Michelangelo Vianello
Dept. FISPPA, School of Psychology, University of Padova

Slides available

You can open them in your device

Contents

- Create consensus around words.
- The Open Science Movement in psychology since its inception: a view from within
- Observed & expected replicability rates
- A fresh look on how to increase replicability (the devil is not only in the details)
- A tentative summary of changes occurred in the last ten years.
- Conclusions

Ice-breaker: survey on open research practices

Glossary: a proposal

Term	Definition
Reproducibility	Getting the *same results* applying the *same analysis* on the *same data (e.g.* computational reproducibility). Failures could be due to either process (lack of transparency) or outcome (error in one of the two studies).
Replicability	Getting the *same results* applying the *same analysis* on *different data*. All replicability tests also inform on generalizability on findings, not the opposite.
Robustness	Getting the *same results* applying a *different analysis* on the *same data*
Generalizability	Only positive results inform credibility (otherwise called «conceptual replications»). Null results do not inform credibility of the original.
Questionable research practices	Activities that intentionally or unintentionally distort results in favor of a researcher's own hypothesis – or omissions in reporting such practices – Examples: «HARKing», selective reporting (e.g. «cherry picking» studies, covariates, or observations); p-Hacking; …

Notes: An effect can be perfectly reproducible, replicable, and robust, yet still not valid. Definitions adapted from Nosek et al., 2022 and Kormbacher et al., 2023

Two schools?
1513: Machiavelli on change

- "And it should be considered how there is nothing more difficult to handle, more doubtful of success, and more dangerous to conduct than to take the lead in introducing new orders. For the introducer has as enemies all those who do well under the old orders, and lukewarm defenders among those who would do well under the new orders. This lukewarmness arises partly from fear of opponents who have the laws on their side and partly from the skepticism of men who do not readily believe in new things unless they have seen a solid experience of them. Hence, whenever those who are enemies have an opportunity to attack, they do it partisans, and the others defend themselves tepidly, so that it becomes dangerous to pair with them."

1967:
Meehl complains about weakness of psychological theories

- "a zealous and clever investigator can slowly wend his way through a tenuous nomological network, performing a long series of related experiments which appear to the uncritical reader as a fine example of 'an integrated research program,' without ever once refuting or corroborating so much as a single strand of the network." (Meehl, 1967, p. 114):

1971:
first signs that NHST is misunderstood

- Researchers (members of the Mathematical Psychology Group and APA) heavily over-estimated the probability of getting a p<.05 on 10 ss after they got a p<.05 on 20 ss (Tversky & Kahneman, 1971)
- Many other scholars warned against the so-called Questionable Research Practices (QRPs), like over-emphasis on statistical significance, publication bias, inadequate power, weak theories and lack of replications (Cohen, 1973, 1994; Greenwald, 1975; Rosenthal, 1979; Sterling, 1959).
- They have been largely ignored for 40 years.

2011: Psych Science has been bombed

1. JPSP publishes article that claims evidence for extrasensory perception (Bem, 2011)

2. Revelations of Fraud in dozens of papers by Stapel, Smeesters and Sanna (among others). As of 2023, 70 articles by these three authors were retracted (58+7+5, respectively).

3. «False-positive psychology» article: seemingly innocent and common methodological behaviors allows presenting anything as significant (Simmons et al., 2011).

4. Concerns over the use of questionable research practices were (finally) publicly shared (John et al., 2012)

- Psychology entered the *credibility revolution* (Kormbacher et al., 2023; Vazire, 2018)

Questionable research practices

https://app.wooclap.com/ESLYRP

Questions from Gopalakrishna et al. (2022)

2012: The reaction starts

1. «Social Psychology» journal invites Registered Replication Reports: Fifteen studies were ultimately published, *most* of them f*ailed to find evidence of replicability* (Nosek & Lakens, 2014).

2. A blog post criticized replication methods (Schnall, 2014) and sparked the *#repligate:* replicators were called «shameless little bullies» and «replication police» by prominent psychologists in Harvard, Princeton and Cambridge. It got hot.

2013: the reaction does not stop

- The Open Science Collaboration was built using a google discussion group. After two years, results of the largest replication study were published (OSC, 2015):
  - 100 effects taken from leading psych journals
  - Replication effects were half the magnitude of originals
  - Across many estimates of replicability, 36% to 60% were successfully replicated
  - Great impetus to Open Scholarship
- Prominent psychologists criticized this paper as well (Gilbert et al., 2016)

OSC*: a huge collaborative effort
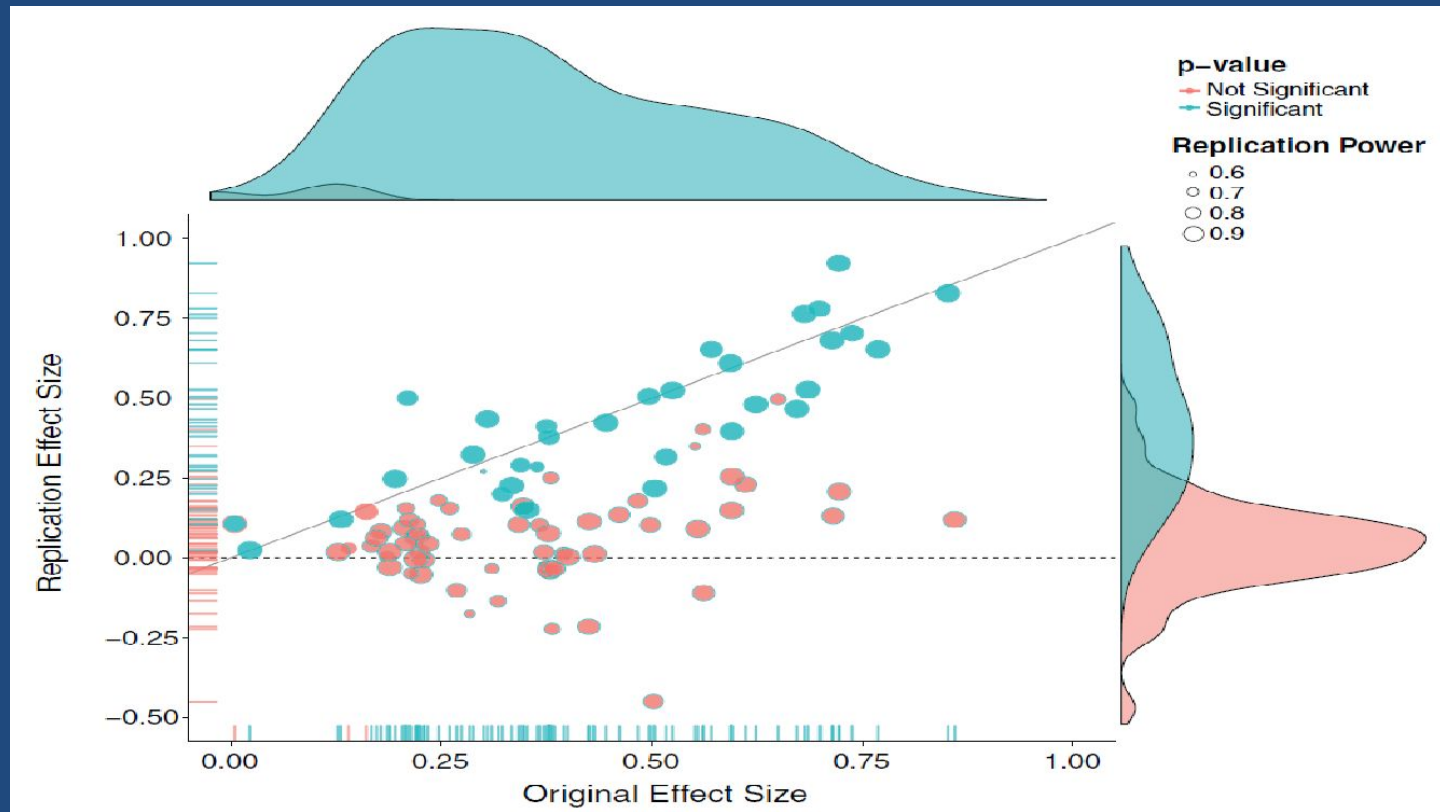
Most pessimistic estimate: 36% replicated.

Most optimistic estimate: 59.87% of original effect sizes were within the replication 95% confidence interval (i.e., 47 out of an expectation of 78.5 given study design and random error).

*Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.

2014: The «Many Labs» series begin

- *Many data collection sites* (~10 to 30) spread across a determined geographical area
- *Many studies* (roughly 10 to 30) are conducted on the same participants
- Large sample size (N>1000)
- All labs provide participants with *the same data collection protocol* and stimuli (in their mother tongue, translation is often necessary).
- *Direct replications* rather than conceptual (goal is to use exactly the same original materials and study design) ☐ minimize procedural differences!
- Research design, hypotheses and analysis plan are *pre-registered*. No room for the "garden of forking paths" (Gelman & Loken, 2014) in confirmatory analysis.
- Clearly *identify exploratory analyses* and eventually provide them after the confirmatory tests.
- *Involve original researchers* in study design and make them review the registered protocol, study materials, and analysis plan (not all of them will accept, but the majority is supportive in my experience).
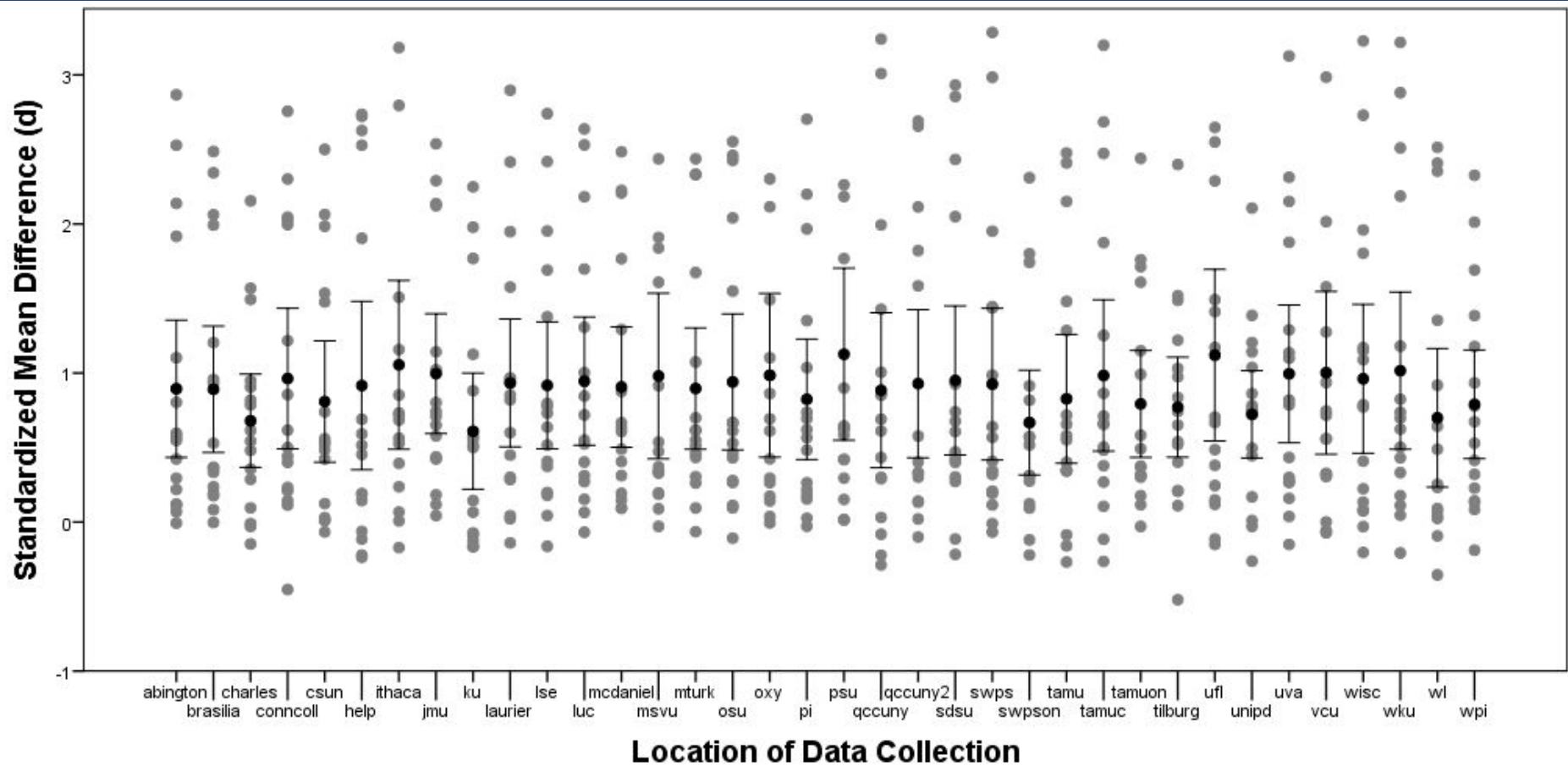
Many Labs 1
(https://osf.io/wx7ck/)

13 effects

36 sites

N=6344
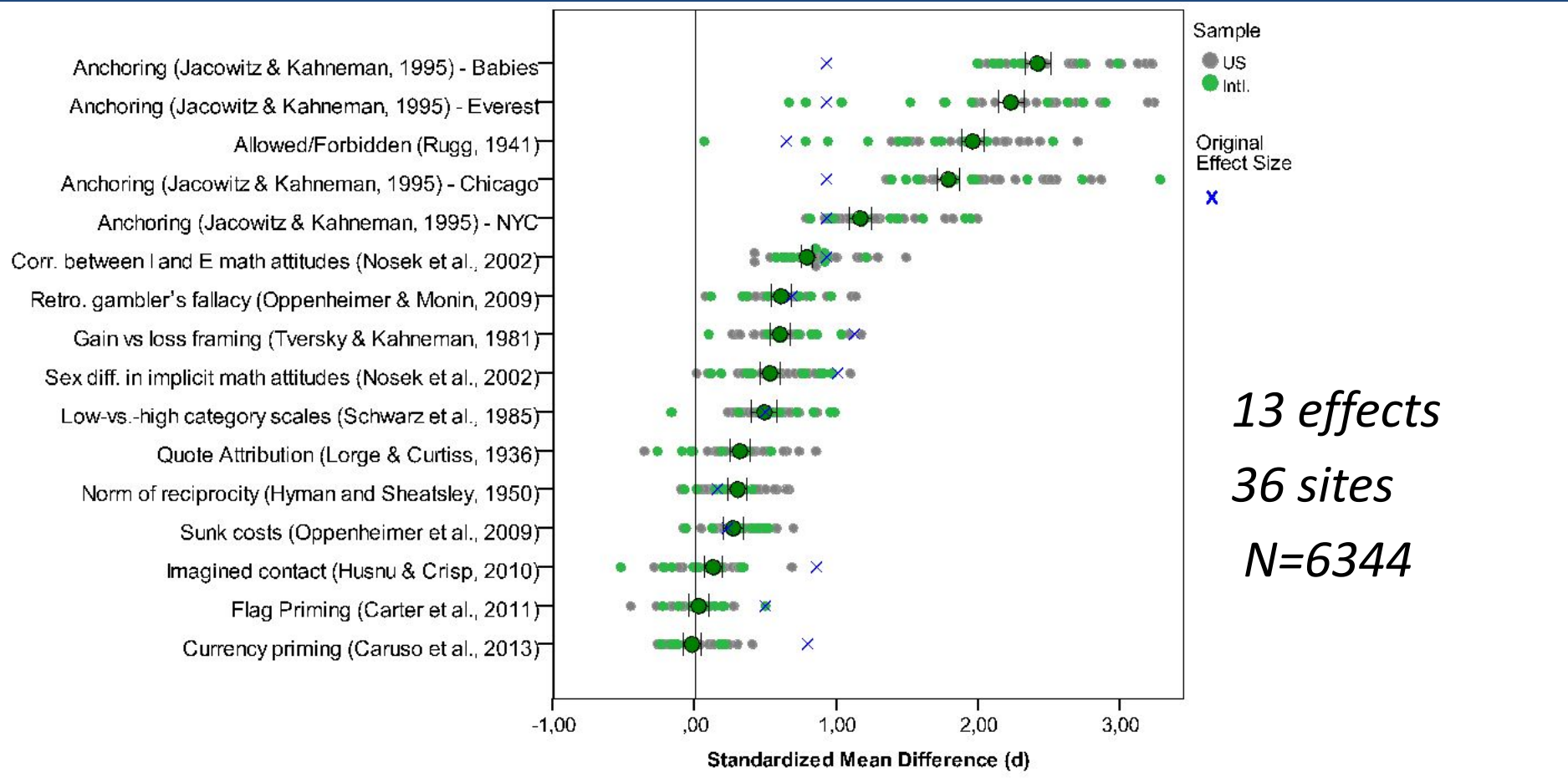
If we consider the anchoring studies as one effect, ML1 successfully replicated 10 out of 13 effects. Both priming effects were not replicable (77%).

Many Labs 1: No variation across sites

2018: Many Labs 2
(https://osf.io/8cd4r/)

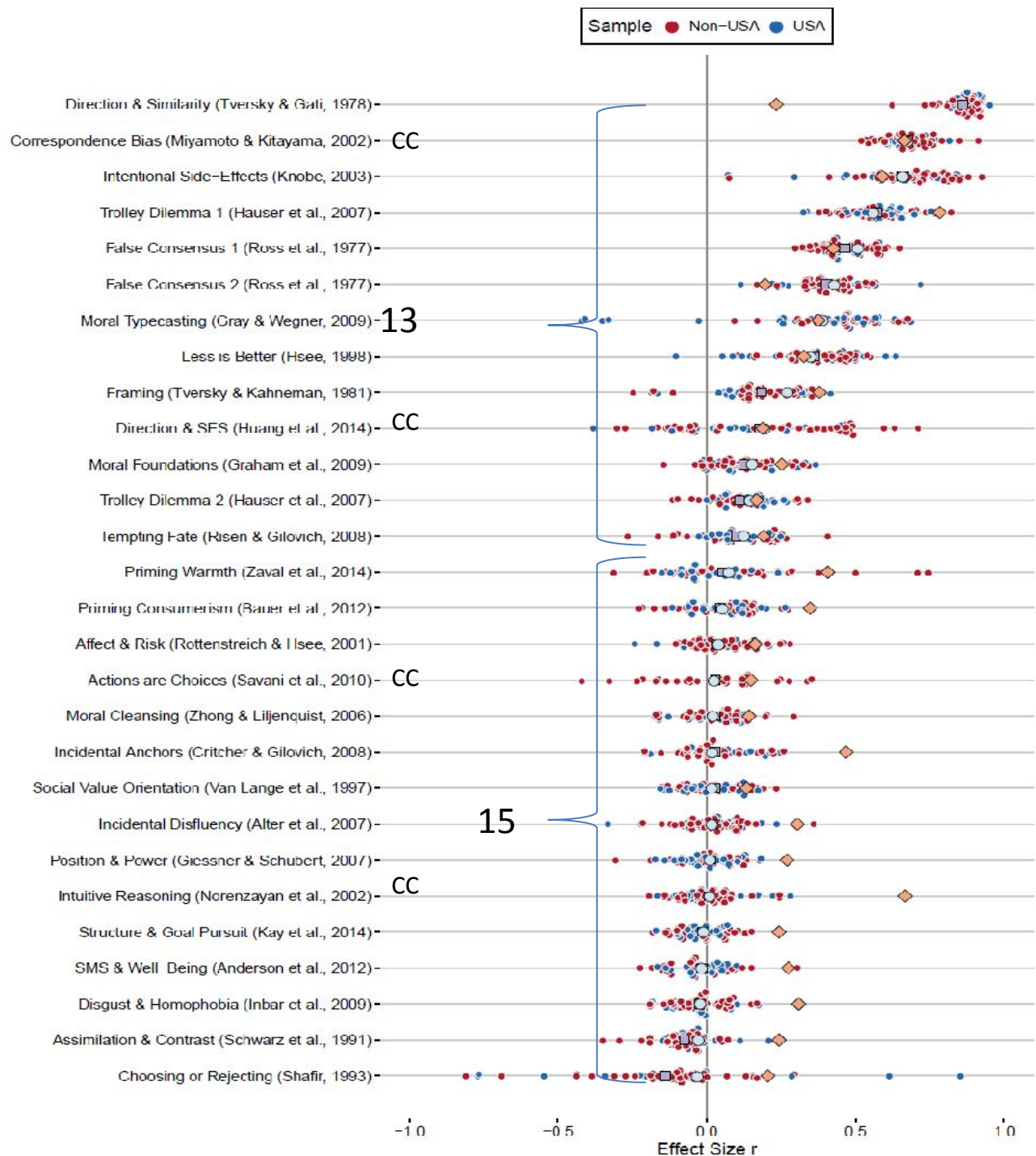28 studies

62 sites

N=15791

Approximate
N per
study=7800

ML 2

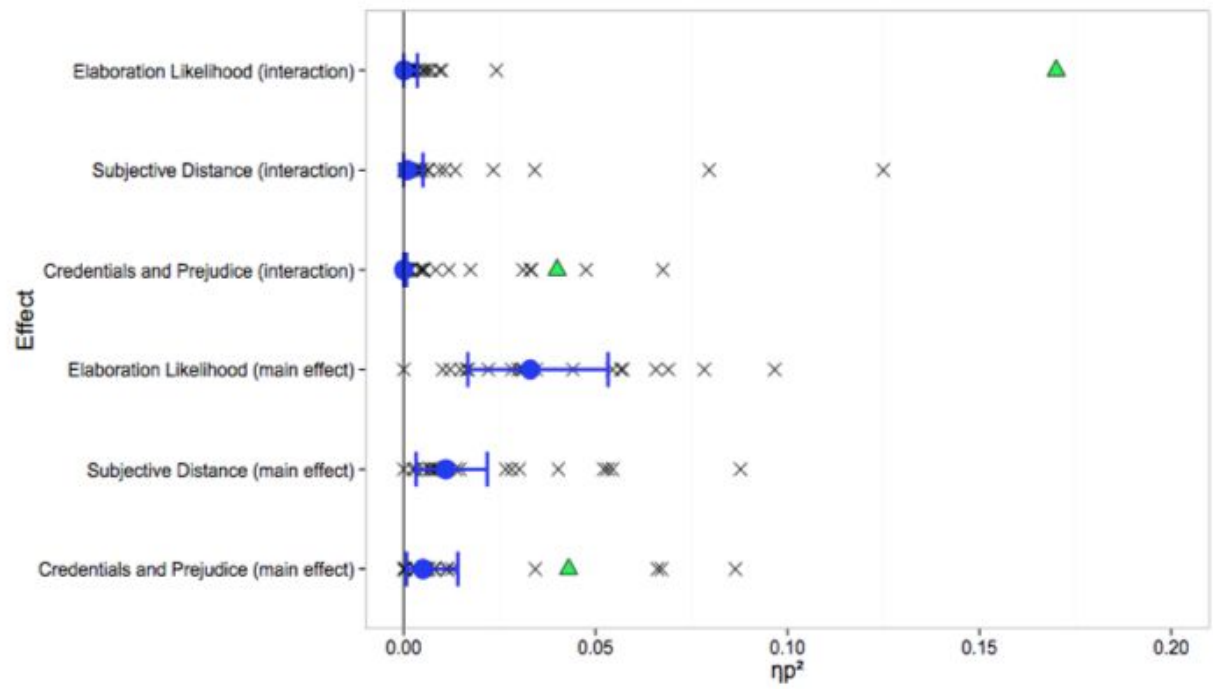13/28 successful replications (46%)
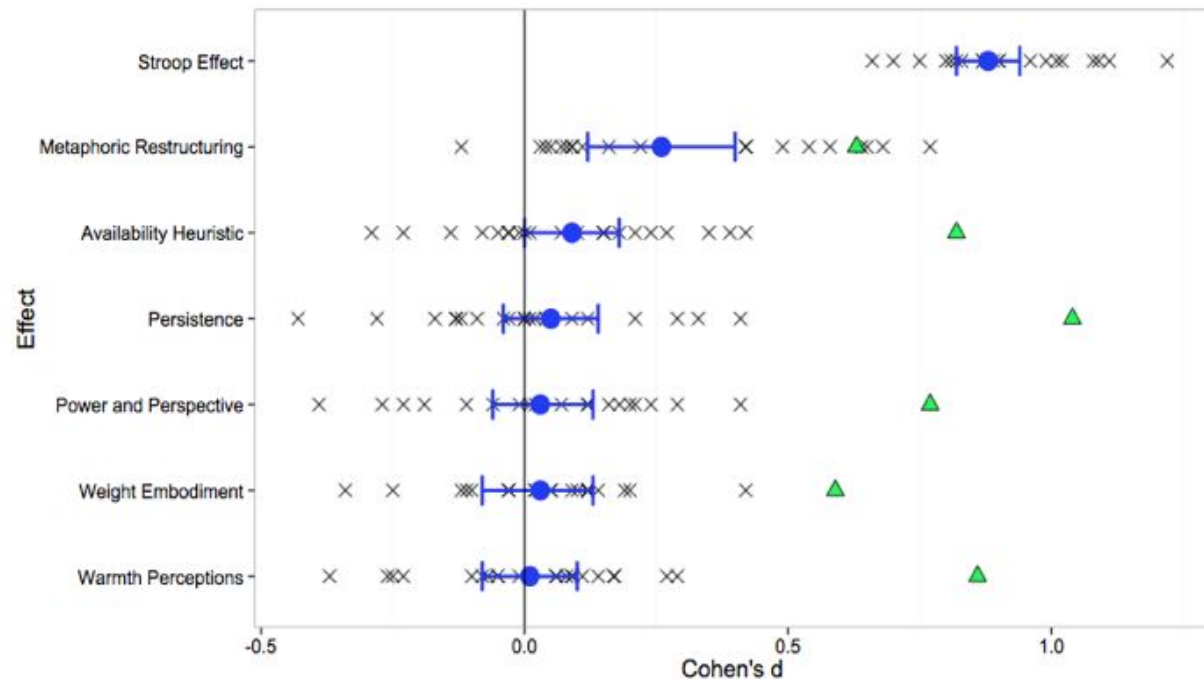
2016:
Many Labs 3

10 studies

2 waves

N=3433

7/10 did not
replicate

No support for
time-of-semester
effect

Other Many Labs Projects

- ## 2020: ML5 (N=13995)
  - Replicated (twice) studies by OSC (2015) for which original authors had expressed concerns on design before data collection: *with and without formal peer review* by original authors.

- ## 2022: ML4 (N=1550)
  - Fails to replicate mortality salience effect (a pillar of terror management theory) with and without *original author involvement* in study design
    - Original authors criticized the analysis, arguing that TMT replicates if: 1) all samples below N=60 are discarded; 2) protocols that were not advised by original authors were discarded; 3) a one-tail test (p<.10) was employed… but effect size is still 14 times smaller than in the original (d=.09 vs 1.34)!
    - Remember the «False-positive psychology» paper?

Summary of all replic. efforts (N=307)
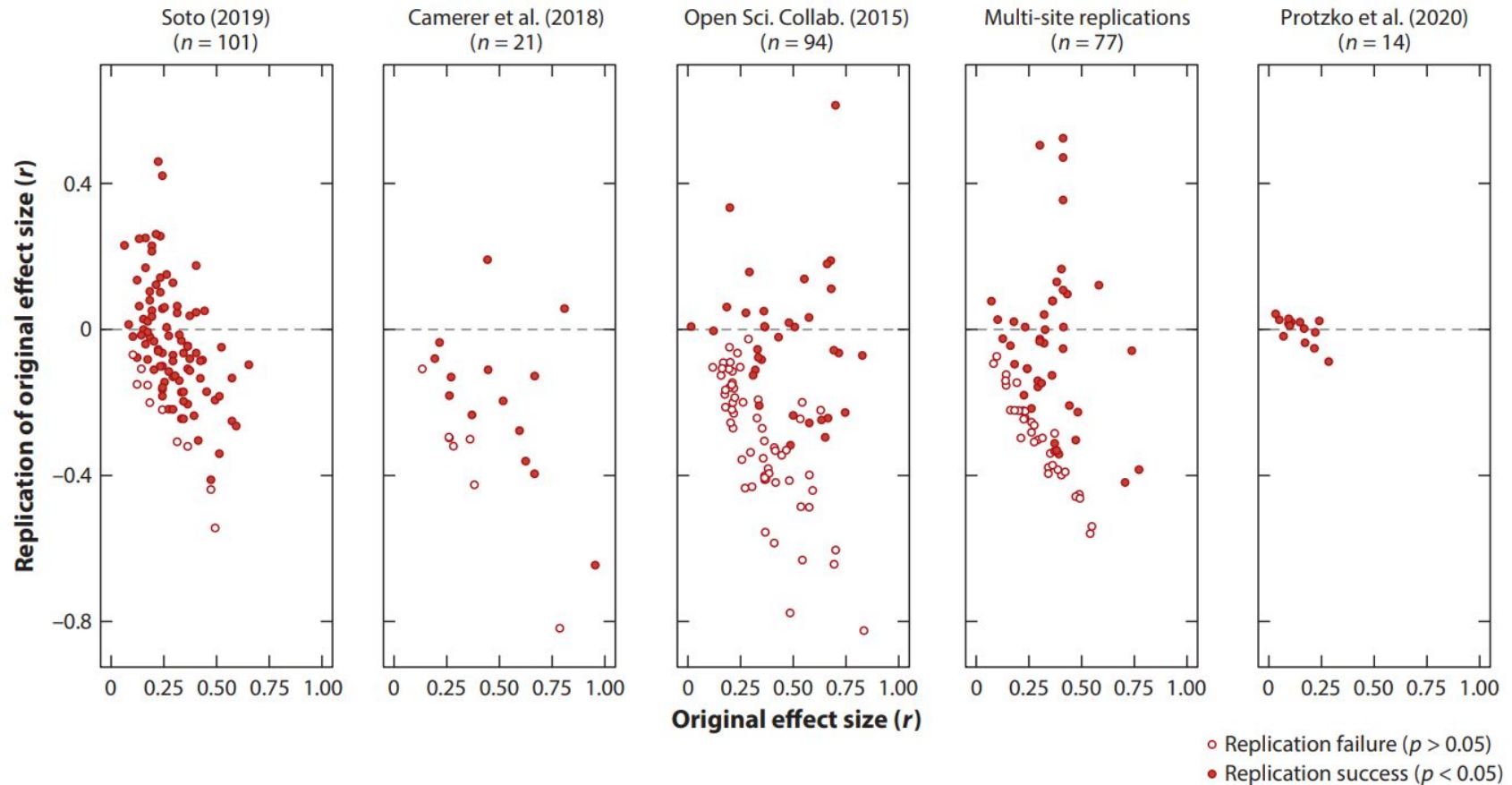
Replic. Rate: 90% 62% 36% 56% 86%

Notes: Soto (2019) investigated the personality-behavior relation; Camerer (2018) replicated 21 soc. sc. studies from Nature and Science; Protzko (2020) conducted a prospective best-practice replication study.
Source: Nosek et al., (2022).

What is the current replicability rate in Psychology?

- Considering all replications (n = 307):
  - 64% reported statistically significant evidence in the same direction,
  - effect sizes were 68% as large as in the original studies.
  - Sample size in the replication studies was on average 15.1 times the size used in the original studies.

Back to the wooclap!

Replicability in Psychology: A tale of two schools

Utopia science

- Reproducibility rate (same data, same analysis) should be close to 100%. It only depends on accepted false positive rate (e.g. 5%), integrity, openess and transparency

- In psychology, empirical reproducibility rate (same data, same analysis) is 70% (Artner et al., 2020, N=232 findings).

- Replicability rate is expected to be around 89% (Field et al., 2019; OSC, 2015;).

- Protzko (2020) provides the most likely maximum replicability rate that is empirically achievable in psychology: 86%.

How to increase replicability?
(Nosek et al., 2022)

- *#1, 2 & 3: Improve the incentive structure!*
  - Get rid of the «publish or perish» culture (Smaldino et al., 2016):
    - «key challenges are rooted in the cultural and structural features of science, non in the mind of scientists» (Nosek et al., 2021, p. 19).
  - Evaluate scientists based on methods and rigor (being right) rather than results (getting it right).
  - Devote journal space to pre-reg replications.
  - Enforce and Reward Openness and Transparency in grant proposals (ERC has made some progress) and promotions

How to increase replicability?
(Nosek et al., 2022)

- *#4:* Improve original studies!
  - *Calibrate prior odds* (currently 8.8%; Dreber et al., 2015) against the current strong preference for innovative and groundbreaking results (Fanelli, 2010)
  - *Improve the strenght of methods*: increase sample sizes, avoid selective reporting & analysis, avoid QRPs, preregister confirmatory tests, use stonger measures and manipulations, use pilots (Ioannidis, 2005; John et al., 2012; Gelman & Loken, 2013; Rosenthal, 1979; Simmons et al., 2011, Vazire, 2020; Wagenmakers, 2012)
  - *Increase transparency*! (also affects reproducibility, which is currently at 70% in Psychology; Artner t al. 2020)
 - FAIR data ☐ move toward born-open data sharing.

How to increase replicability?
(Nosek et al., 2022)

- #5: Develop better theories!
  - Mature theories define clear predictions, boundary conditions, auxiliary hypotheses (OSC, 2015).

- *#6: Improve replication studies!*
  - All that applies to original studies applies to replication studies as well, although in practice there is much less room for imrovements.
  - Fidelity to the original and skills of replicators were frequently cited as causes of replicability (e.g. Gilbert et al., 2016), but this assertion lacks any evidence (e.g. Ebersole et al., 2020).

Did something change?

- The old school does not critique/complain anymore: a sign that constructive change toward robust research practices overcame personal interests and defensive positions

- Changes can be differentiated in three categories:
  - Community change
  - Structural change
  - Procedural change

Source: Krombacher et al., 2023
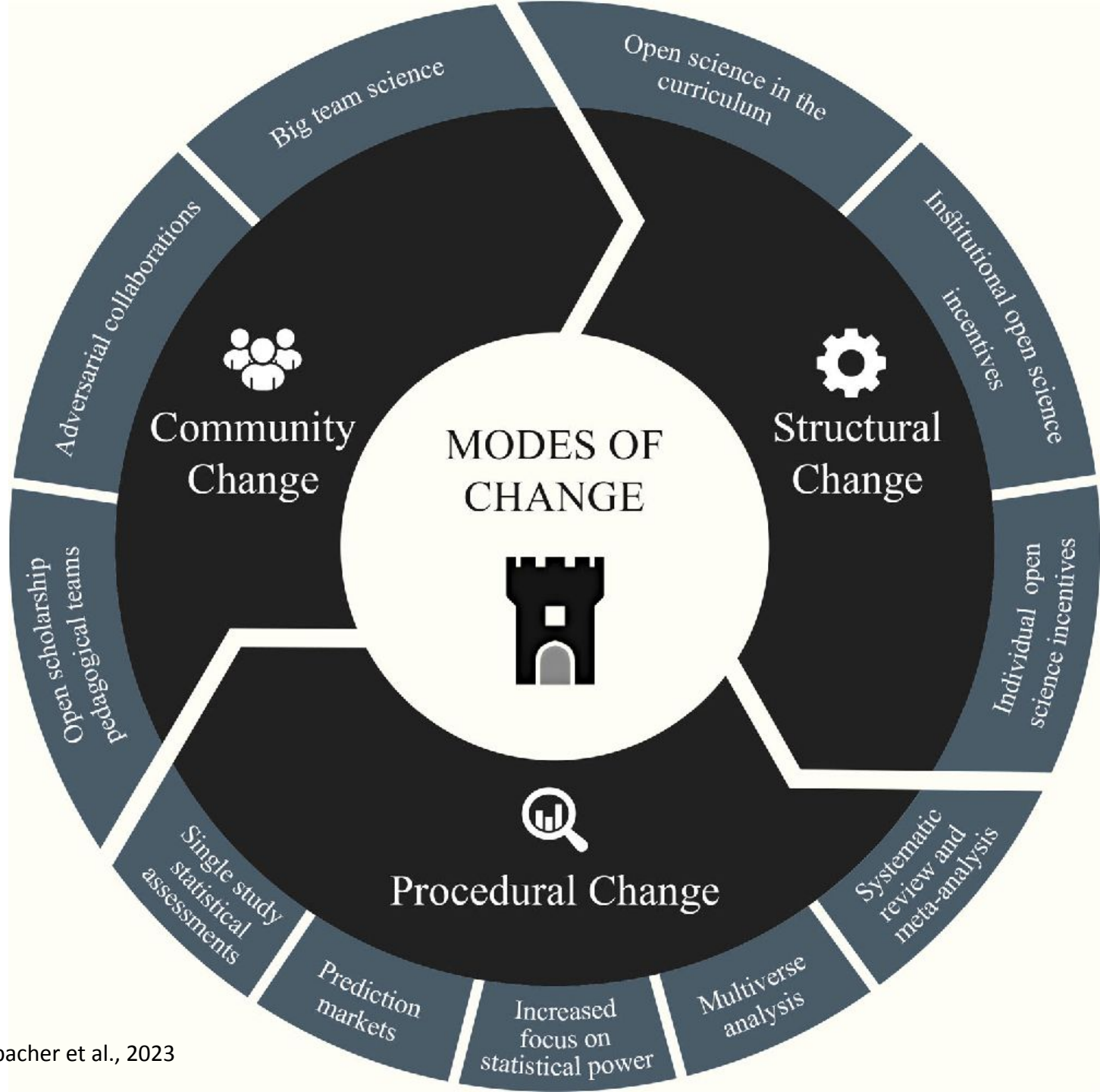
Did something change?

- TOP guidelines and TOP factor (topfactor.org): 80% of most open journals (TOP factor > 22), are in the psychology field.

- % of psych scientists who shared data was 20% in 2011, and 51% in 2017.

- Academic job offers that explicitly mentioned open science increased from 1% (2017) to 3.8% (2020; Nosek et al., 2021).

Did outcomes change?

- Trends cannot be detected in average self-reported use of *QRPs* among 14 surveys because comparability is low, but numbers seems stable around 10% (Fanelli, 2009; Gopalakrishna et al., 2022)

- No evidence that sample size increased from 2011 to 2021 (Nosek et al., 2021)

Conclusions

- The change has started, but there is still much to do
- Institutions: you are lagging behind in promotion policies: reward *being* right rather than *getting it* right!
- Policy makers: *decrease pressure to publish*! (Gopalakrishna et al., 2022)
- Scientist: care A LOT about *theories* & use Open and Transparent Research Practices.
- *Love Science*, not career. It pays off for society and (arguably) for the individual as well

All our science, measured against reality, is primitive and childlike—and yet it is the most precious thing we have.
(A. Einstein)

Back to the wooclap!

Replicability in Psychology: A tale of two schools

Grazie!

Useful resources

- Checklist on Open Research
- Guide to Open Science: Kathawalla et al. (2021)
- Guide to Registered Reports: Henderson (2022)
- Authorship: Coles et al., (2023)
- Peer review Guidelines: Davis et al. (2018)
- Codebook guidelines: Horstmann et al., (2020)
- Commented bibliography: Bastart et al. (2018)

Useful resources

- https://forrt.org/resources/ (perhaps the largest knowledge base)
- http://forrt.org/reversals (check if effect is replicable)
- https://www.cos.io/blog/educating-in-the-open-scholarship-eara
- https://osf.io/wfc6u/
- https://curatescience.org/
- https://guides.lib.uw.edu/research/reproducibility/teaching
- https://en.unesco.org/science-sustainable-future/open-science/recommendation
- https://www.crep-psych.org/
- https://itrn.org