THE CREDIBILITY REVOLUTION IN PSYCHOLOGICAL SCIENCE

Simine Vazire

Psychology

UC Davis

WHAT MAKES SCIENCE SCIENCE?

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

WHAT DOES A SELF-CORRECTING SYSTEM LOOK LIKE?

- **Universalism**
  - The validity of a scientific claim does not depend on who is making it.
  - No hierarchy. Status should not matter.
- **Communality**
  - 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.
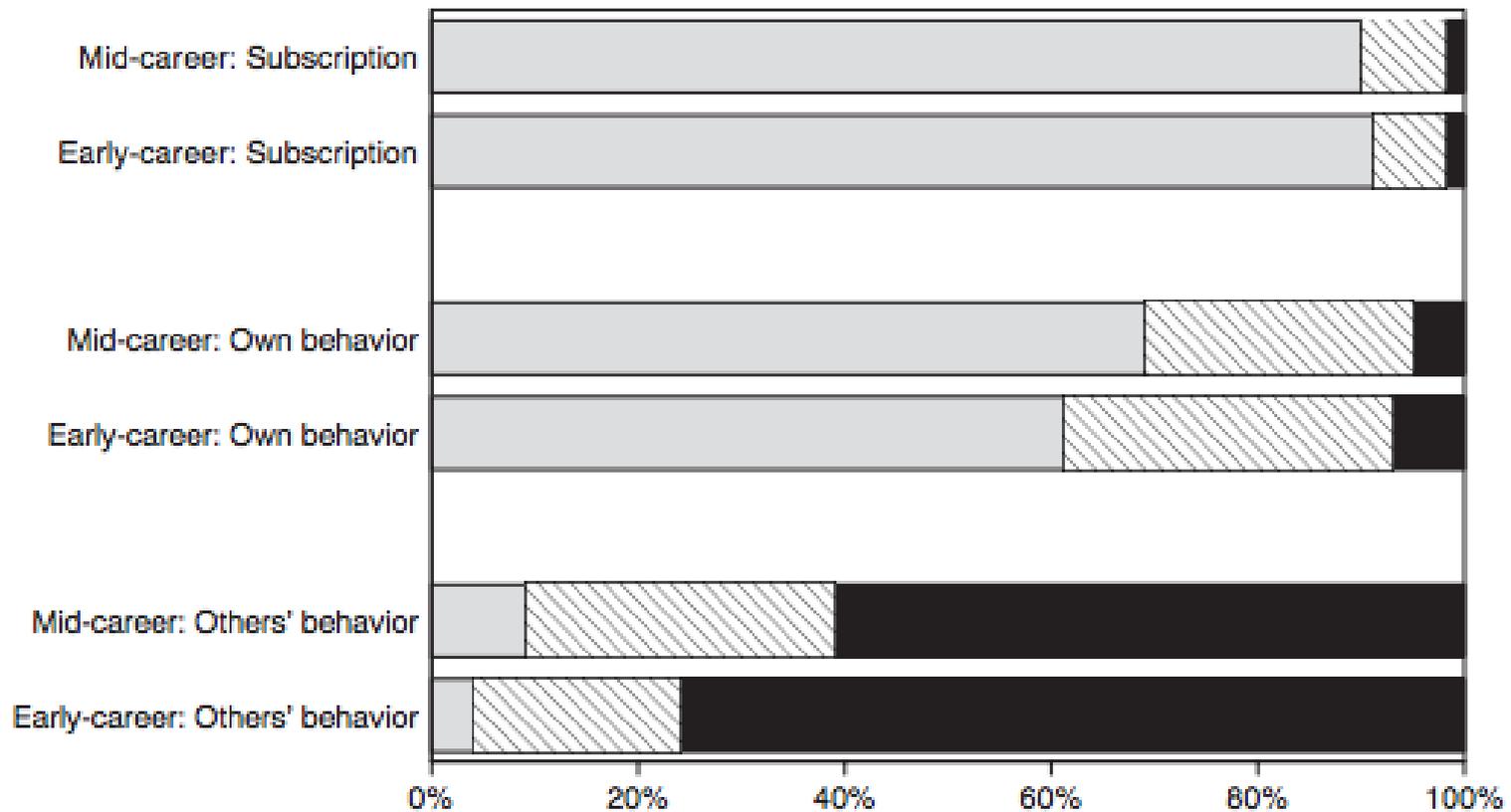
DO SCIENTISTS FOLLOW MERTON'S NORMS?

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

HOW ARE WE DOING?

Sanjay Srivastava
@hardsci

This is no shade at the replication researchers. But "science is self-correcting" will be an empty slogan if we cannot collectively get past the place where it takes 22 labs and >7,000 subjects to counter a study that nobody should have believed in the first place

THE ONE-TWO PUNCH

THE PROBLEM:

Common research practices violate rules of NHST and increase the rate of false positives
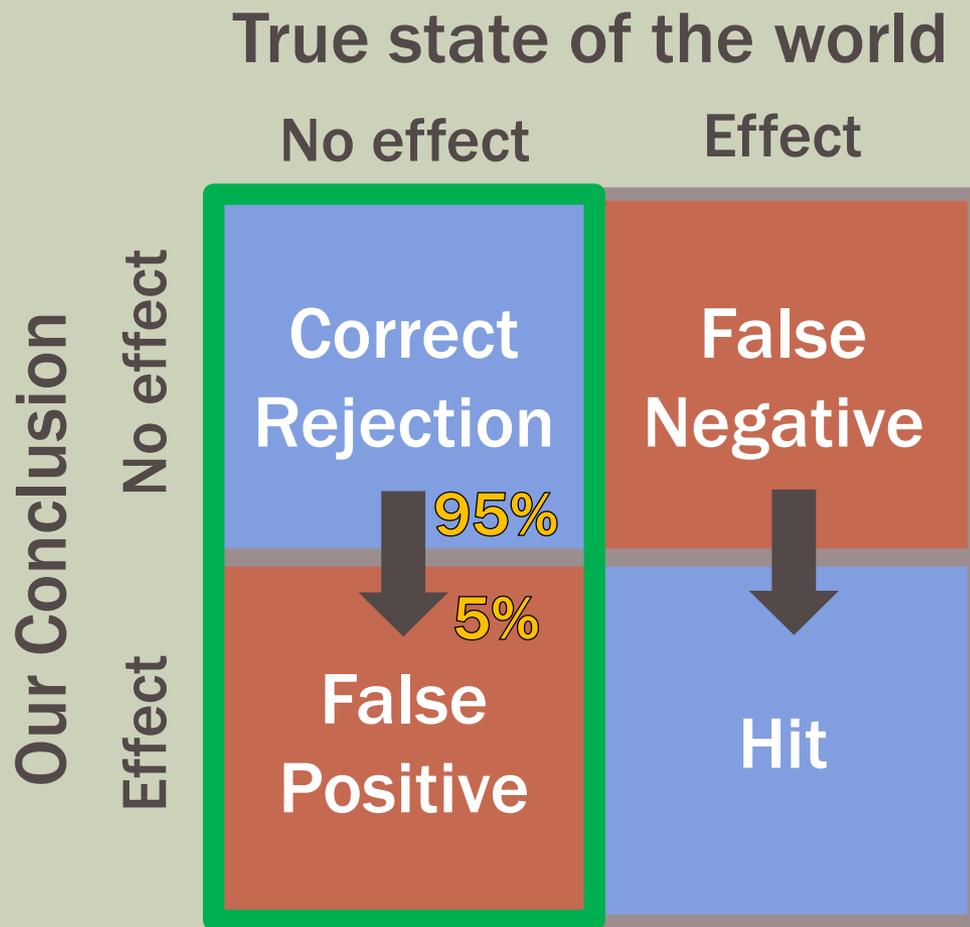
THE CONSEQUENCE:

The False Discovery Rate is unacceptably high

"Hypotheses cannot be tested using the same data that were used to generate the hypotheses in the first place"

-Wagenmakers/De Groot/Pierce

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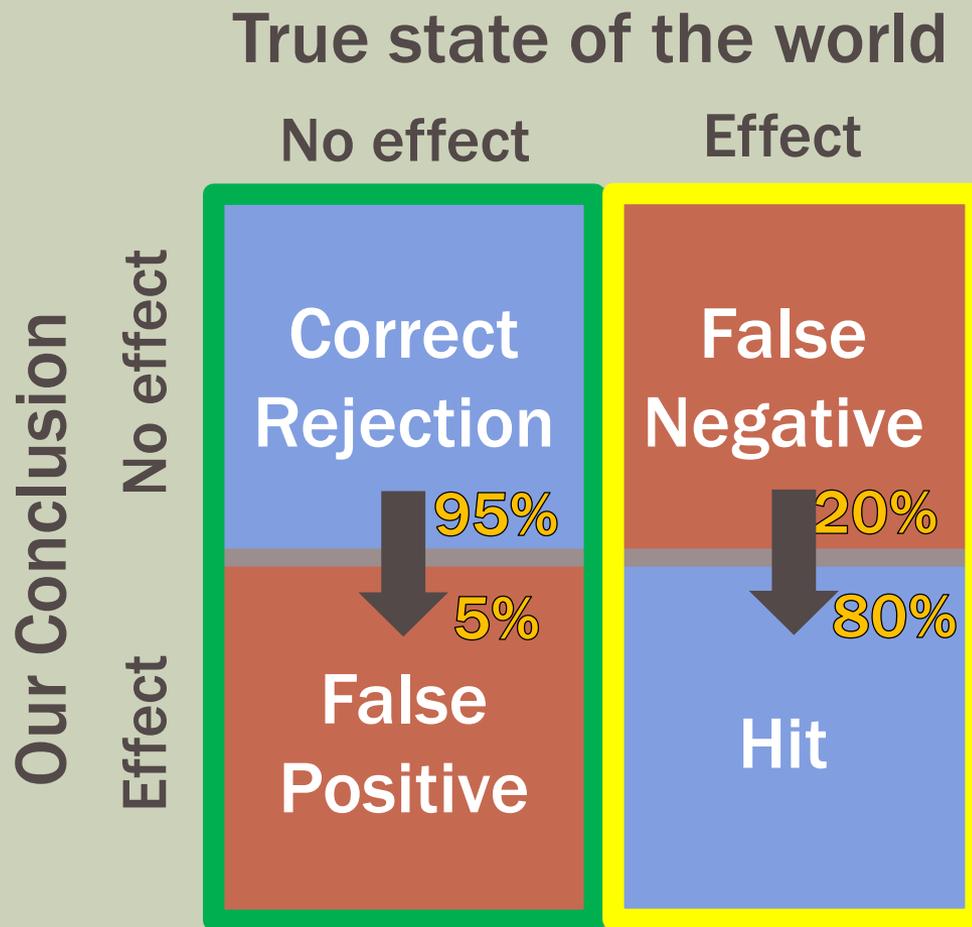
DON'T THEORIES CONSTRAIN US?

"[T]heories are so flexible that just about any comparison can be taken to be consistent with theory.
Remember sociologist Jeremy Freese's characterization of some hypotheses as '*more vampirical than empirical—unable to be killed by mere evidence*.'"

-Andrew Gelman, Feb 2018

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

THE ONE-TWO PUNCH

THE PROBLEM:

Common research practices violate rules of NHST and increase the rate of false positives

THE CONSEQUENCE:

The False Discovery Rate is unacceptably high

WHAT IS THE FALSE DISCOVERY RATE?

Across the social science:

- 39/100 in RP:P (Psychology)
- 11/18 in EERP (Economics)
- 10/13 in Many Labs 1 (Psychology)
- 14/28 in Many Labs 2 (Psychology)
- 3/10 in Many Labs 3 (Psychology)
- 13/21 in Science & Nature (Social Sciences)

= 87/190 = 46% replicability rate

= 54% false discovery rate

SO WHAT IS THE FALSE DISCOVERY RATE?

HYPOTHESES NOT SUPPORTED BY RESEARCH PAPERS (%)

Estimates from general literature **5–20%**

Registered reports for novel studies **55%***

Registered reports for replication studies **66%***

©nature

*Sample size: 296 hypotheses across 113 studies in biomedicine and psychology

SO WHAT IS THE FALSE DISCOVERY RATE?

40-60% ???

THE ONE-TWO PUNCH

THE PROBLEM:

Common research practices violate rules of NHST and increase the rate of false positives

THE CONSEQUENCE:

The False Discovery Rate is unacceptably high

An article [...] in a scientific publication is not the scholarship itself,
it is merely advertising of the scholarship.

-David Donoho (1998)

THE CREDIBILITY REVOLUTION

1. Transparency
2. Strong methods
3. Calibrated claims

TRANSPARENCY IS NECESSARY FOR CREDIBILITY
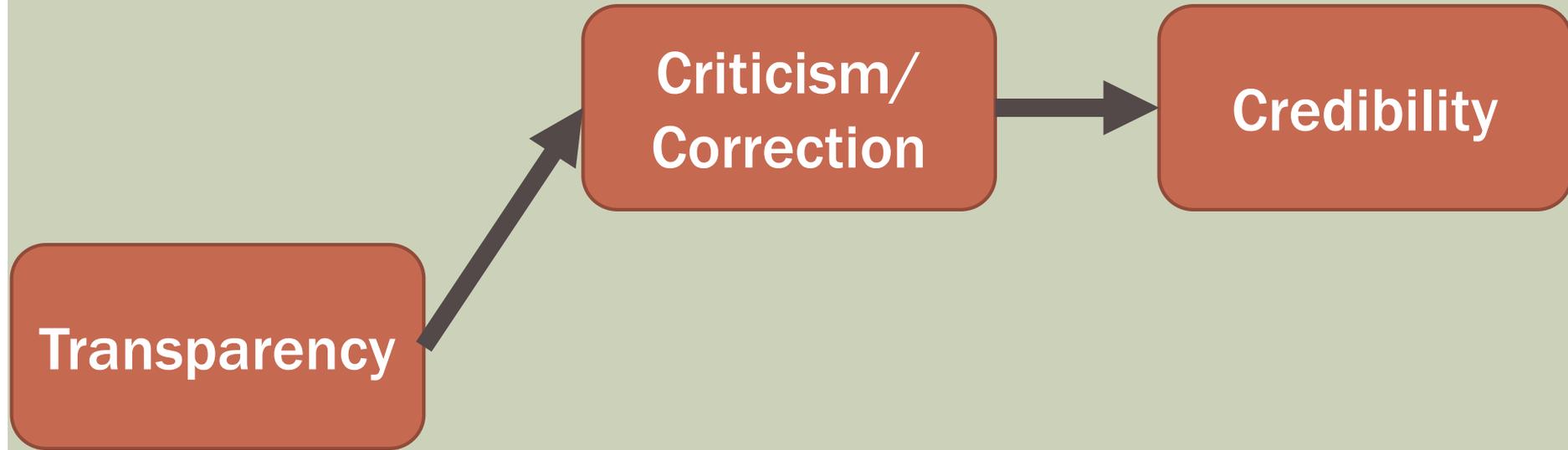

Perspective/Opinion

Quality Uncertainty Erodes Trust in Science

Author: Simine Vazire

The Market for "Lemons": Quality Uncertainty and the Market Mechanism
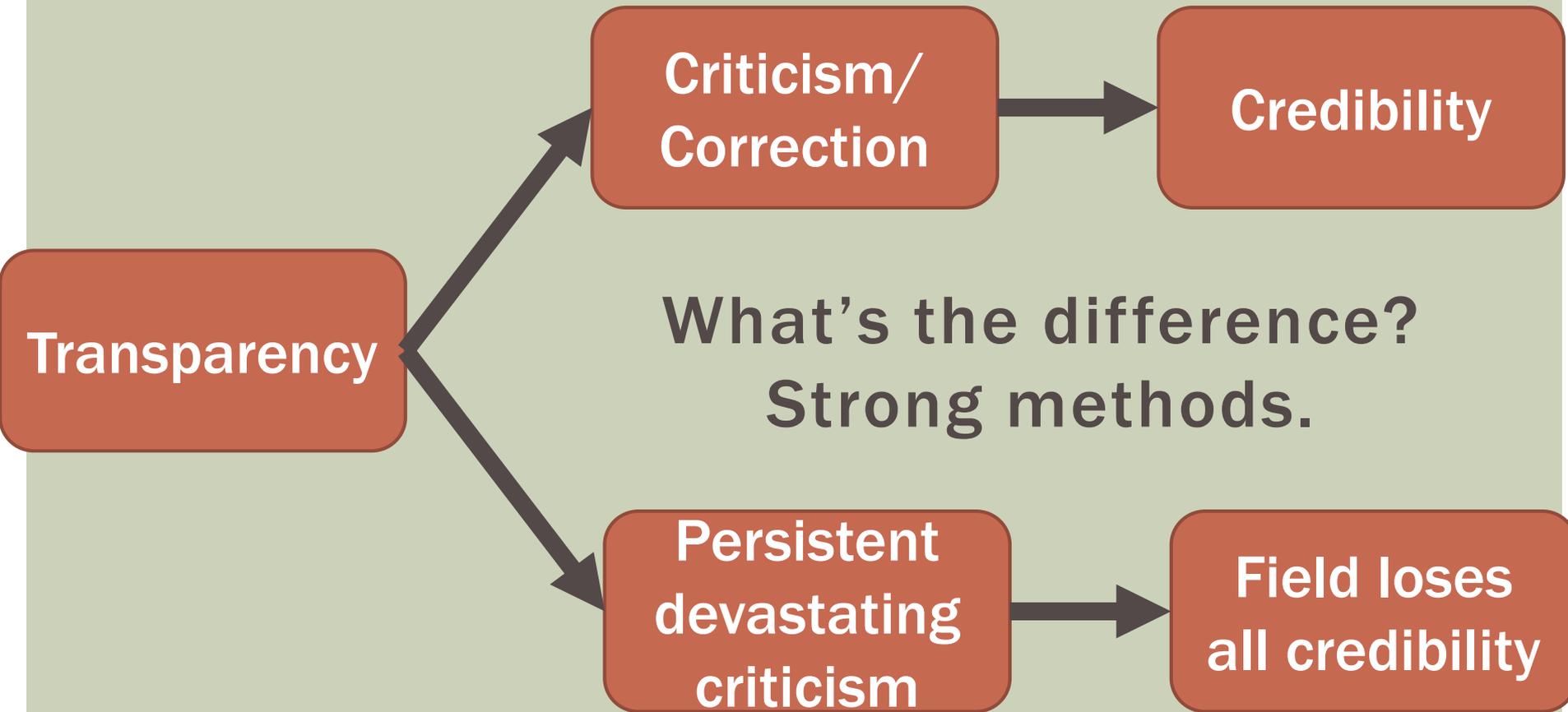
G. AKERLOF

THE CREDIBILITY REVOLUTION

THE CREDIBILITY REVOLUTION

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

Transparency gives our critics ammunition.

THE CREDIBILITY REVOLUTION

Transparency

Criticism/ Correction → Credibility

Persistent devastating criticism → Field loses all credibility

What's the difference? Strong methods.

Gelman (2017) Honesty and Transparency Are Not Enough

Consider the practical consequences for a researcher who eagerly accepts the message of ethical and practical values of sharing and openness, but does not learn about the importance of data quality.

He or she could then just be driving very carefully and very efficiently into a brick wall, conducting transparent experiment after transparent experiment and continuing to produce and publish noise.

The openness of the work may make it easier for a later researcher to attempt—and fail—to replicate the resulting published claims, but little if any useful empirical science will be done by anyone concerned.

I do not think we are doing anybody any favors by having them work more openly using data that are inadequate to the task.

THE CREDIBILITY REVOLUTION

1. **Transparency**
2. **Strong methods**

THE CREDIBILITY REVOLUTION

Strong methods:

- Research methods 101
- Precision (large sample)
- Replication
- Should produce a consistent pattern of results (mostly small p-values)

THE CREDIBILITY REVOLUTION

1. **Transparency**

2. **Strong methods**

3. **Calibrated claims**

WHERE TO GO FROM HERE?

As individuals: An oath for scientists

THE PROBLEM IS BIGGER THAN INDIVIDUALS

WHAT STANDS IN THE WAY OF PRIORITIZING CREDIBILITY?

- Status
- Secrecy
- Self-interest
- Dogma

PRIORITIZING CREDIBILITY

PAPER A	PAPER B
Discloses measures/studies	No disclosures
Shares materials	No materials
Shares data	No data
Shares pre-analysis plan	No pre-analysis plan

It must be ok to criticize Paper A for having flaws despite extreme transparency

It must be ok to criticize Paper B for lacking transparency despite obvious flaws

PRIORITIZING CREDIBILITY

When evaluating scientific claims:

- Blind yourself to authors and institutions
- Avoid and disclose conflicts of interest
- Ask for the information you need
- Use this information, transparency is for accountability!
- Ask for evidence of robustness, and calibrated claims
- Tolerate uncertainty and messiness when necessary
- Value incremental contributions
- Value corrections and critiques

WHO WILL WATCH THE WATCHERS?

Journals, editors, and societies enjoy:

- Few consequences for publishing low credibility science
- Monopolies on prestige

Solutions

- More accountability for gatekeepers
  - Public discussion and criticism – call out journal/editor, too
  - More metrics/rankings
- Get rid of gatekeepers
  - Preprints, open review, Plaudit

WHO WILL WATCH THE WATCHERS?

IS PEER REVIEW WORTH IT?

- Peer review cannot live up to its reputation – gives false sense of security
- Truth in advertising: "Between 1 and 5 scientists thought this paper was ok"
- Too negative, but not negative in the right ways
- Conflicts of interest, commissioned articles, and status bias
- Editors motivated to chase impact and popularity
- No obligation to self-correct
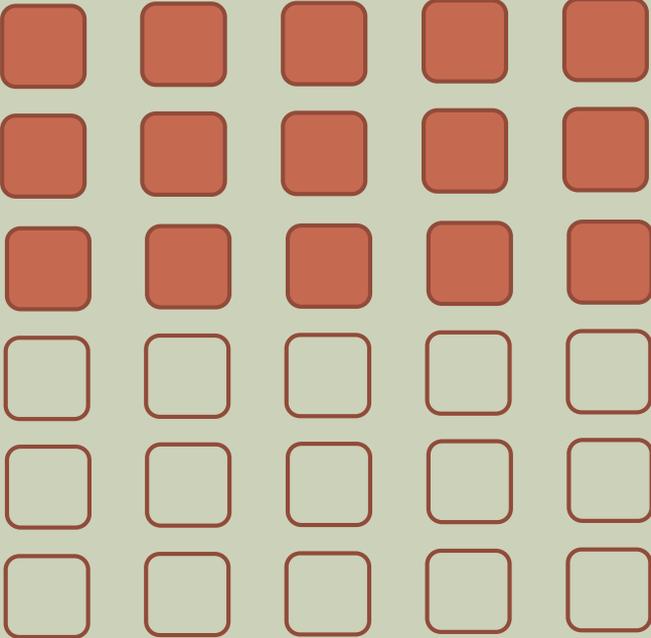
PRIORITIZING CREDIBILITY

Researcher A

Total N: 7,500
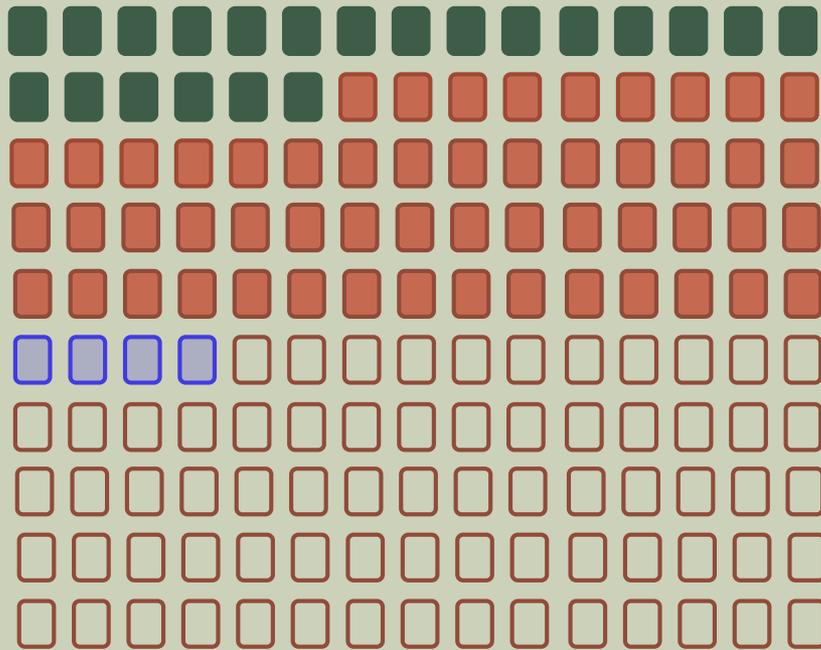Sample size: 50
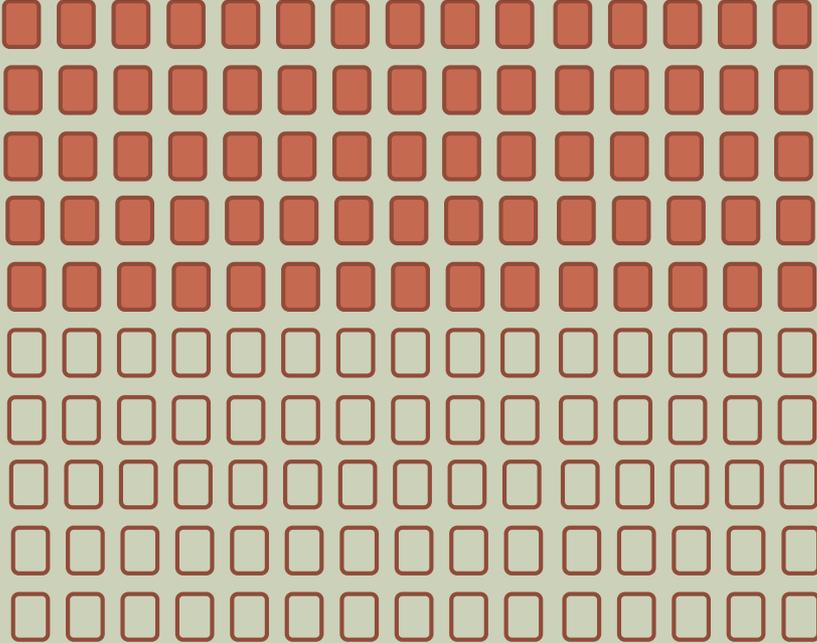
Researcher B

Total N: 7,500
Sample size: 250

Gervais et al., SPPS, 2015

Researcher A

Total N: 7,500
Sample size: 50

true positive

false positive
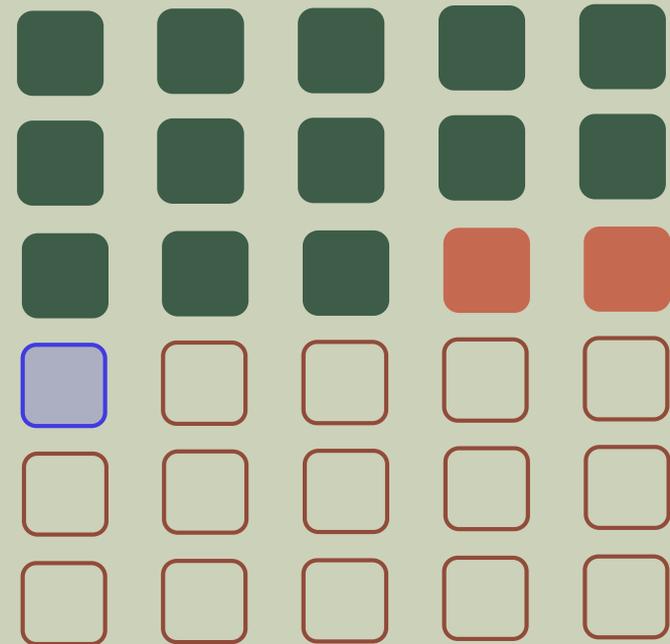
16% of positive results are false positives

28% statistical power

83% of results are negative (file-drawered)

Researcher B

Total N: 7,500
Sample size: 250
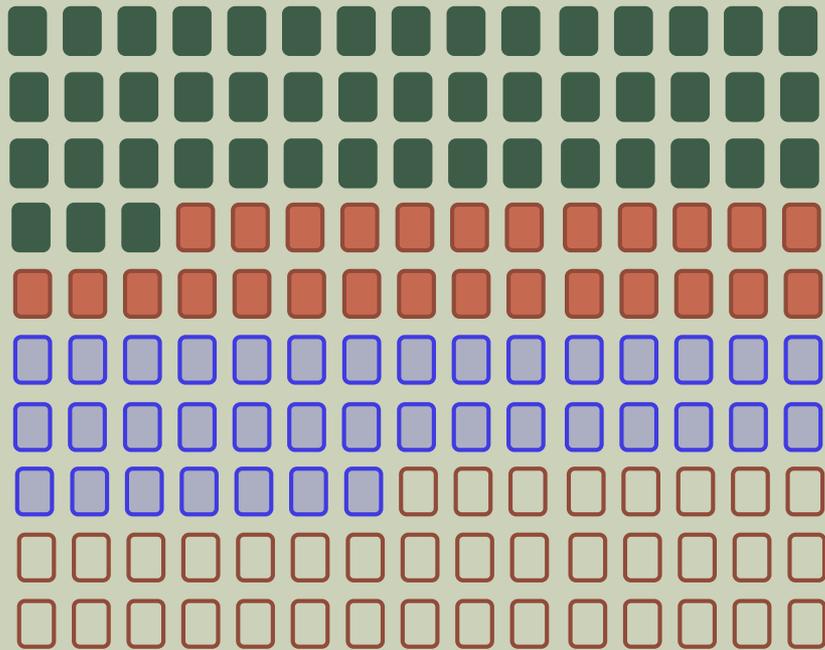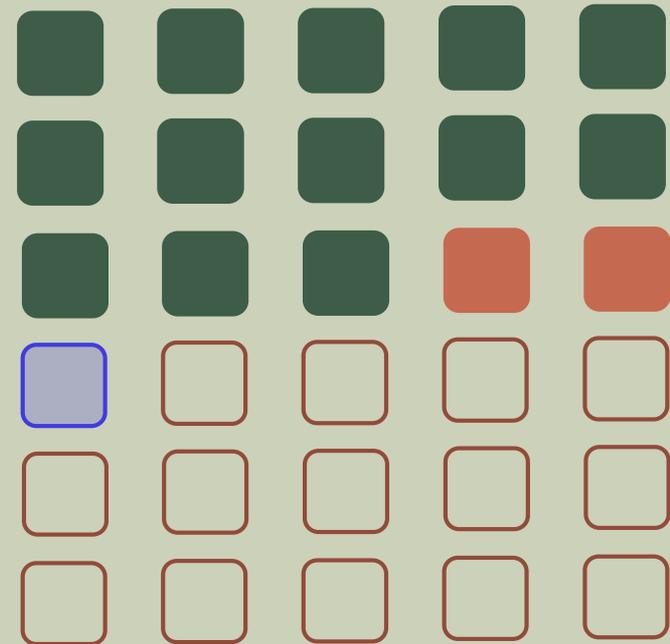
5% of positive results are false positives

88% statistical power

53% of results are negative (file-drawered)

Gervais et al., SPPS, 2015

Researcher A
Total N: 7,500
Sample size: 50

Researcher B
Total N: 7,500
Sample size: 250

true positive

false positive

44% of positive results are false positives

64% statistical power

43% of results are negative (file-drawered)

5% of positive results are false positives

88% statistical power

53% of results are negative (file-drawered)

Gervais et al., SPPS, 2015

THE CREDIBILITY REVOLUTION

The end