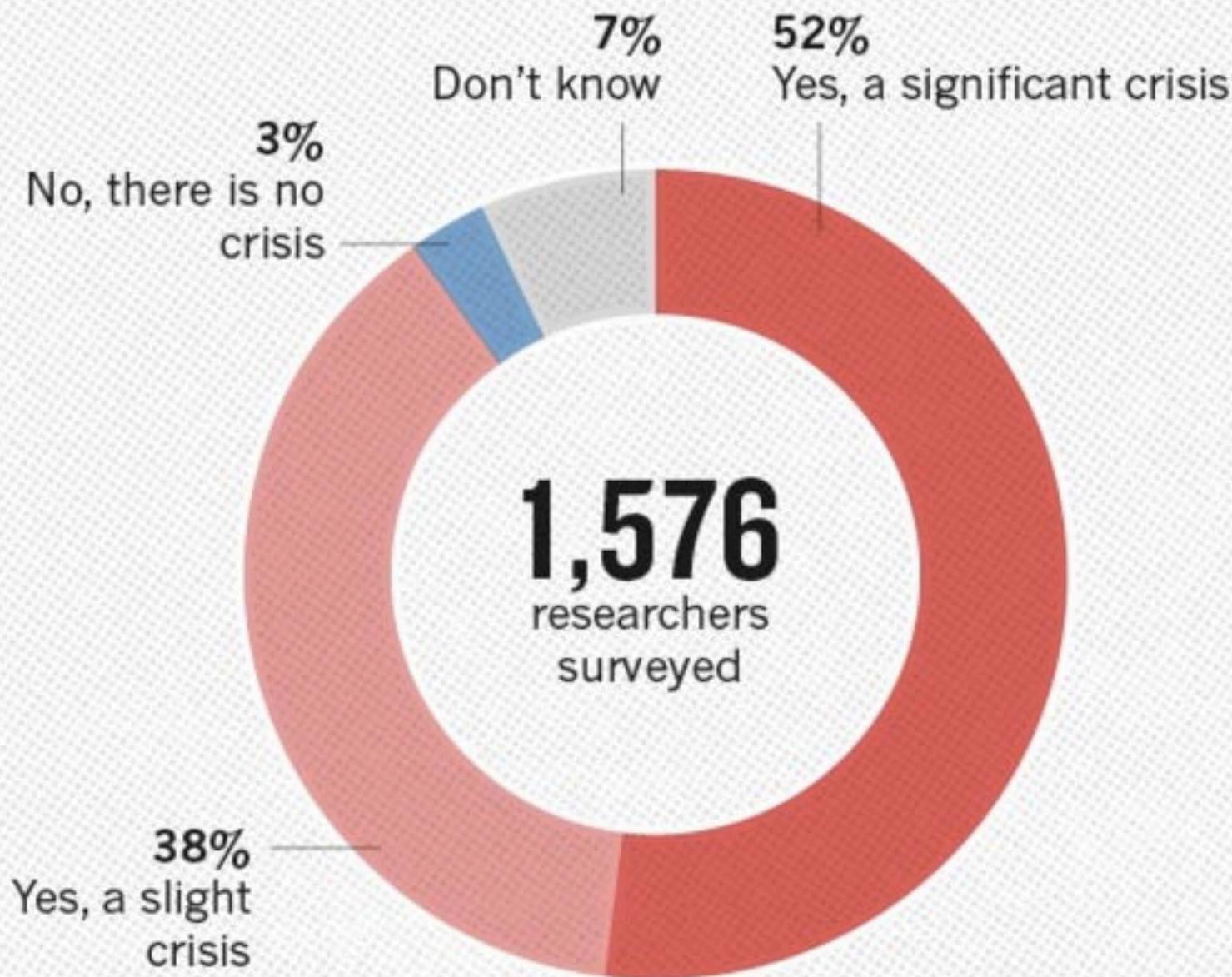


Iran



3
14

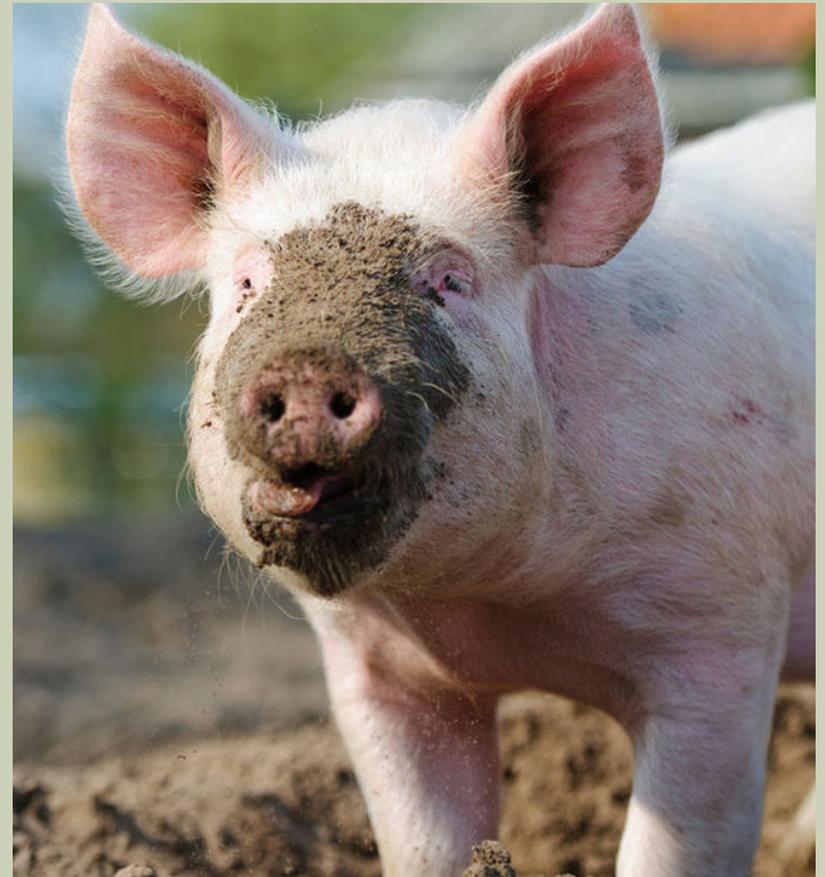
IS THERE A REPRODUCIBILITY CRISIS?



CRISIS: SUMMARY

What we've learned:

- 1.** Common research practices violate rules of NHST and increase the rate of false positives
- 2.** Many published findings are difficult or impossible to replicate



**Open Science Challenge #1:
If you don't p-hack/HARK,
your result will be less pretty.**

SUMMARY SO FAR

FOR open science:

- Our values say we should
- The rules of NHST say we should
- If we don't do it, others won't trust us
- It makes science (potentially) self-correcting

AGAINST open science:

- Your results will be less pretty



PART IV: SOLUTIONS

PART I: Values and norms in science

PART II: Stats review

PART III: Crisis

PART IV: Solutions

SOLUTION: ASSUME WE'RE HUMAN

We want to live up to Merton's norms

We're human

HOW TO KEEP HUMANS HONEST

Transparency:

- Making materials publicly available
- Sharing data
- Disclosing flexibility in data analysis
- Distinguish between what was predicted and what wasn't

WHAT IS PRE-REGISTRATION?

- So what is pre-registration?
 - 1. Determine your sample size, your manipulations, your measures, your analytic strategy, your critical hypothesis test.
 - 2. Write that down.
 - 3. Share it.
 - 4. Collect data.

Source: Leif Nelson's slides

HOW TO KEEP HUMANS HONEST

Transparency doesn't guarantee credibility, but transparency allows others to evaluate the credibility of your scientific claims.

HOW TO KEEP HUMANS HONEST

Transparency

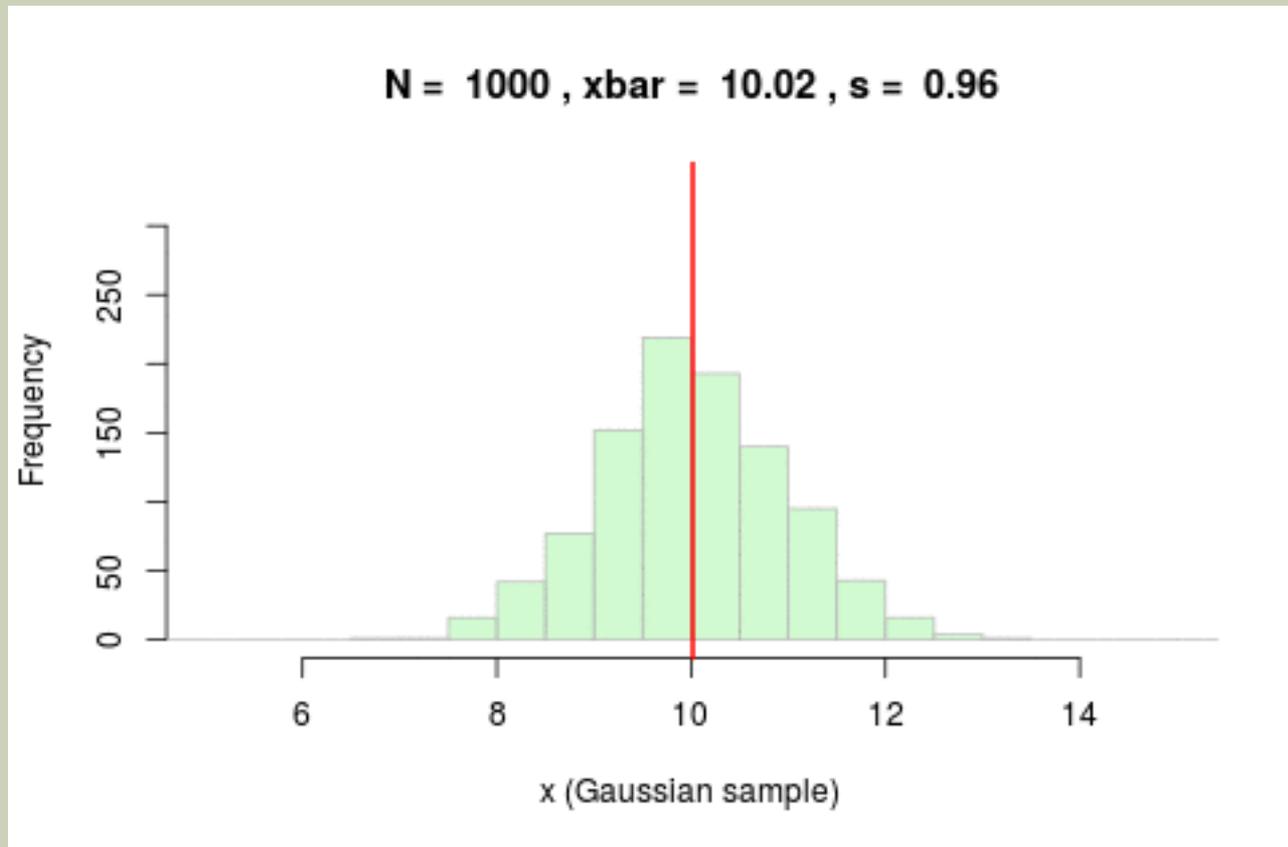
1. Open materials
2. Open data
3. Disclosure of flexibility
4. Pre-registration

High standards

1. Precision (large sample)



WHY ARE LARGE SAMPLES BETTER?



HOW TO KEEP HUMANS HONEST

Transparency

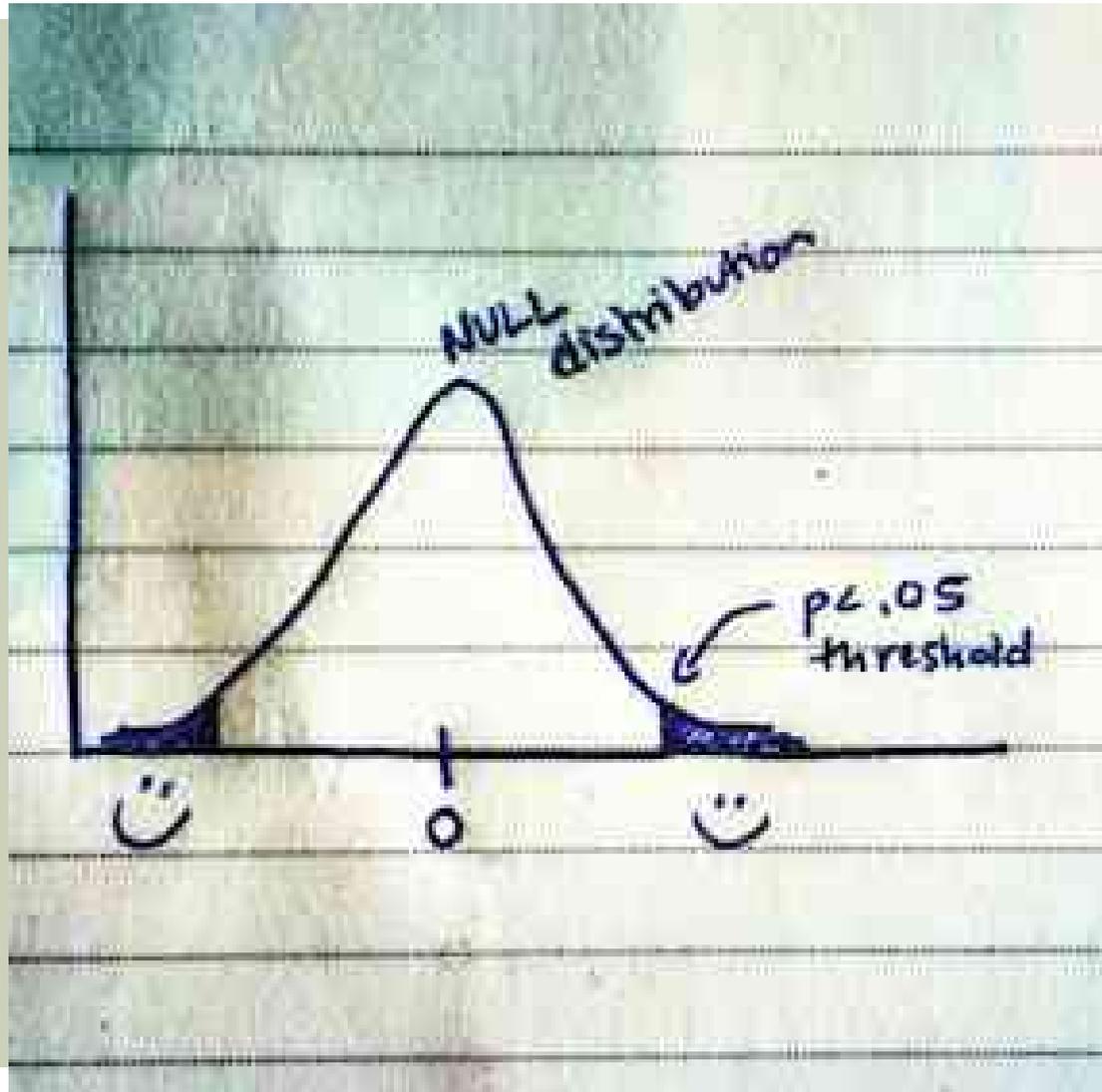
1. Open materials
2. Open data
3. Disclosure of flexibility
4. Pre-registration

High standards

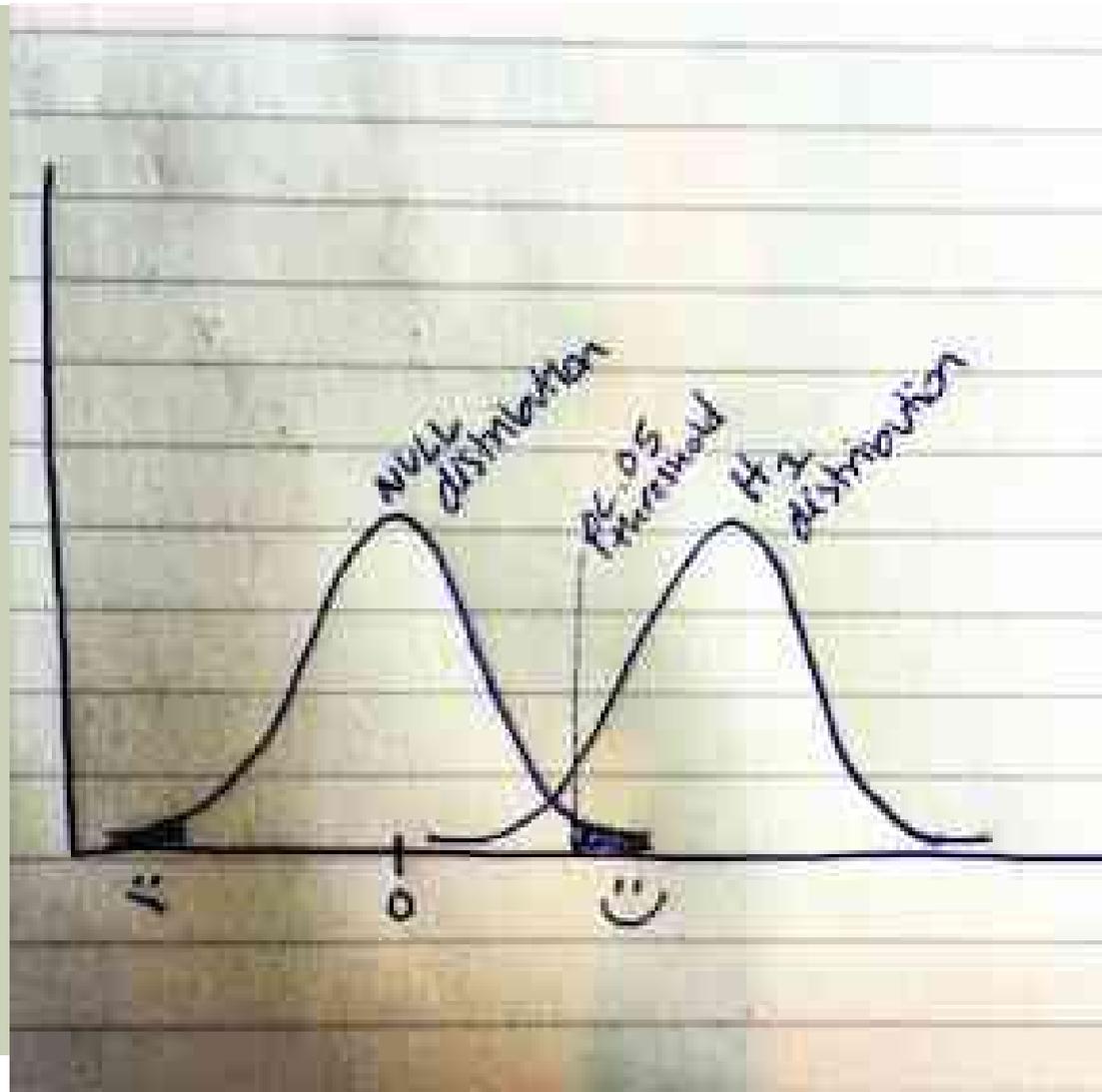
1. Precision (large sample)
2. Calibrated claims (small p-values)



WHY $p = .04$ SHOULD BE RARE



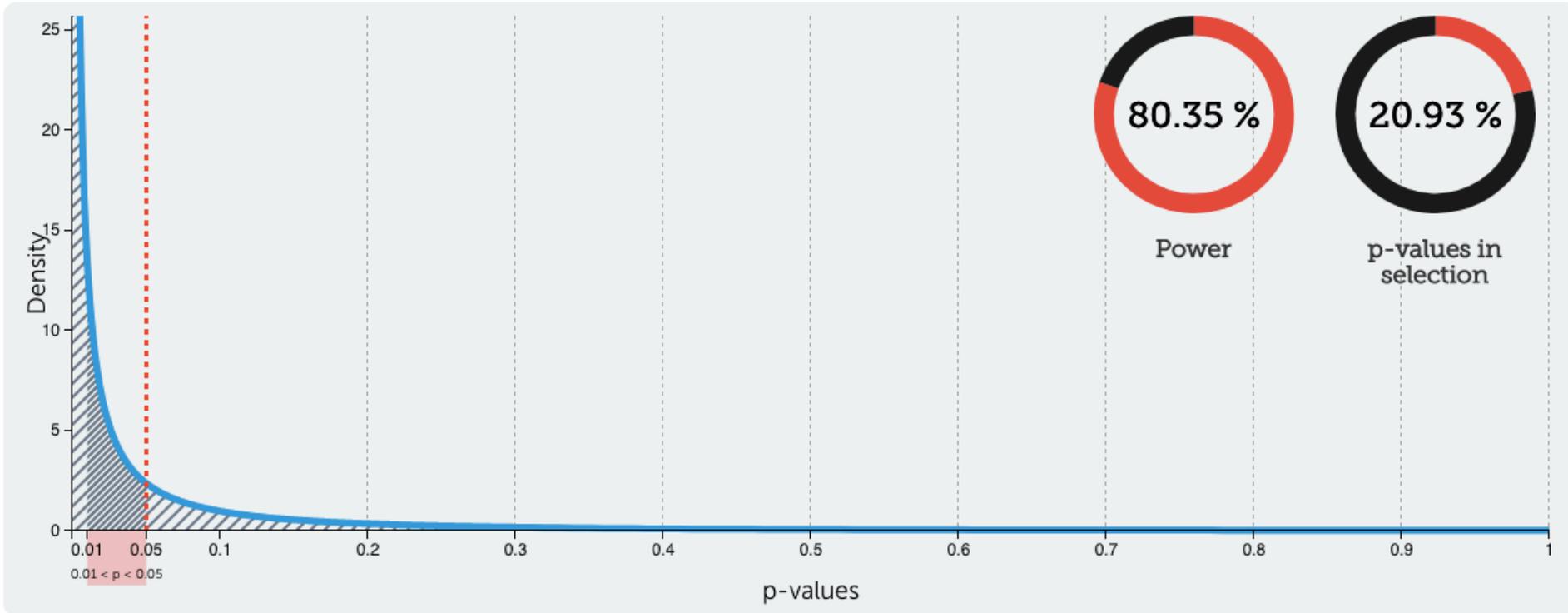
WHY $p = .04$ SHOULD BE RARE



Slide to change ES ($d = 0.4$)



Slide to change sample size ($n = 99$ per group)



<http://rpsychologist.com/d3/pdist/>

HOW TO KEEP HUMANS HONEST

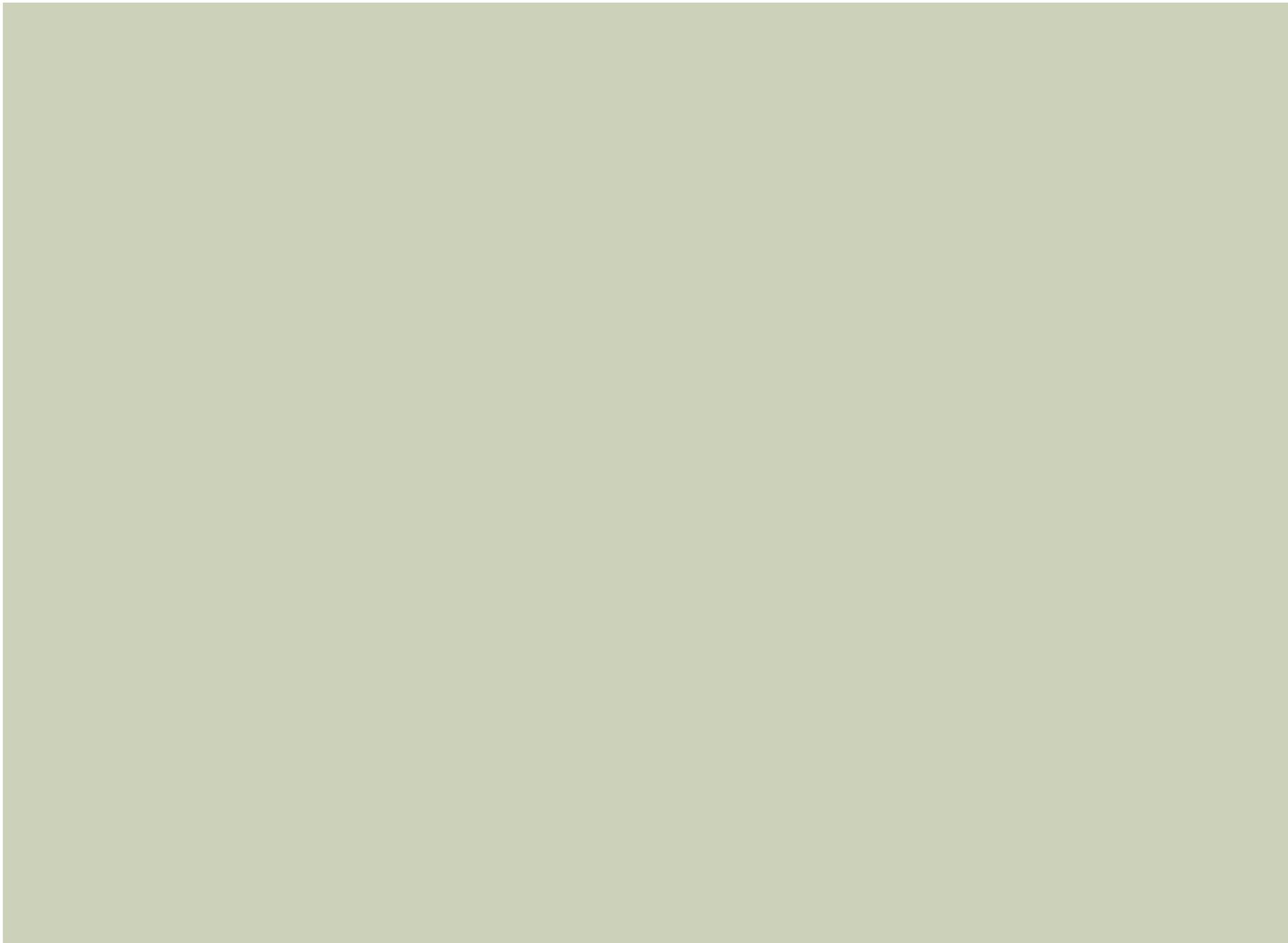
Transparency

1. Open materials
2. Open data
3. Disclosure of flexibility
4. Pre-registration

High standards

1. Precision (large sample)
2. Calibrated claims (small p-values)
3. Replication





SUMMARY

We want credibility.

**Open science is critical to credibility
but it doesn't guarantee credibility.**

**It allows others to evaluate your work
and assess whether your claims are credible.**

You also need high standards.

POST SCRIPT

- What individual researchers can do
 - Be transparent: Let others verify your claims
 - Be skeptical: Try to prove yourself wrong
- What gatekeepers can do
 - Journals, societies, universities, funding agencies
- Value credibility above: impact, status, novelty
 - Publish null results, boring results, replications, errata, critical commentaries
 - Be transparent (esp. about conflicts of interest)
 - Be accountable (PPPR, self-audits, etc.)



**Collabra:
Psychology**



statcheck



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN



SMU

psychology



CENTER FOR
OPEN SCIENCE



**SOCIETY FOR THE
IMPROVEMENT OF
PSYCHOLOGICAL SCIENCE**



Collabra: Psychology

The acceptance criterion for Collabra: Psychology is scientific, methodological, and ethical rigor. While Collabra: Psychology editors and reviewers do not attempt to predict a submission's impact to the field, nor employ any topic bias in accepting articles, they will check for **rigorously and transparently conducted, statistically sound, adequately powered, and fairly analyzed** research worthy of inclusion in the scholarly record. This is a focus on more objective acceptance criteria, and the bar is set high.

Collabra: Psychology supports the principles of Open Science, including a mandatory open data policy, and an option for authors to choose open peer review. We strongly encourage pre-registration of studies and pre-analysis plans.

SCIENCE IS HARD

The end

