TOWARDS A MORE SELF-CORRECTING SCIENCE

Simine Vazire

Psychology

UC Davis

ORCID: 0000-0002-3933-9752

WHAT ARE THE CORE VALUES OF SCIENCE?

The goal of science is not to never make errors.
The goal is to minimize unforced errors,
detect unavoidable errors as soon as possible,
and calibrate our claims.

"Science is self-correcting"

99+%	Trust these
80%	
60%	Verify or correct these
40%	
🫣	

James Heathers
@jamesheathers

Following

"Science is self-correcting" - sure, *when we correct it*, not because of Magical Progress (tm).
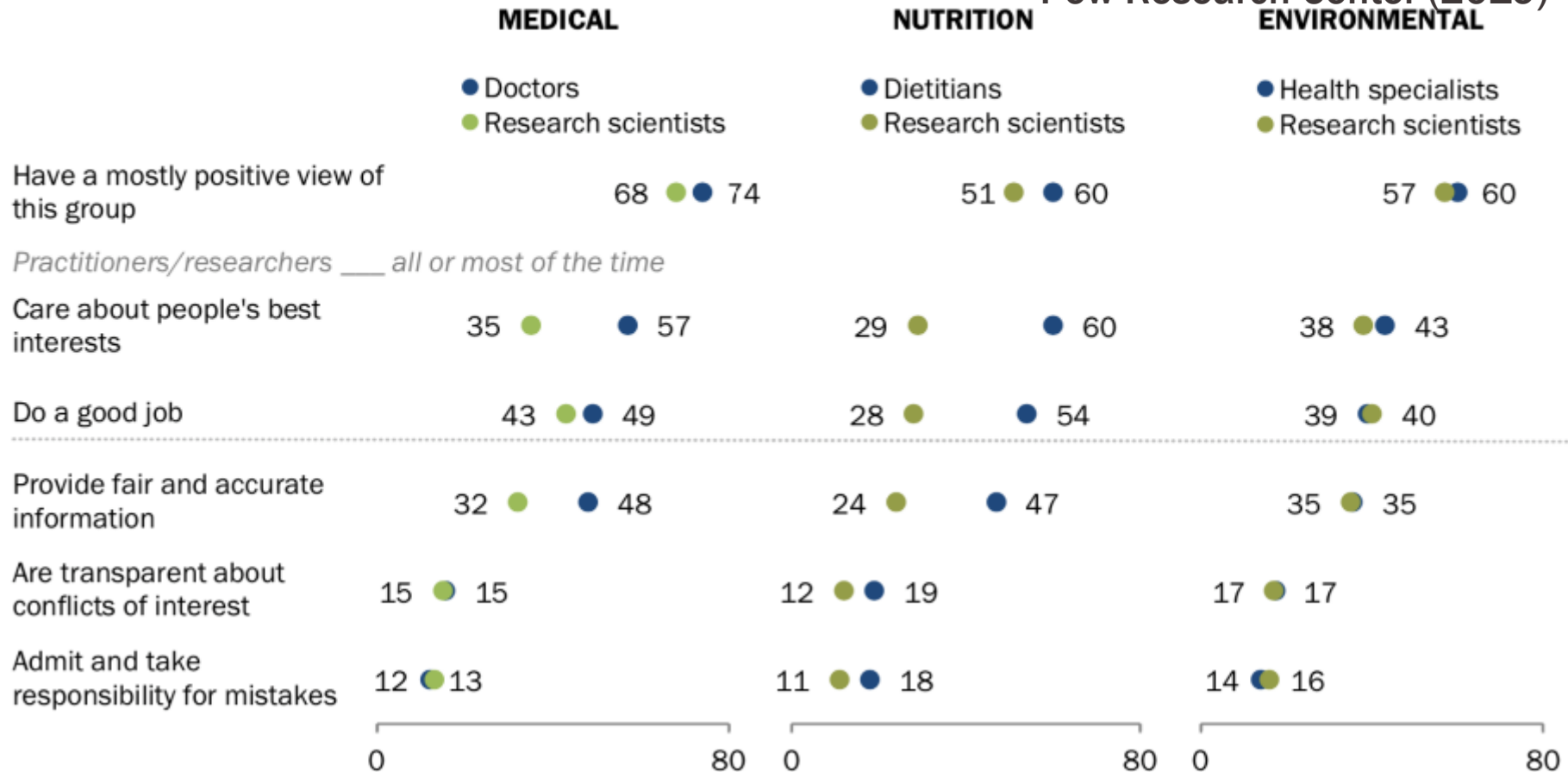
RETWEETS
28

LIKES
76

10:57 AM - 25 Mar 2017

ARE SCIENTISTS SELF-CORRECTING?

Americans trust medical and food science practitioners more than researchers
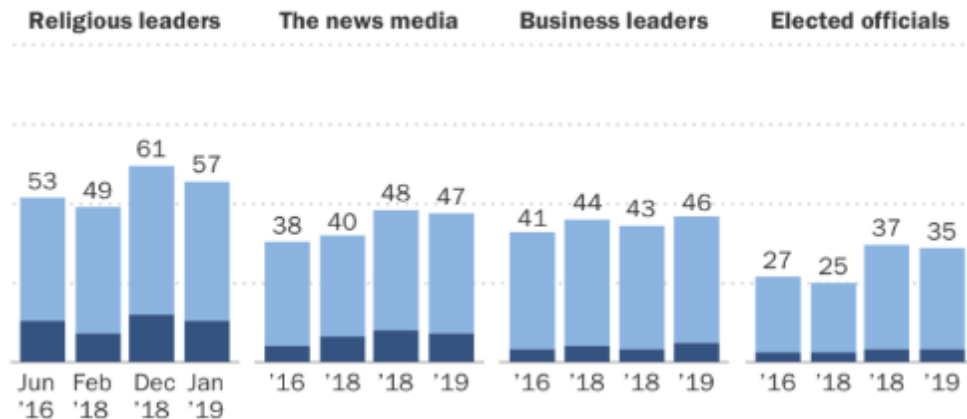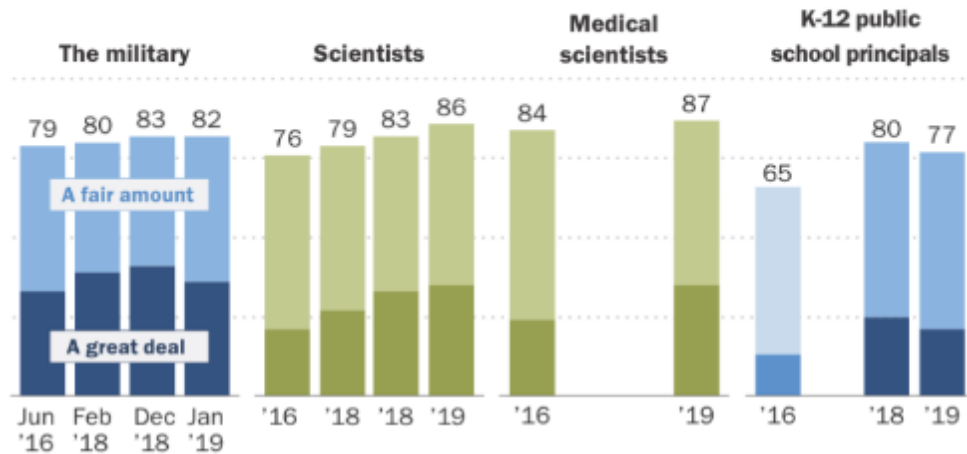
% of U.S. adults who say the following about each of these groups

Pew Research Center (2019)

	MEDICAL		**NUTRITION**		**ENVIRONMENTAL**	
	● Doctors		● Dietitians		● Health specialists	
	● Research scientists		● Research scientists		● Research scientists	
Have a mostly positive view of this group	68 ● ● 74		51 ● ● 60		57 ● ● 60	

Practitioners/researchers ___ all or most of the time

Care about people's best interests	35 ● ● 57		29 ● ● 60		38 ● ● 43	
Do a good job	43 ● ● 49		28 ● ● 54		39 ● ● 40	
Provide fair and accurate information	32 ● ● 48		24 ● ● 47		35 ● ● 35	
Are transparent about conflicts of interest	15 ● ● 15		12 ● ● 19		17 ● ● 17	
Admit and take responsibility for mistakes	12 ● ● 13		11 ● ● 18		14 ● ● 16	
	0 80		0 80		0 80	

AND YET…

% of U.S. adults who say they have a great deal or fair amount of confidence in each of the following groups to act in the best interests of the public
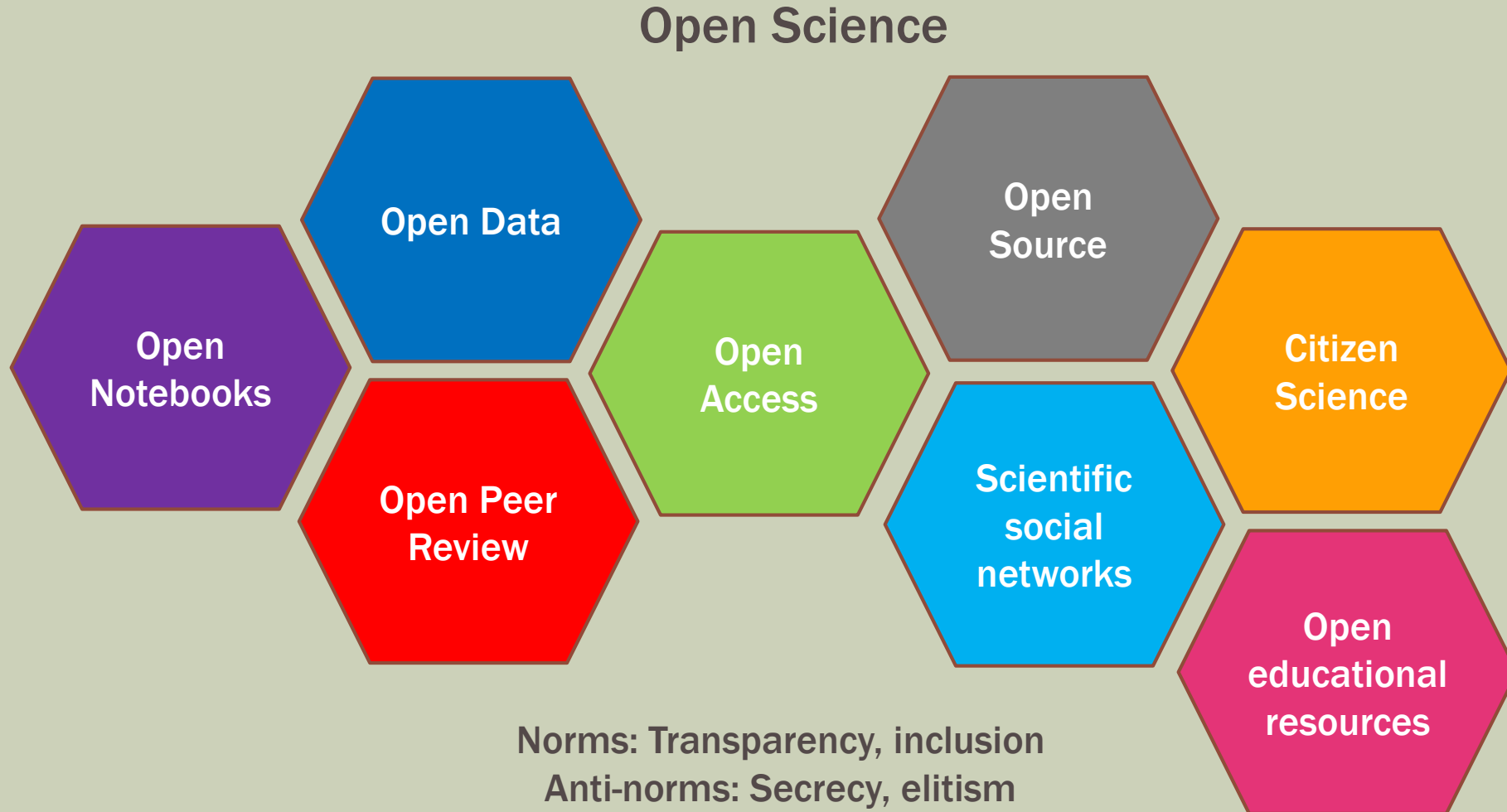
WHY TRUST SCIENCE?

"The resulting dependability of reports [...] comes from a social process rather than from dependence upon the honesty and competence of any single experimenter. [...]
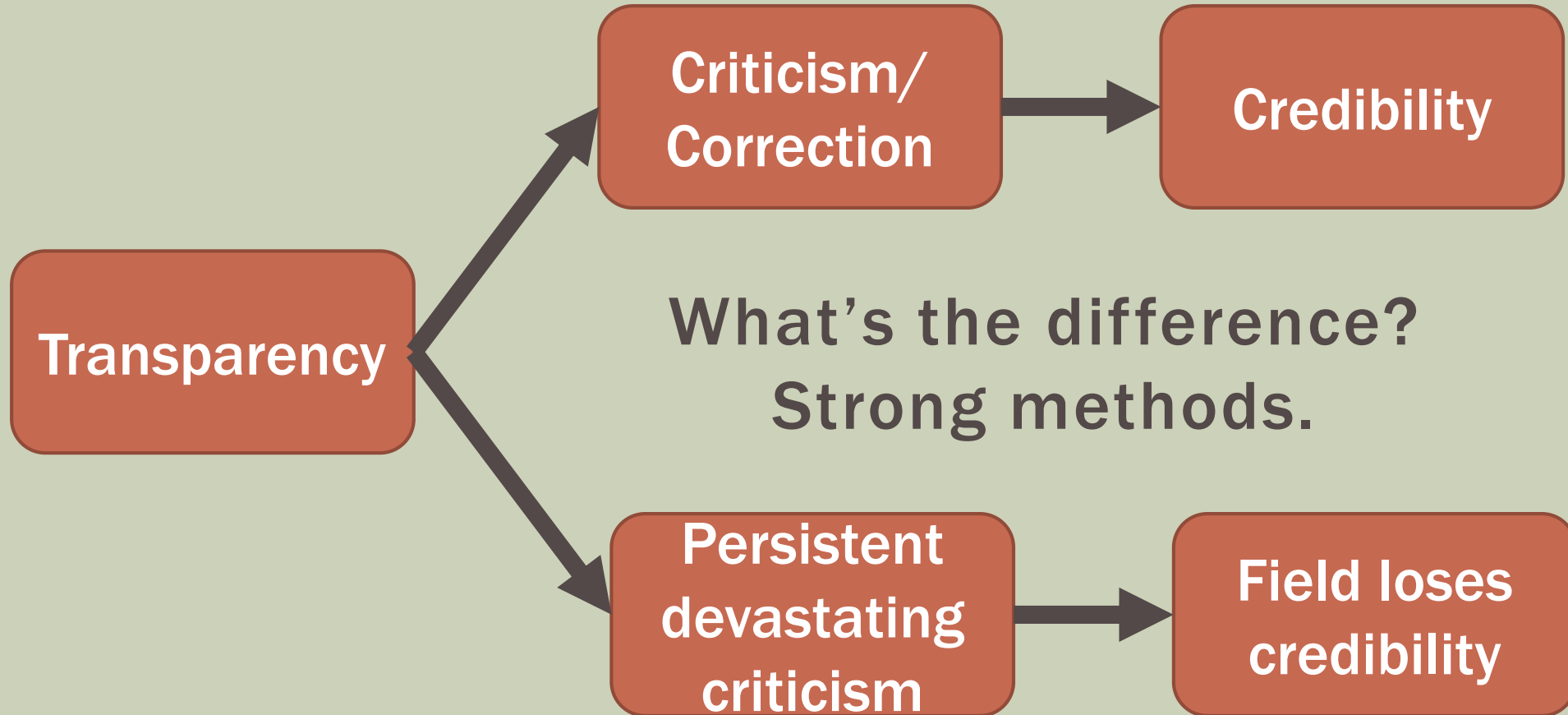
Organized distrust produces trustworthy reports."

-Donald Campbell (1984)

THE CREDIBILITY REVOLUTION

Open Science

Open Notebooks

Open Data

Open Access

Open Source

Citizen Science

Open Peer Review

Scientific social networks

Open educational resources

Norms: Transparency, inclusion
Anti-norms: Secrecy, elitism

THE CREDIBILITY REVOLUTION

THE CREDIBILITY REVOLUTION

Open Science

Quality Control

+

Norms: Transparency, inclusion
Anti-norms: Secrecy, elitism

Norms: Organized skepticism
Anti-norms: Dogmatism, deference, credulity

THE CREDIBILITY REVOLUTION

Open Science

Norms: Transparency, inclusion
Anti-norms: Secrecy, elitism

+

Quality Control

Norms: Organized skepticism
Anti-norms: Dogmatism, deference, credulity

THE CREDIBILITY REVOLUTION

99+%

80%

60%

40%

Open Science Quality Control
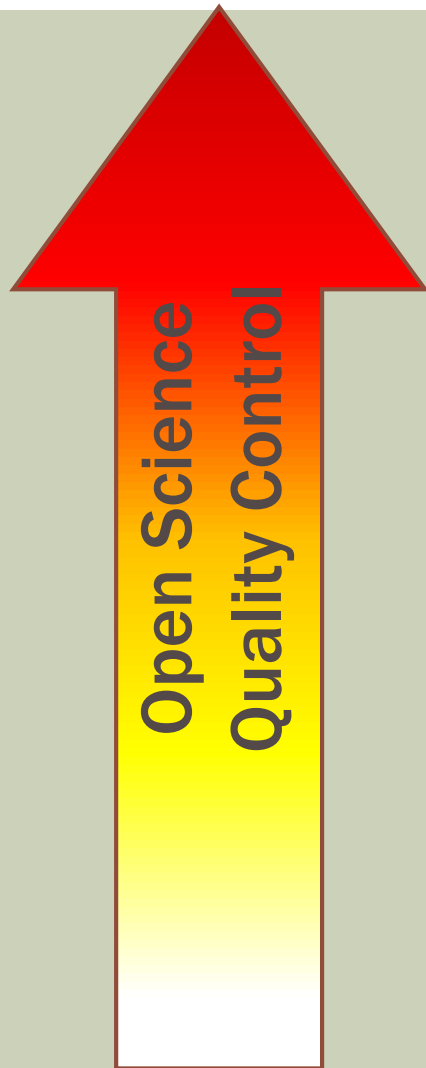
Trust these

Verify or correct these

THREATS TO CREDIBILITY

DEFINITIONS
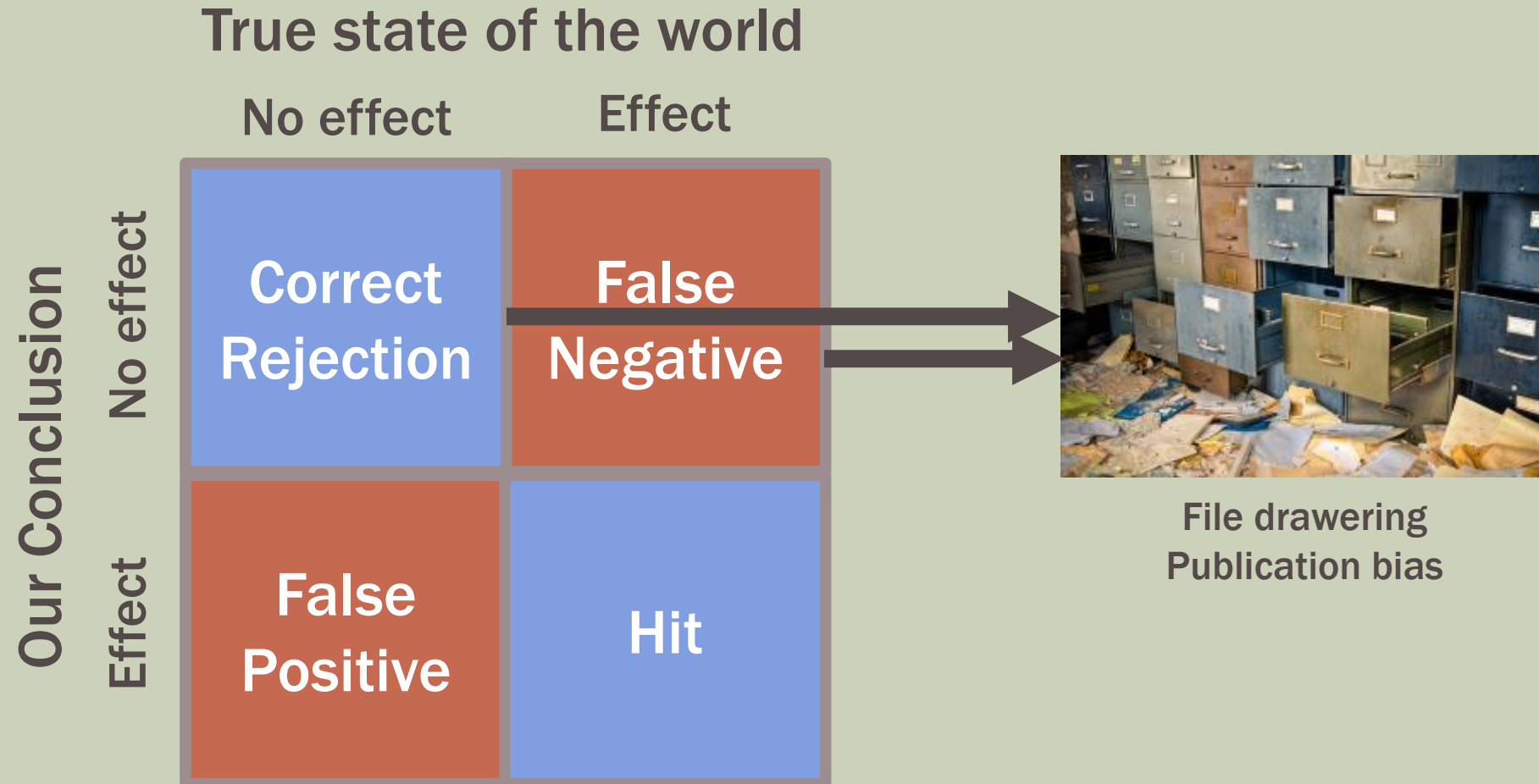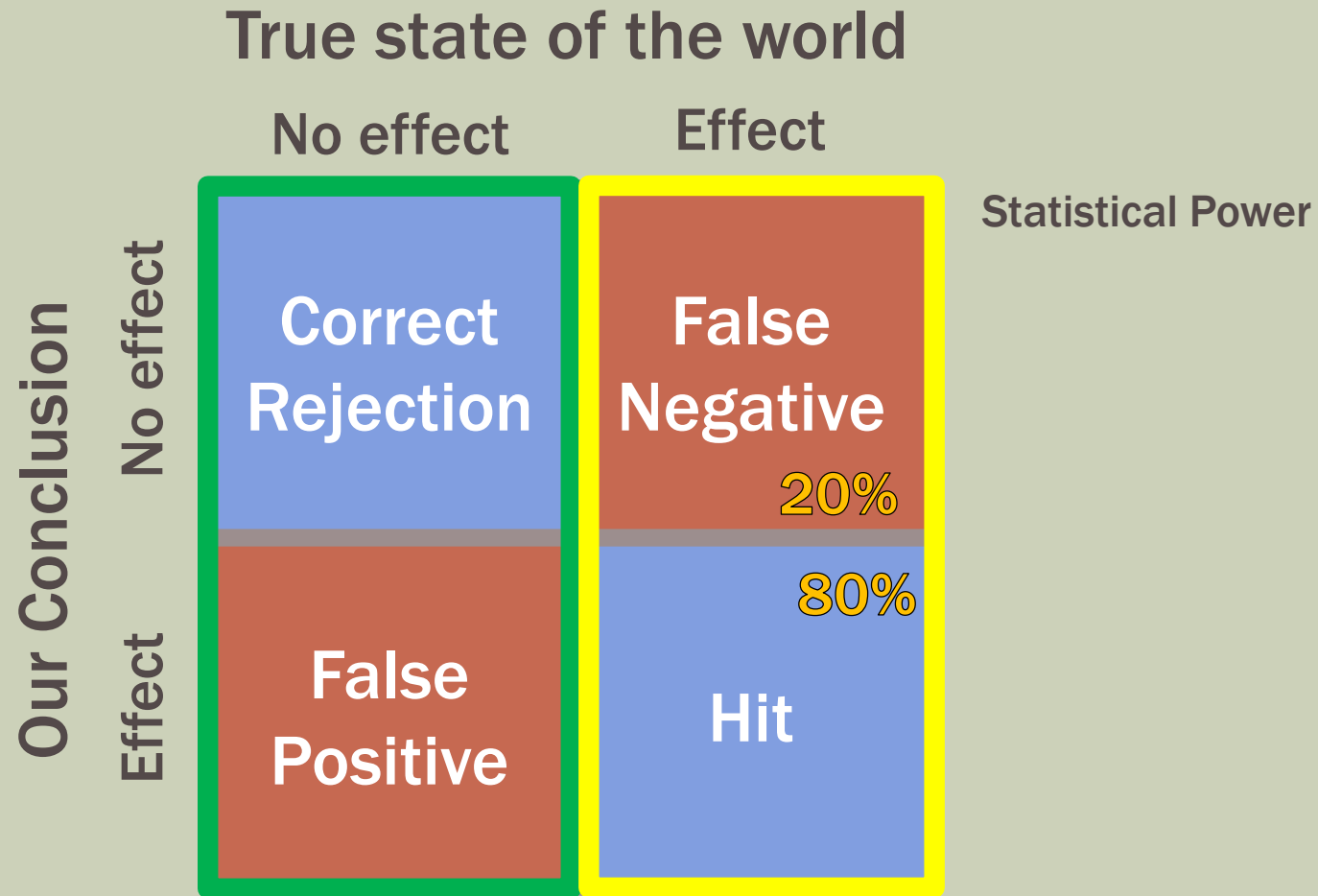
True state of the world

Our Conclusion

	No effect	Effect
No effect	Correct Rejection	False Negative
Effect	False Positive	Hit

File drawering
Publication bias

DEFINITIONS

DEFINITIONS

True state of the world

DEFINITIONS

DEFINITIONS

True state of the world
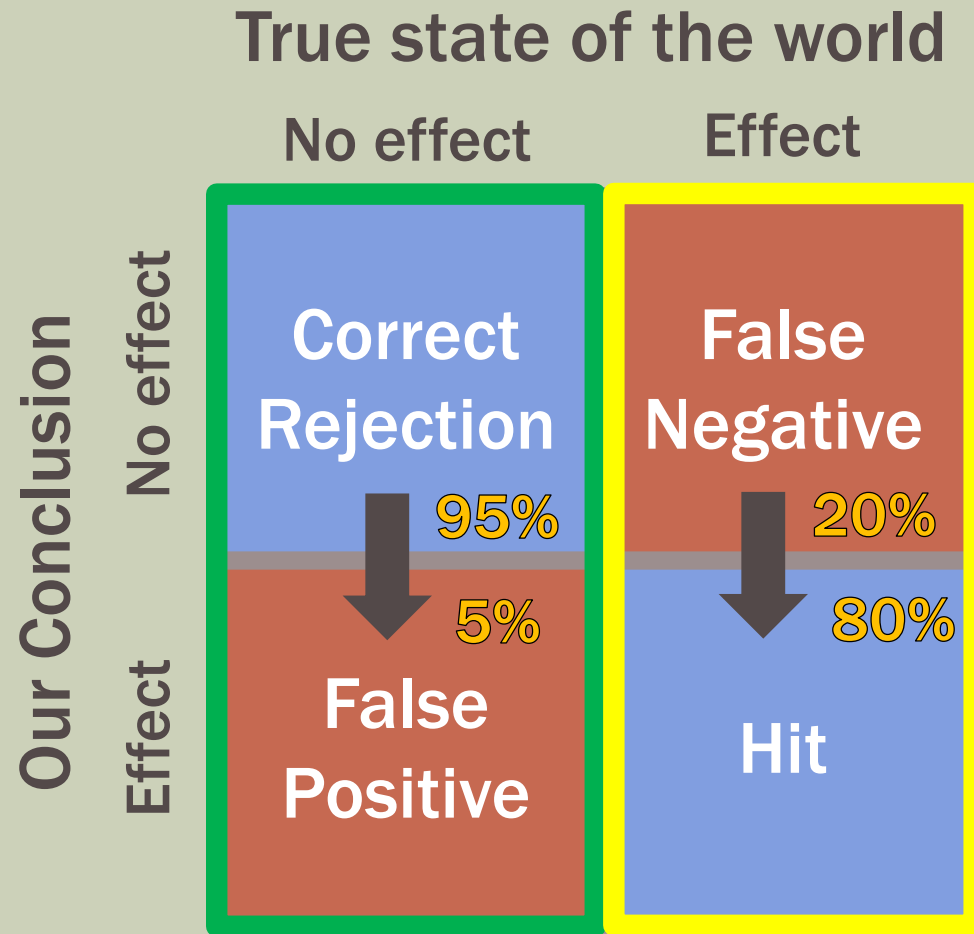
Questionable Research Practices:
Giving yourself many chances &
not disclosing that flexibility
p-hacking
HARKing

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 $p<.05$.

Leif Nelson's slide

DEFINITIONS: HARKING
(HYPOTHESIZING AFTER RESULTS ARE KNOWN)
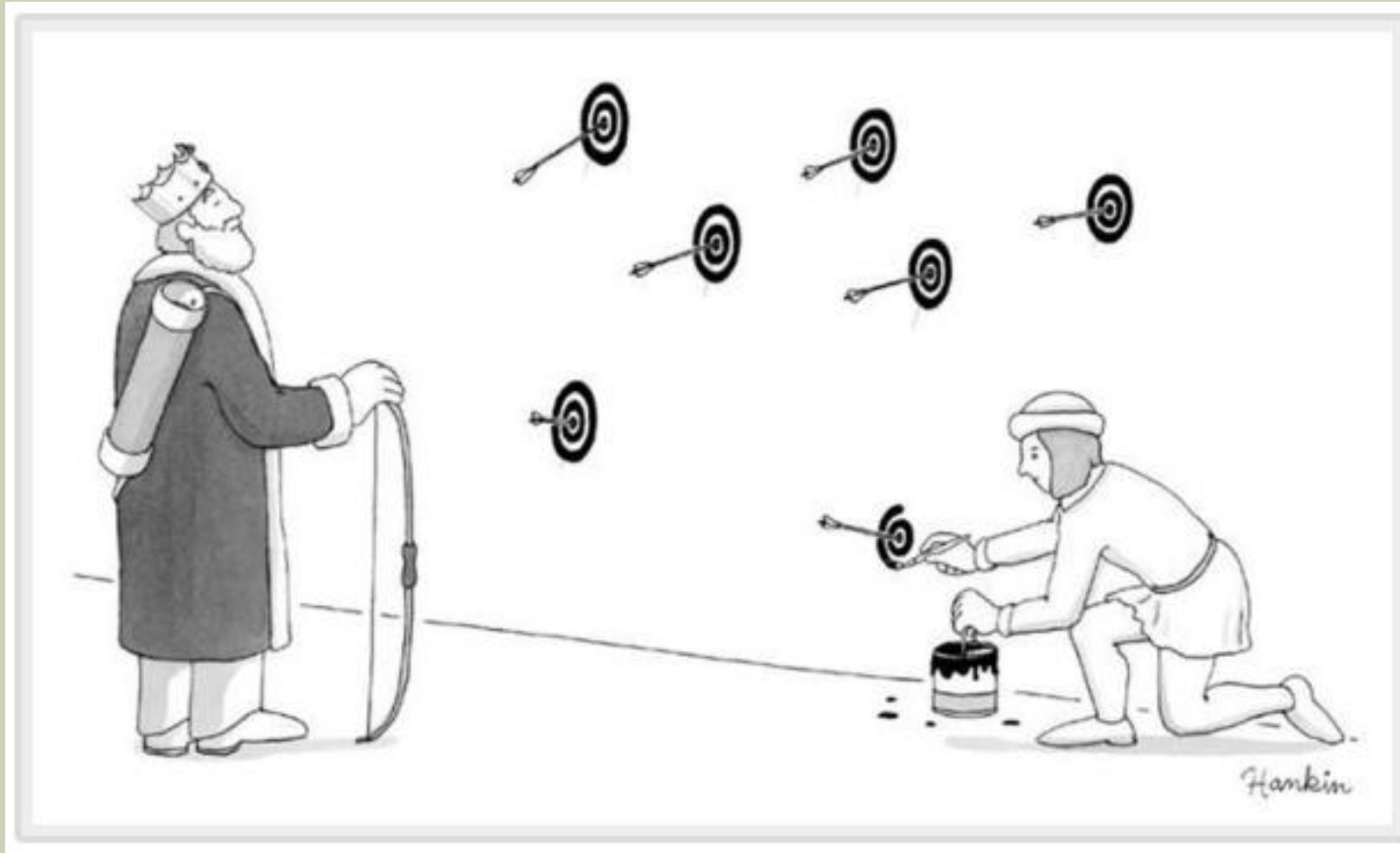
QUESTIONABLE RESEARCH PRACTICES

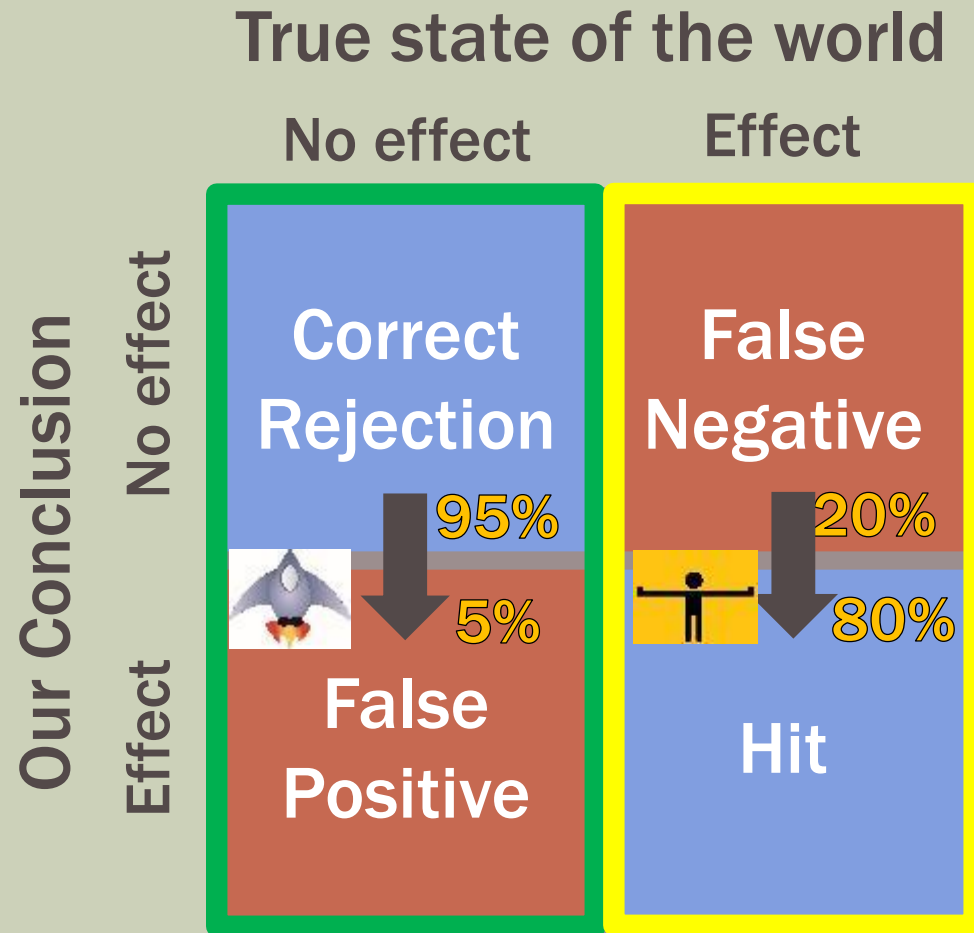|	Self-admission rate (%)
Item	Control group
1. In a paper, failing to report all of a study's dependent measures	63.4
2. Deciding whether to collect more data after looking to see whether the results were significant	55.9
3. In a paper, failing to report all of a study's conditions	27.7
4. Stopping collecting data earlier than planned because one found the result that one had been looking for	15.6
5. In a paper, "rounding off" a p value (e.g., reporting that a p value of .054 is less than .05)	22.0
6. In a paper, selectively reporting studies that "worked"	45.8
7. Deciding whether to exclude data after looking at the impact of doing so on the results	38.2
8. In a paper, reporting an unexpected finding as having been predicted from the start	27.0
9. In a paper, claiming that results are unaffected by demographic variables (e.g., gender) when one is actually unsure (or knows that they do)	3.0
10. Falsifying data	0.6

p-hacking (items 1–5)

File drawering (item 6)

p-hacking (item 7)

HARKing (item 8)

John et al. (2012)

DEFINITIONS

THE CREDIBILITY REVOLUTION

Open Science
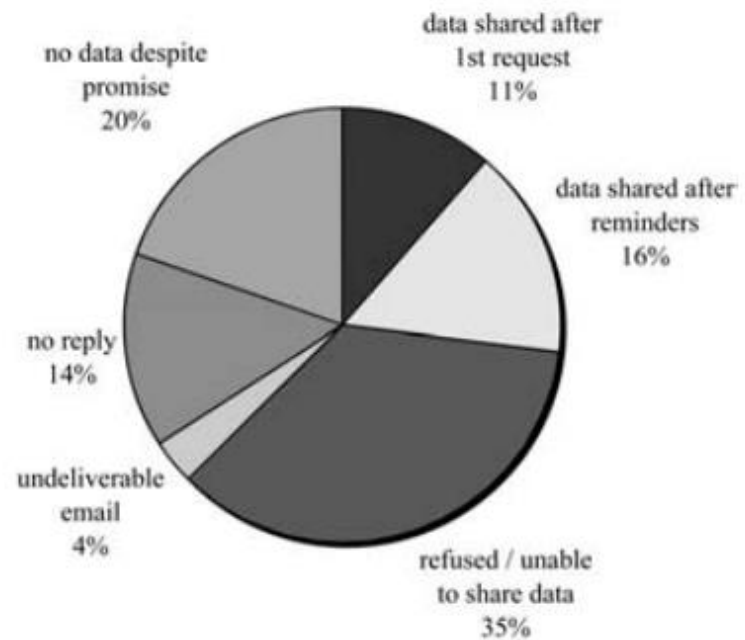
Quality control

THE CREDIBILITY REVOLUTION

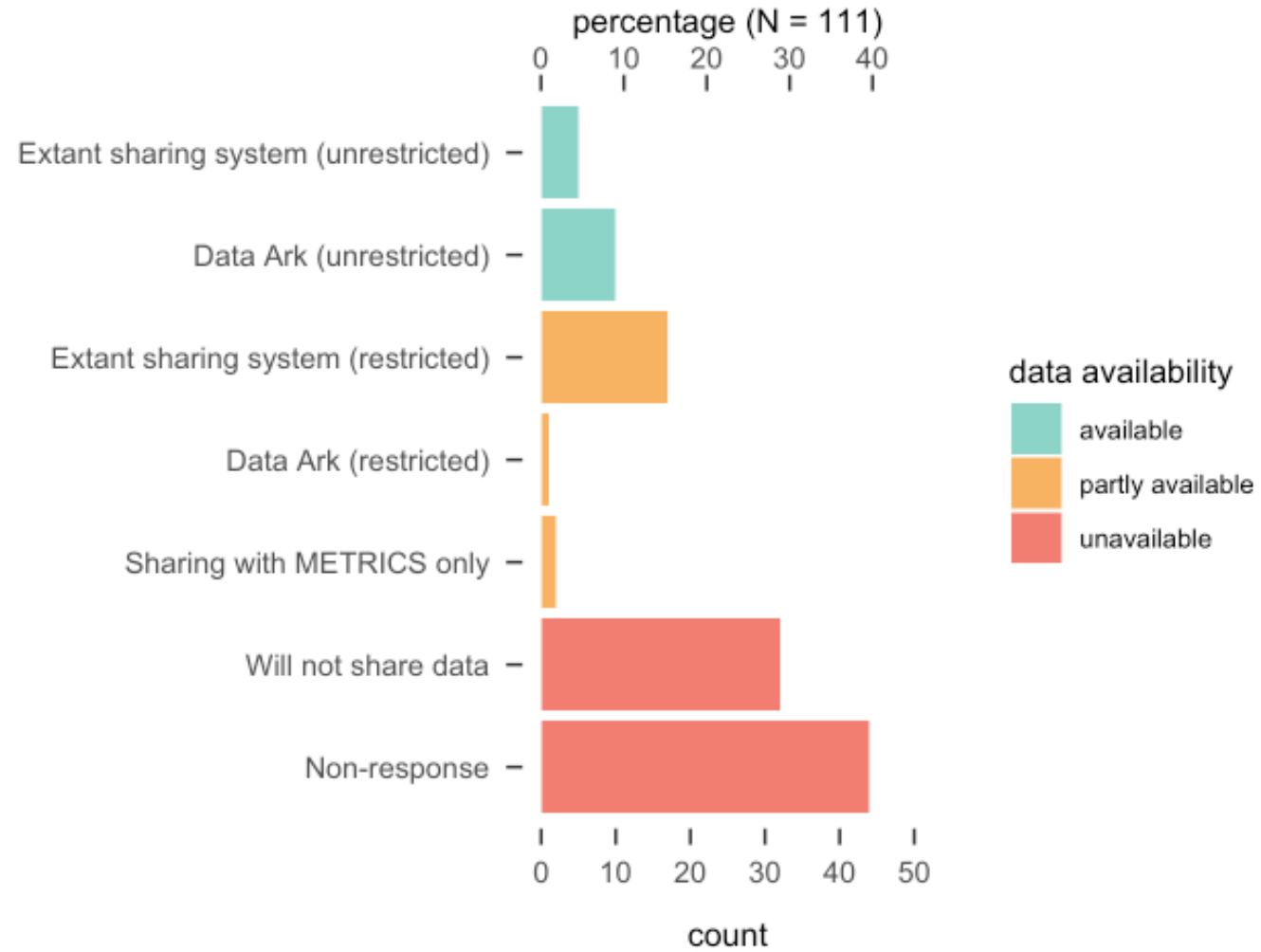Open Science
- Open data and code
- Open materials/notebooks

HOW ARE WE DOING?
DATA SHARING
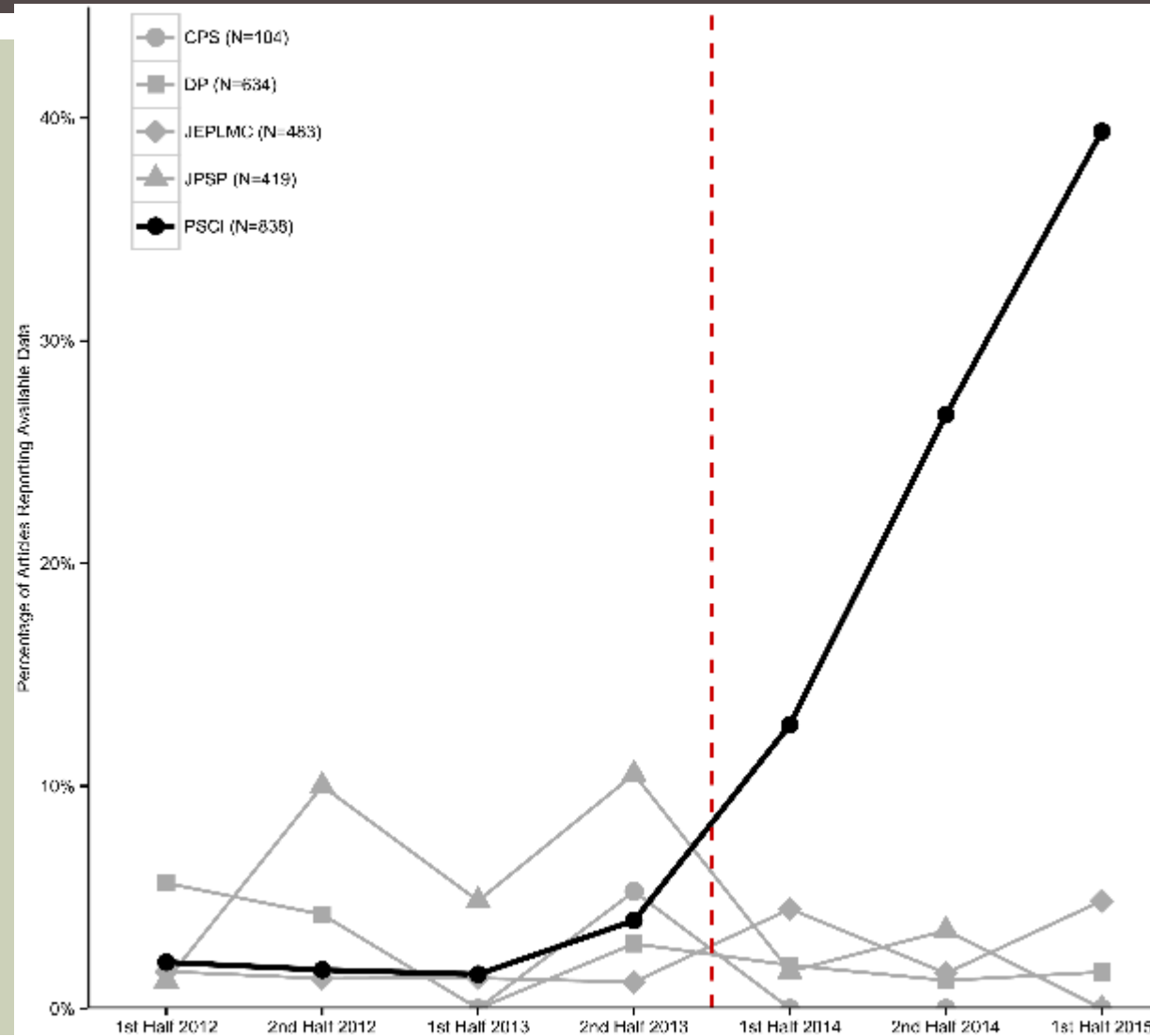

Figure 1.
Percentages of Empirical Articles' Corresponding Authors in Different Response Categories

- no data despite promise 20%
- data shared after 1st request 11%
- data shared after reminders 16%
- no reply 14%
- undeliverable email 4%
- refused / unable to share data 35%

Wicherts et al. (2006)

HOW ARE WE DOING?
DATA SHARING

Hardwicke & Ioannidis (2018)

HOW ARE WE DOING?
DATA SHARING

Kidwell et al. (2016)

THE CREDIBILITY REVOLUTION

Open Science

- Open data and code
- Open materials/notebooks
- Pre-registration

HOW ARE WE DOING?
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.

Leif Nelson's slide

Planning ahead

Study preregistrations on the Open Science Framework (OSF) are doubling every year; more than 120 journals have introduced registered reports.
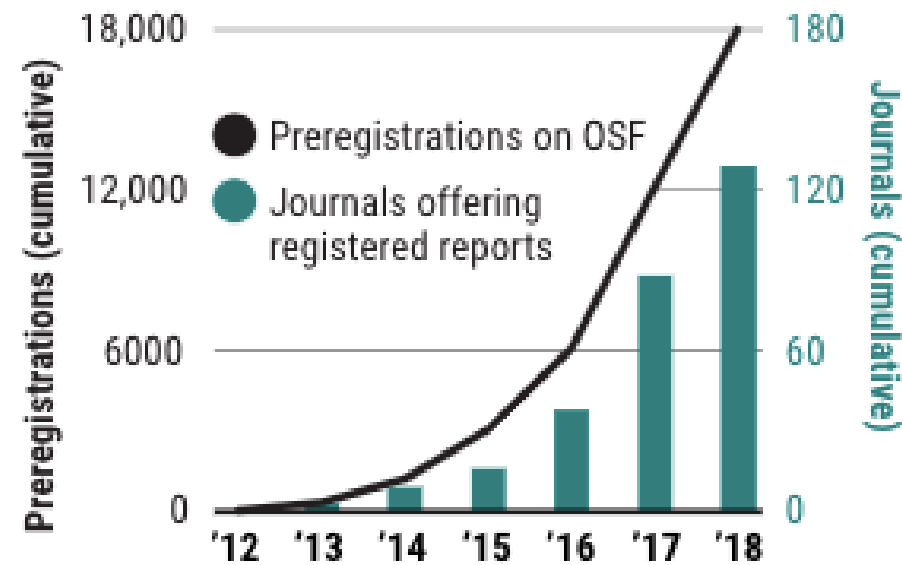
Kupferschmidt (2018)

THE CREDIBILITY REVOLUTION

Open Science

- Open data and code
- Open materials/notebooks
- Pre-registration
- Open access/preprints
- Open review
- Open source
- No barriers to entry
- Declaring conflicts of interest
- Contributorship instead of authorship

THE CREDIBILITY REVOLUTION

Open Science
- Open data and code
- Open materials/notebooks
- Pre-registration
- Open access/preprints
- Open review
- Open source
- No barriers to entry
- Declaring conflicts of interest
- Contributorship instead of authorship

Quality control
- Error detection

HOW ARE WE DOING?
STATISTICAL ERRORS

Nuijten et al. (2016)

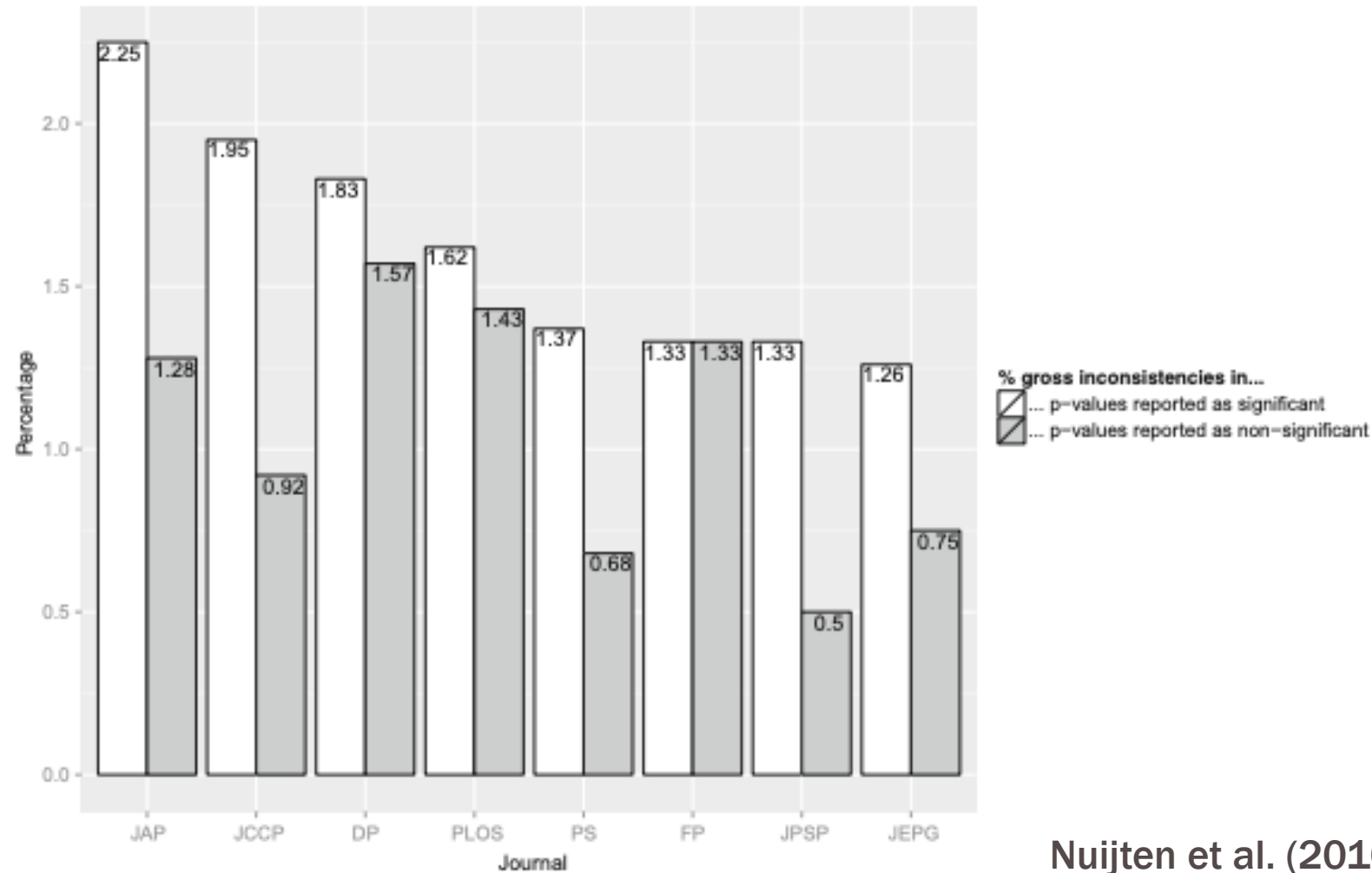# THE CREDIBILITY REVOLUTION

Open Science

- Open data and code
- Open materials/notebooks
- Pre-registration
- Open access/preprints
- Open review
- Open source
- No barriers to entry
- Declaring conflicts of interest
- Contributorship instead of authorship

Quality control

- Error detection
- Reproducibility: can it be repeated using same data?
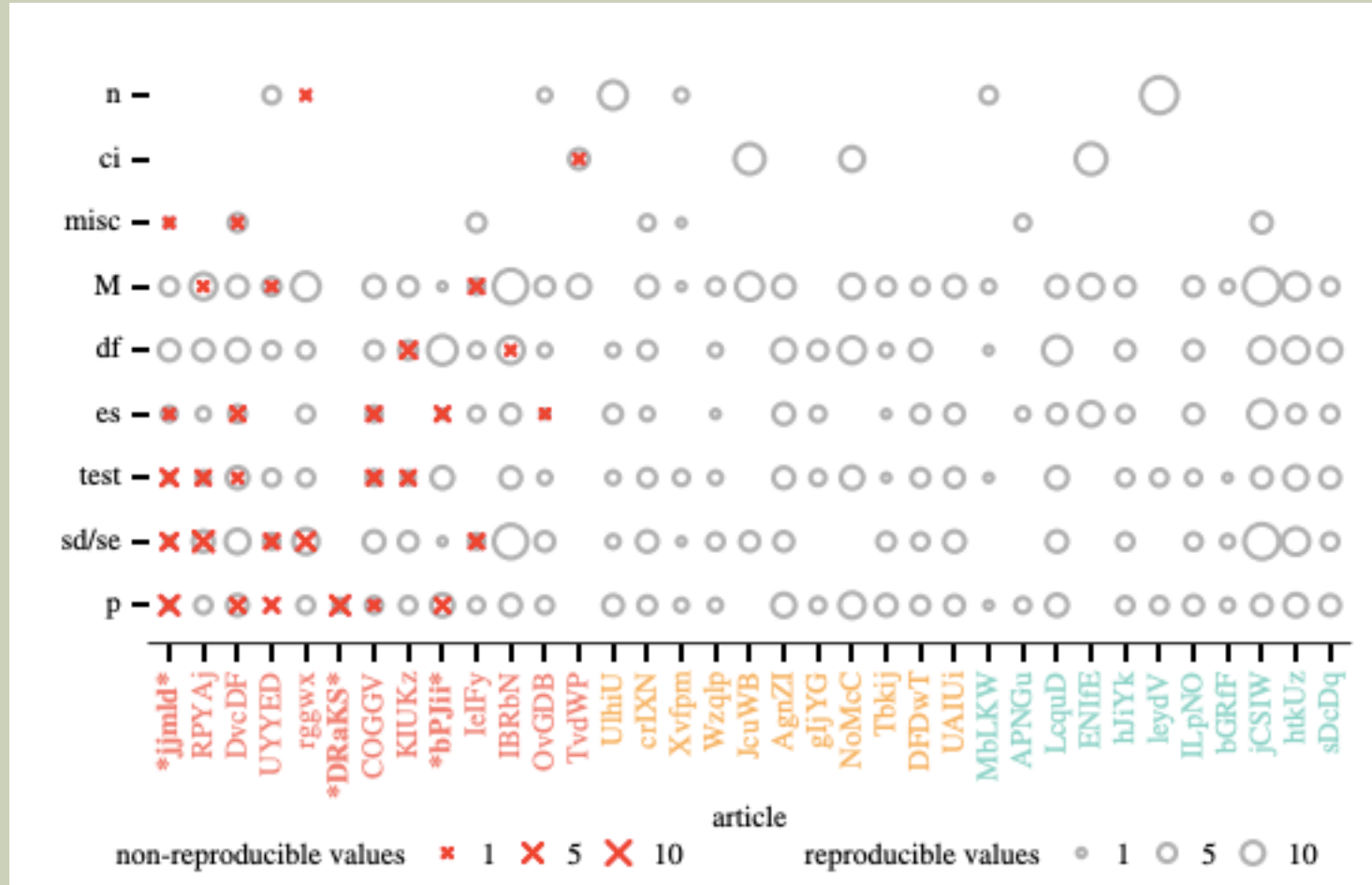
Hardwicke et al. (2018)

THE CREDIBILITY REVOLUTION

Open Science
- Open data and code
- Open materials/notebooks
- Pre-registration
- Open access/preprints
- Open review
- Open source
- No barriers to entry
- Declaring conflicts of interest
- Contributorship instead of authorship

Quality control
- Error detection
- Reproducibility: can it be repeated using same data?
- Replicability: can it be repeated from scratch?

HOW ARE WE DOING? REPLICABILITY

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)
- 2/9 among RRRs (Psychology)

= 89/199 = 45%* replicability rate

= 55%* false discovery rate

* with large uncertainties!
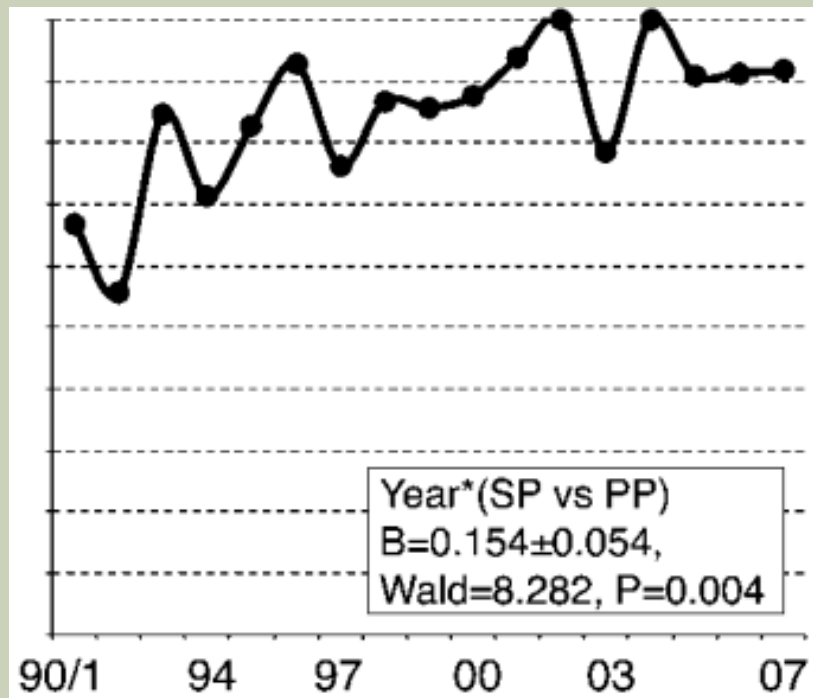
THE CREDIBILITY REVOLUTION

Open Science
- Open data and code
- Open materials/notebooks
- Pre-registration
- Open access/preprints
- Open review
- Open source
- No barriers to entry
- Declaring conflicts of interest
- Contributorship instead of authorship

Quality control
- Error detection
- Reproducibility: can it be repeated using same data?
- Replicability: can it be repeated from scratch?
- Publication of negative results, corrections, criticisms
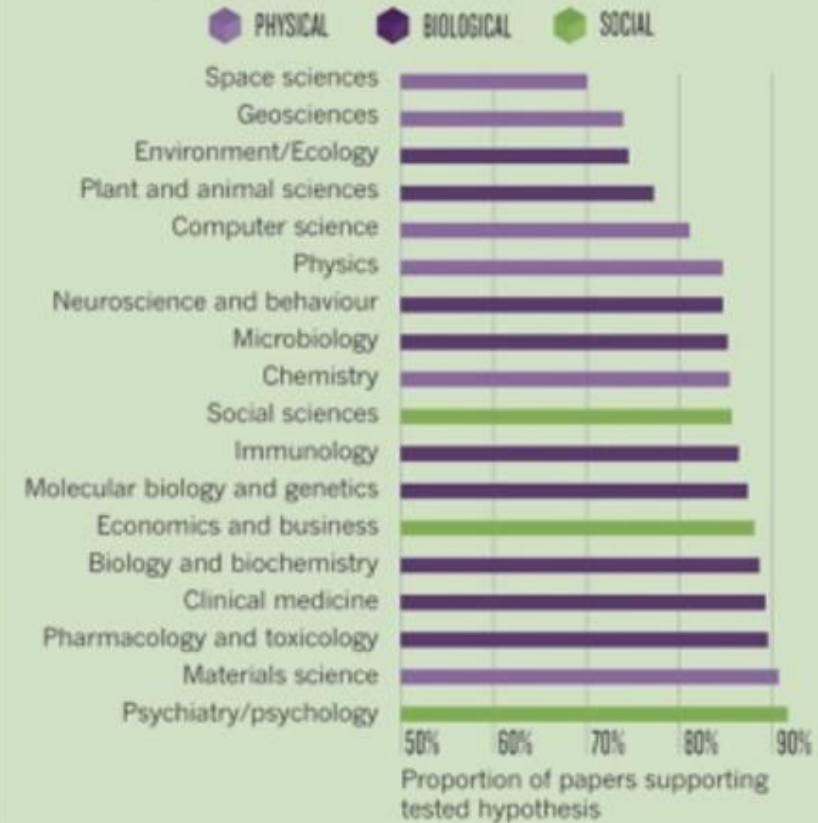
HOW ARE WE DOING?
PUBLISHING NULL RESULTS

% significant (Psychology)

Year*(SP vs PP)
B=0.154±0.054,
Wald=8.282, P=0.004

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.

● PHYSICAL   ● BIOLOGICAL   ● SOCIAL

Space sciences
Geosciences
Environment/Ecology
Plant and animal sciences
Computer science
Physics
Neuroscience and behaviour
Microbiology
Chemistry
Social sciences
Immunology
Molecular biology and genetics
Economics and business
Biology and biochemistry
Clinical medicine
Pharmacology and toxicology
Materials science
Psychiatry/psychology

50%   60%   70%   80%   90%

Proportion of papers supporting tested hypothesis

THE CREDIBILITY REVOLUTION

Open Science

- Open data and code
- Open materials/notebooks
- Pre-registration
- Open access/preprints
- Open review
- Open source
- No barriers to entry
- Declaring conflicts of interest
- Contributorship instead of authorship

Quality control

- Error detection
- Reproducibility: can it be repeated using same data?
- Replicability: can it be repeated from scratch?
- Publication of negative results, corrections, criticisms
- Post Publication Peer Review (PPPR)

THE CREDIBILITY REVOLUTION

Open Science

+

Quality Control

THE CREDIBILITY REVOLUTION
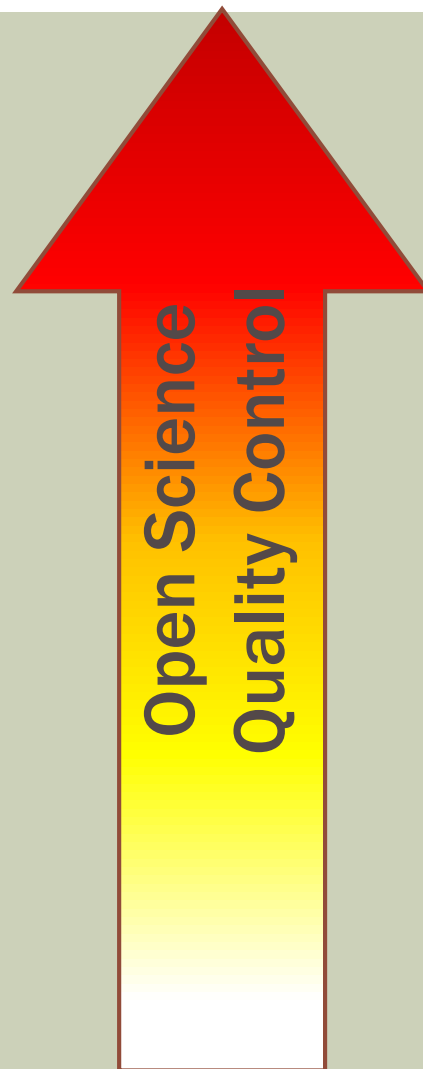
Open Science

+

Quality Control

"Another flaw in the human character
is that everybody wants to build and
nobody wants to do maintenance."

-Kurt Vonnegut

THE CREDIBILITY REVOLUTION

Open Science Quality Control

99+%

80%

60%

40%

Trust these

Verify or correct these

WHAT'S AT STAKE

"If we present our resulting improved truth claims
as though they were definitive achievements comparable to those in the physical sciences,
and thus deserving to override ordinary wisdom
when they disagree, we can be socially destructive.
We can be engaged in the political misuse of the authority of science that has not been
fully earned in our own field."

-Donald Campbell (1984)

THE END