Kredibilita spoločenskovedného výskumu:
Úloha otvorenej vedy a meta-výskumu

Matúš Adamkovič

Centrum sociálnych a psychologických vied SAV

University of Jyväskylä, Fínsko

matho.adamkovic@gmail.com | matus.m.adamkovic@jyu.fi

APVV-22-0458

O mne...

- Metodológia spoločenskovedného výskumu
- Meta-výskum
- Otvorená veda
- Psychopatológia a sieťový prístup

Replikačná kríza (nielen) v sociálnych vedách

- Zopakovanie (replikovanie) štúdie často vedie k rôznym výsledkom
  - Známe efekty sú častokrát nadhodnotené
  - Niektoré fenomény možno ani neexistujú
  - Nízka hodnovernosť poznatkov

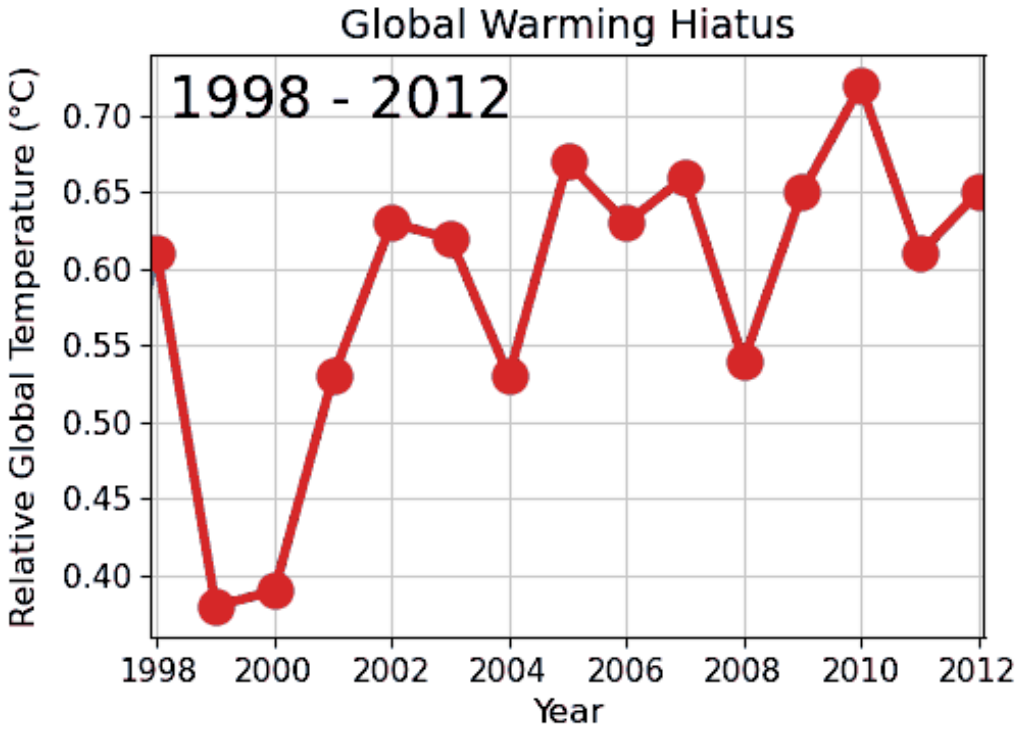
- Príčiny
  - Systém incentív v akademickom svete (publish or perish kultúra)
  - Politiky časopisov (nové a prelomové zistenia)
  - Questionable research practices (cherry picking, p-hacking, HARK-ing…)
  - Neuchopiteľné fenomény (napr. “measuring” the mind)
  - Ľudské skreslenia (napr. konfirmačné skreslenie)
  - Nízka transparentnosť
  - Kolektívna apatia

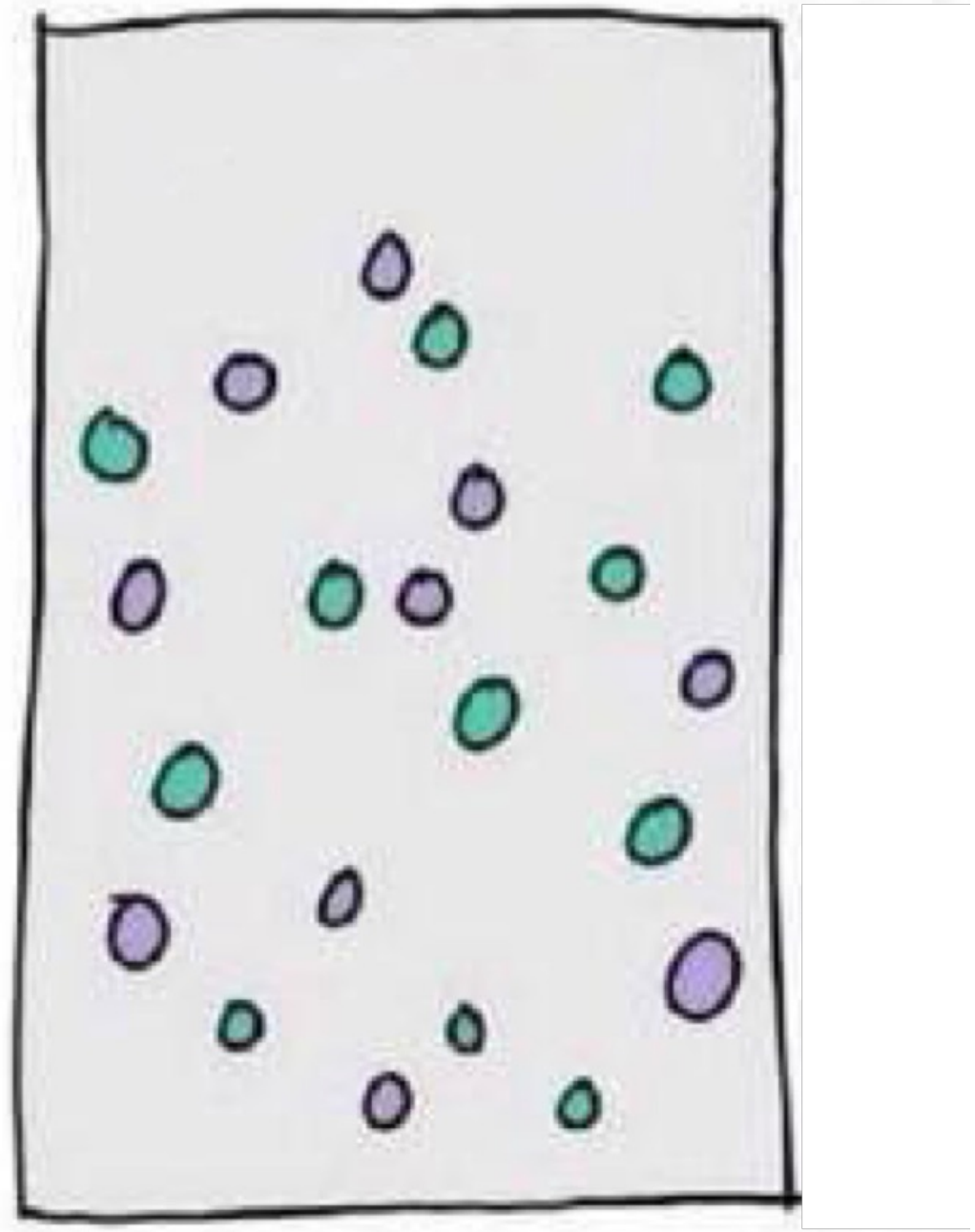
Table 1. Likelihood of Obtaining a False-Positive Result

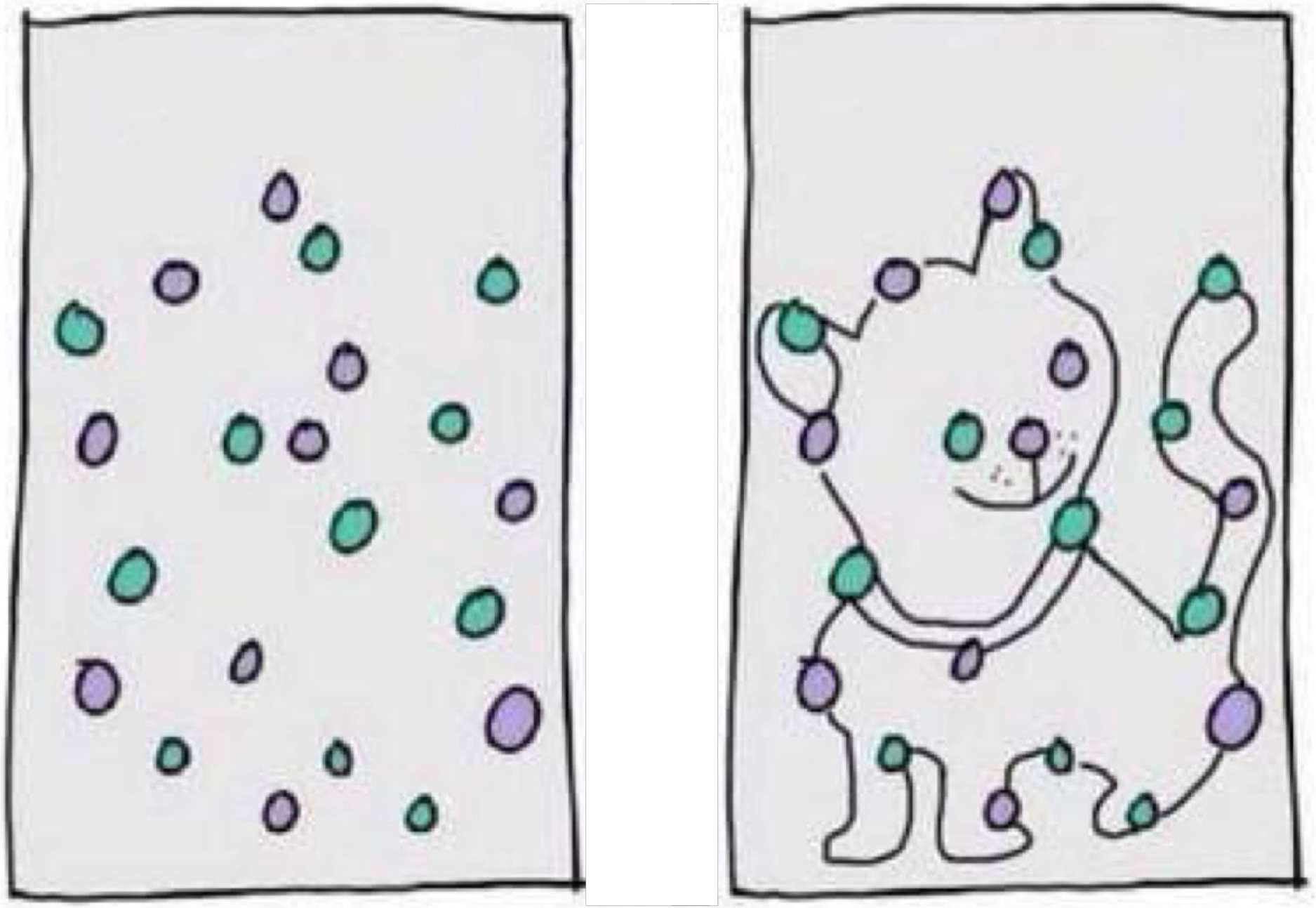
Simmons et al. (2011)

Researcher degrees of freedom	Significance level $p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

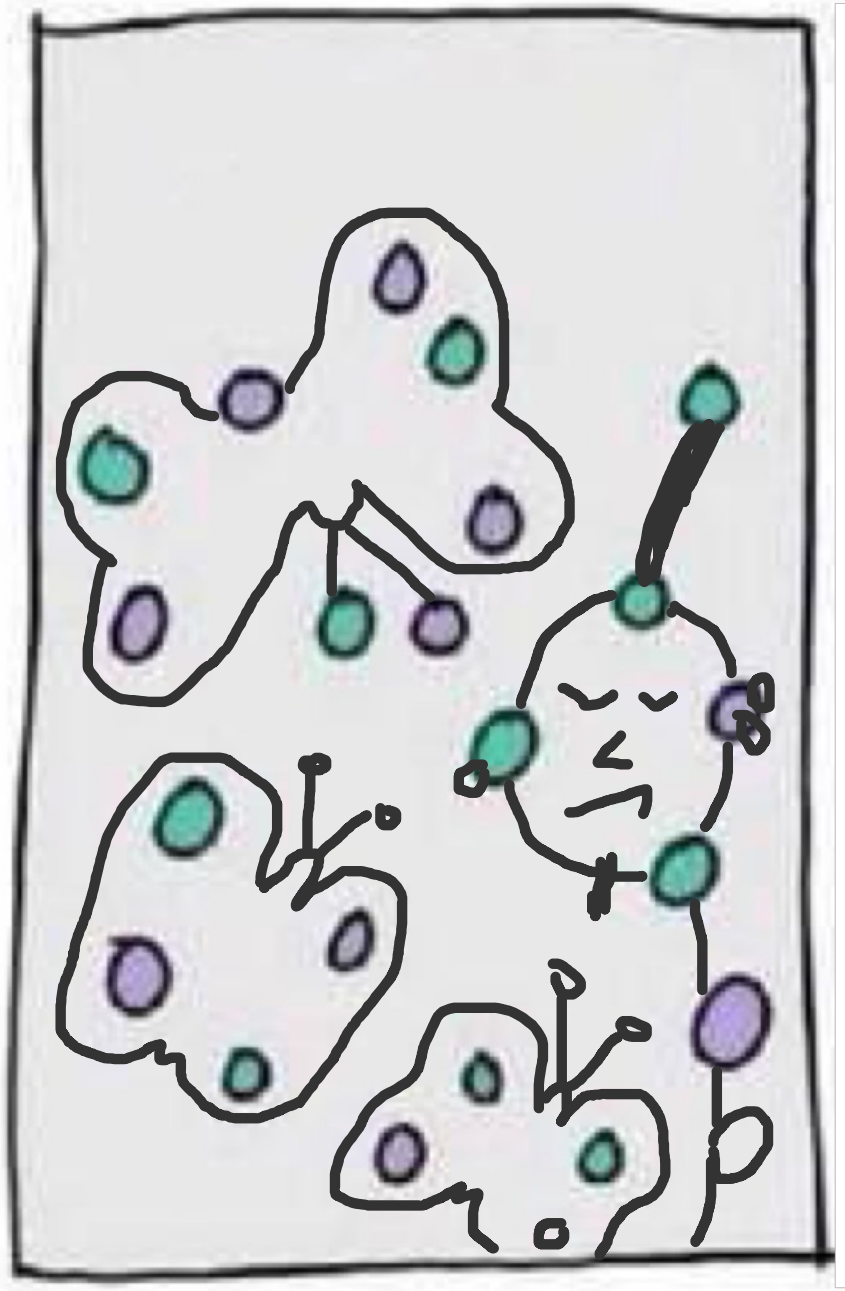
"If you don't reveal some insights soon, I'm going to be forced to slice, dice, and drill!"

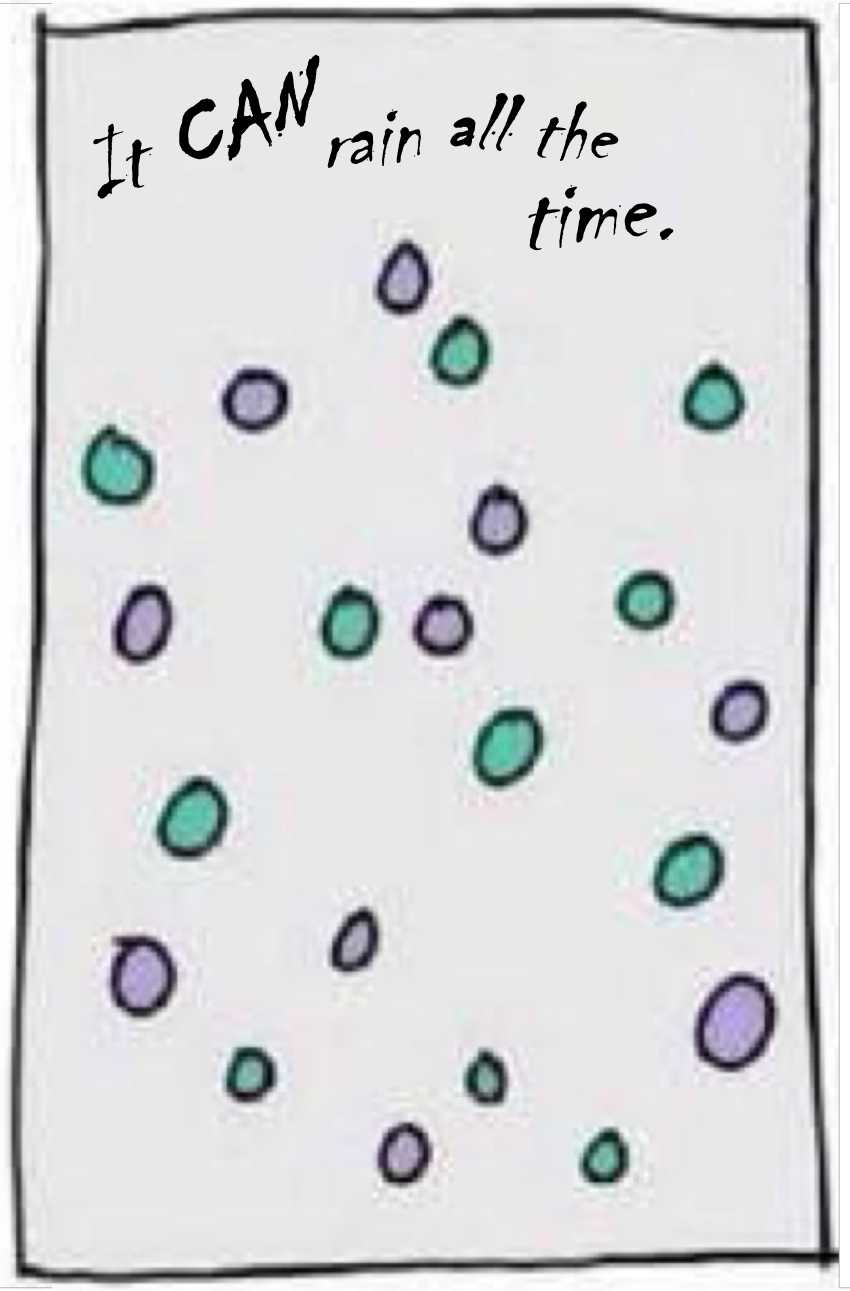


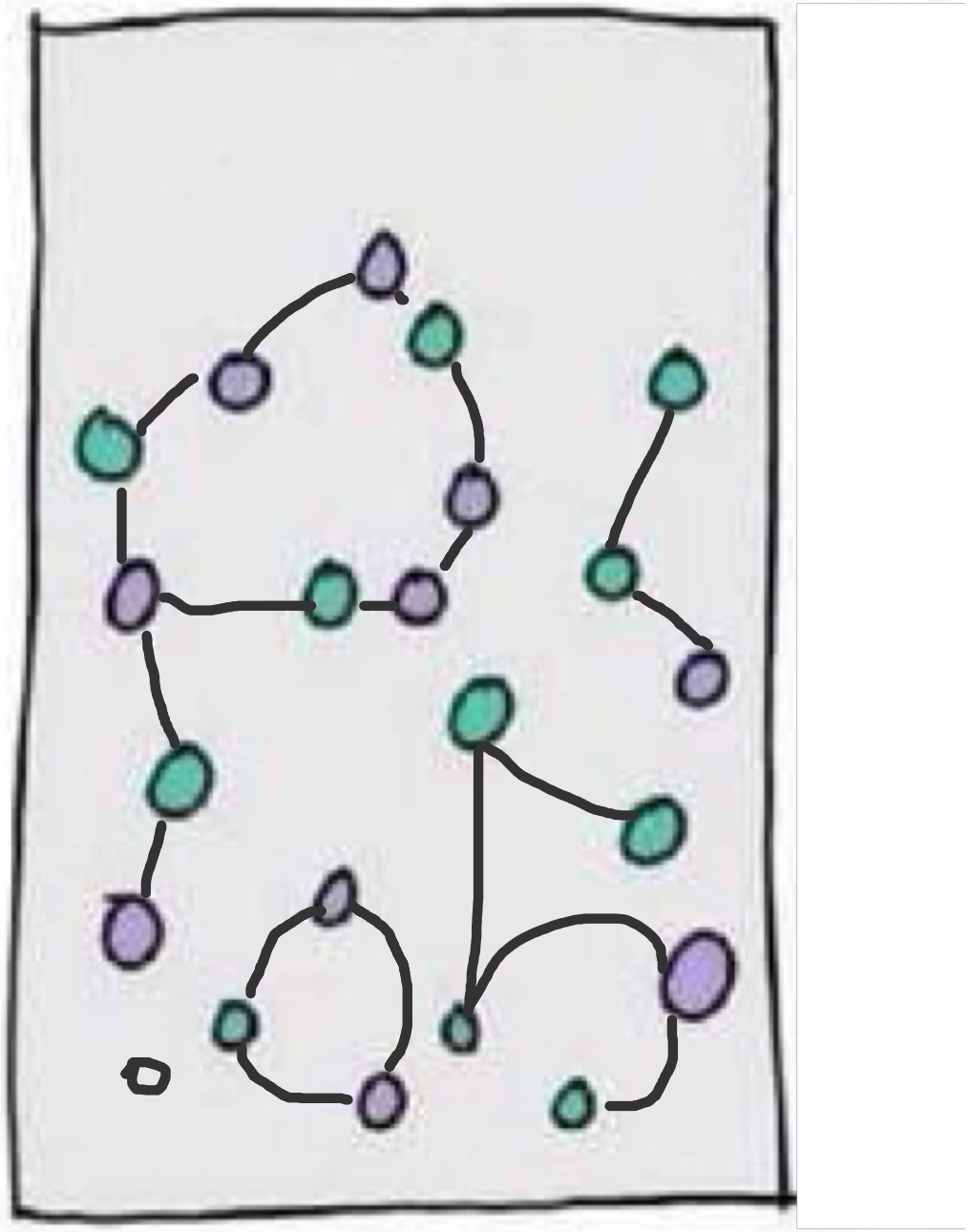


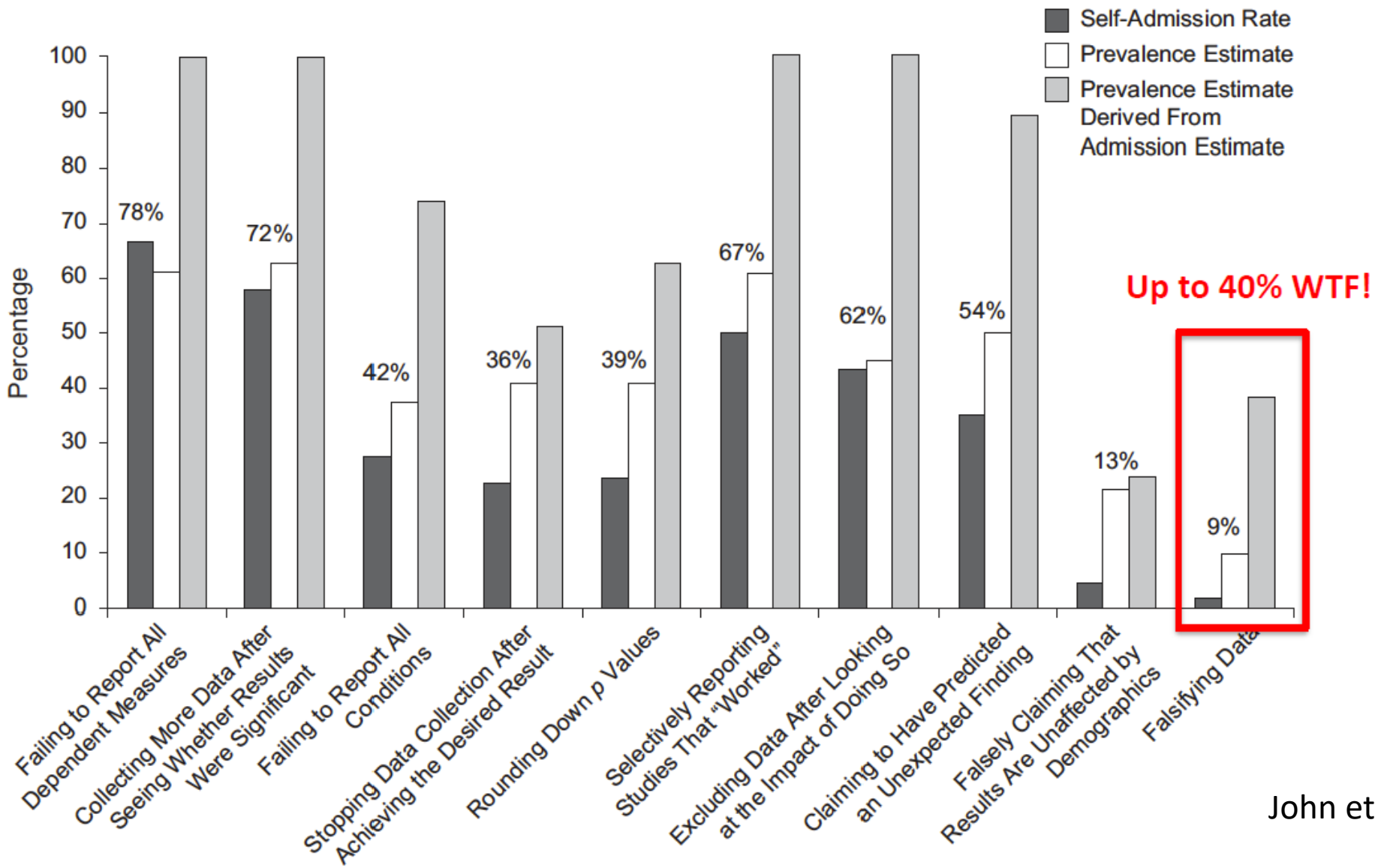


Credit: Daniel Lakens




It CAN rain all the time.

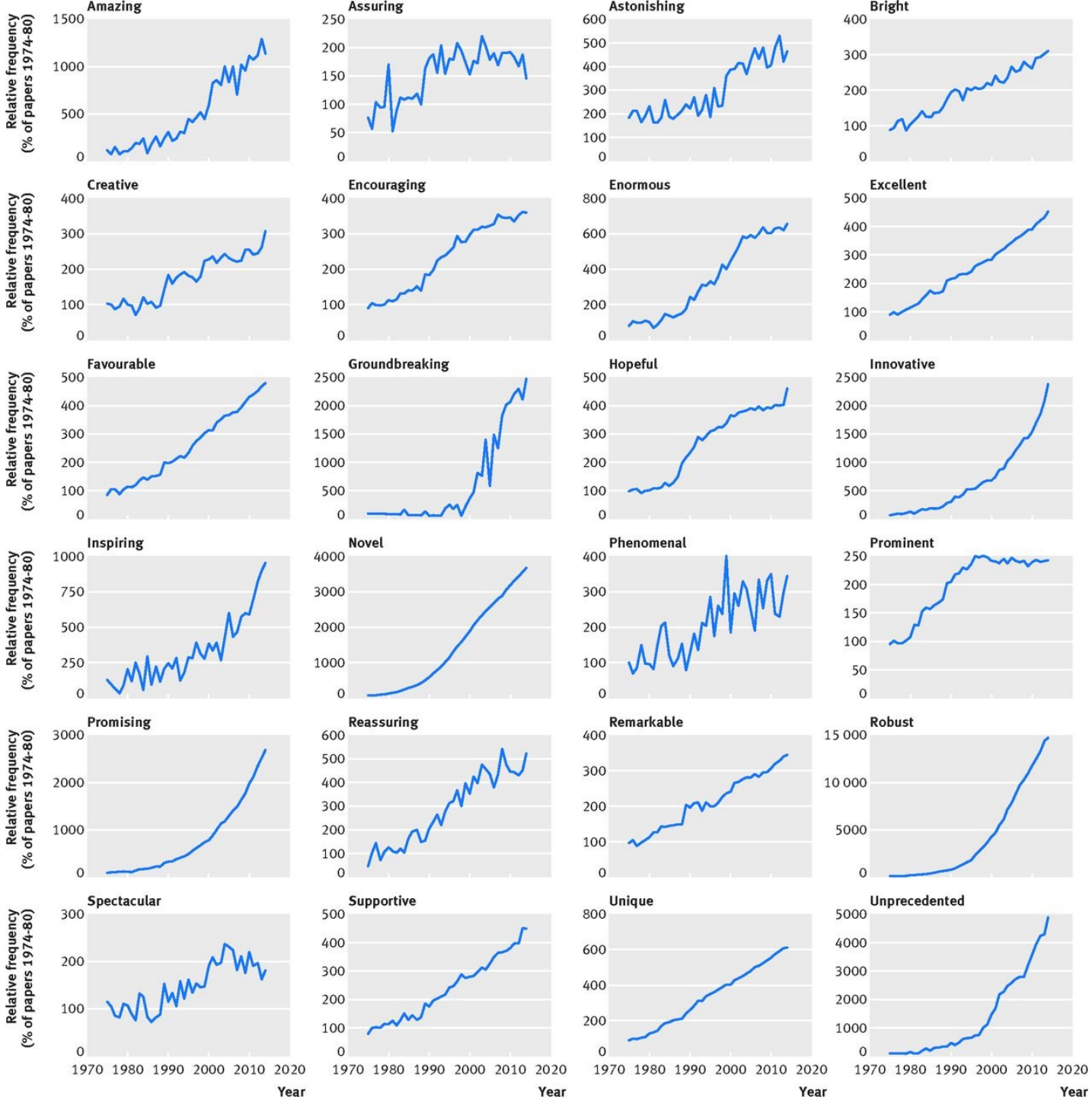




John et al. (2012)

Fig. 1. Results of the Bayesian-truth-serum condition in the main study. For each of the 10 items, the graph shows the self-admission rate, prevalence estimate, prevalence estimate derived from the admission estimate (i.e., self-admission rate/admission estimate), and geometric mean of these three percentages (numbers above the bars). See Table 1 for the complete text of the items.

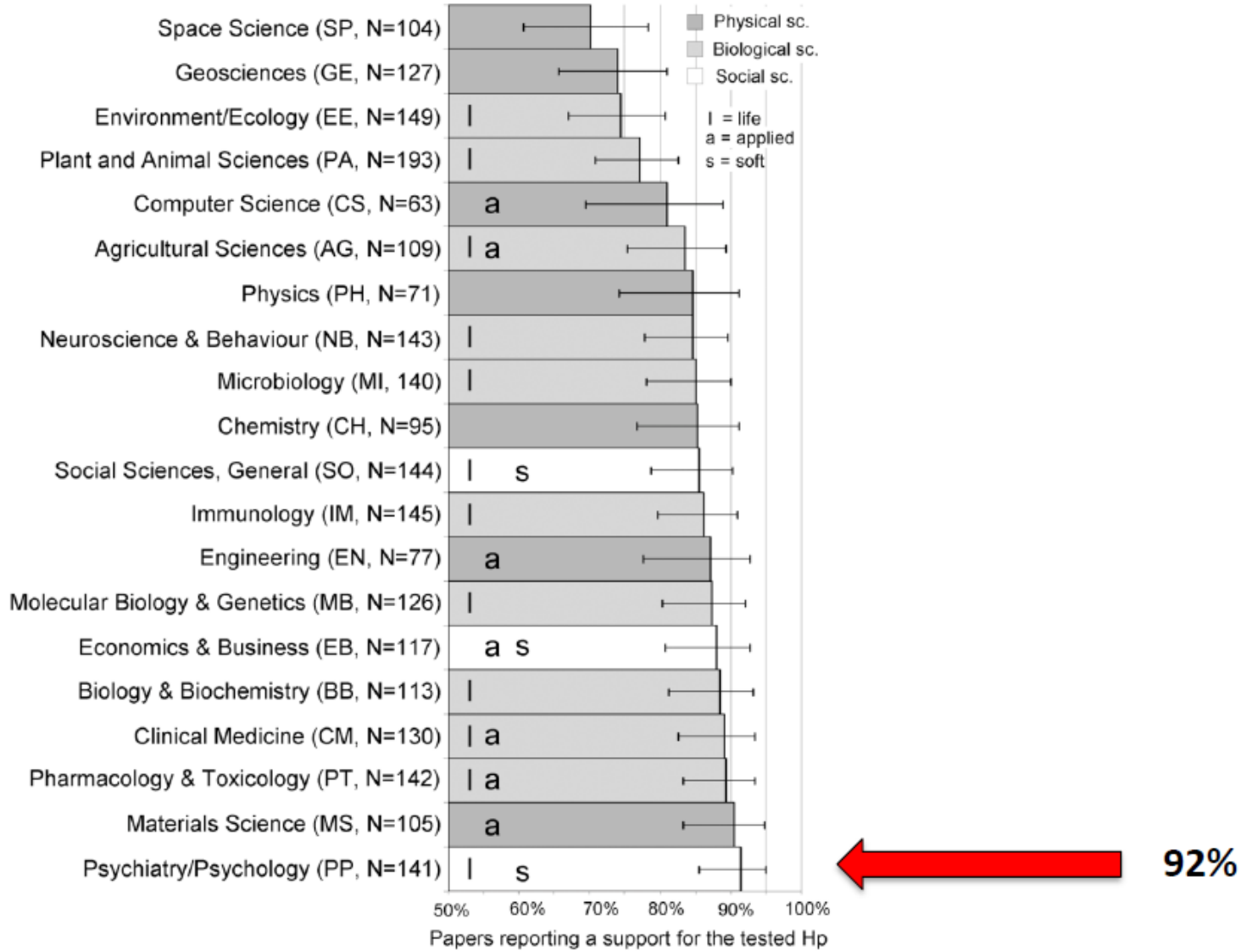
- Relatívna frekvencia pozitívnych slov (1974-2014)
- V priemere +880%
- robust, novel, innovative, unprecedented, unique (Vinkers et al., 2015)
- Negatívne zistenia miznú z publikovanej literatúry (Fanelli, 2011)



"Positive" Results Increase Down the Hierarchy of the Sciences

Daniele Fanelli*

INNOGEN and ISSTI-Institute for the Study of Science, Technology & Innovation, The University of Edinburgh, Edinburgh, United Kingdom



92%

Scheel et al. (2020)

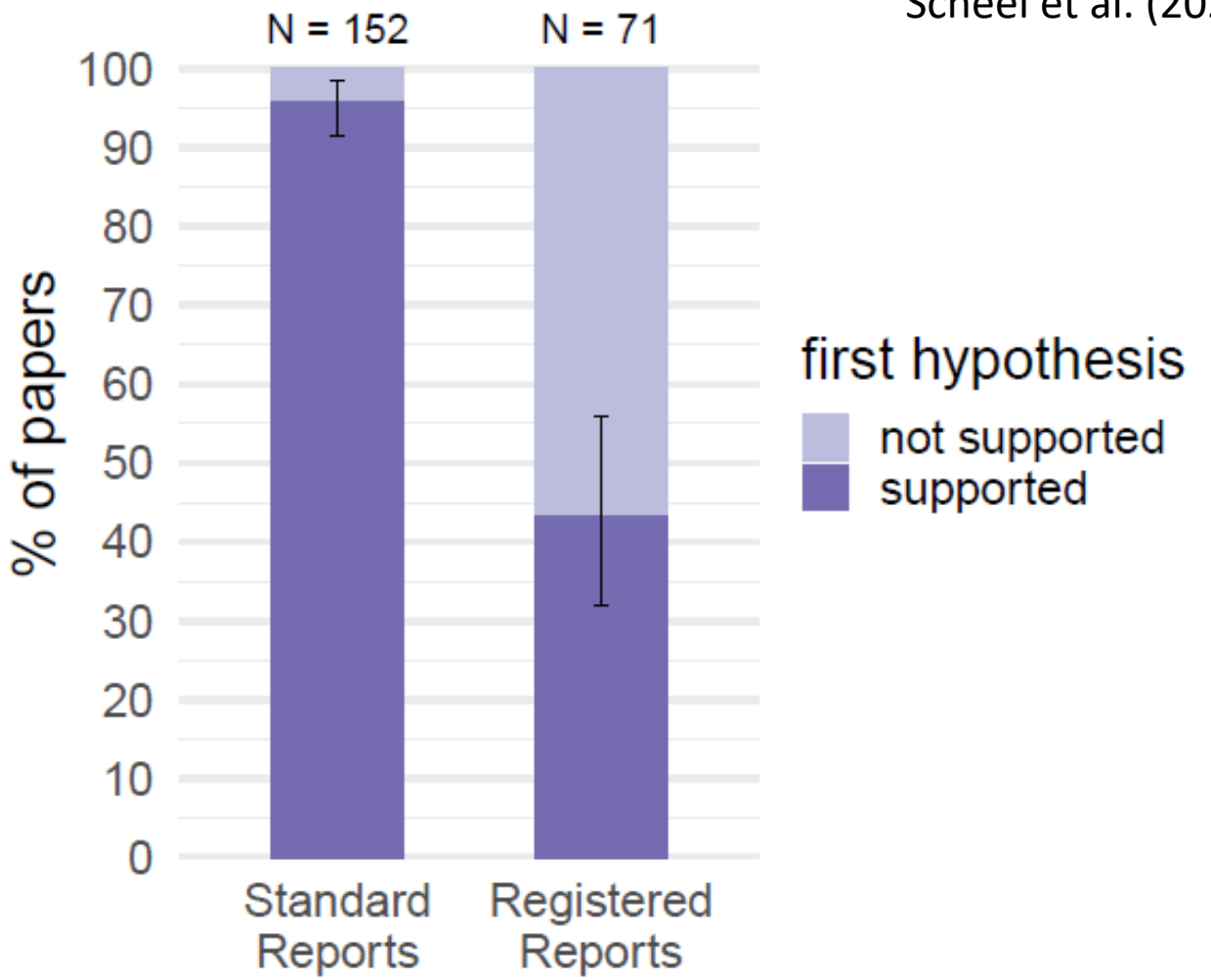
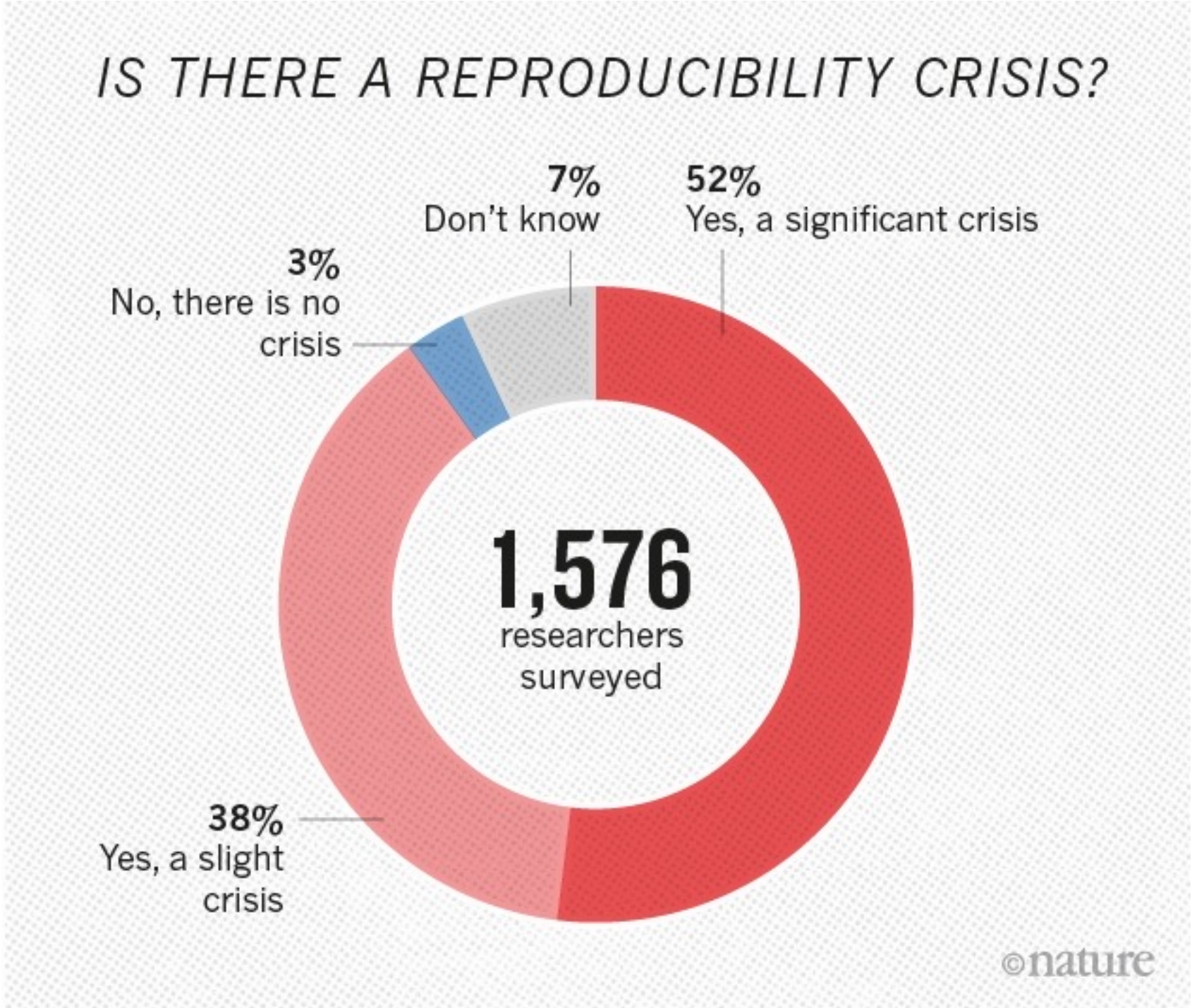


Figure 2. Positive result rates for standard reports and Registered Reports. Error bars indicate 95% confidence intervals around the observed positive result rate.


IS THERE A REPRODUCIBILITY CRISIS?
7%
Don't know
52%
Yes, a significant crisis
3%
No, there is no crisis
1,576
researchers surveyed
38%
Yes, a slight crisis
©nature

RESEARCH ARTICLE

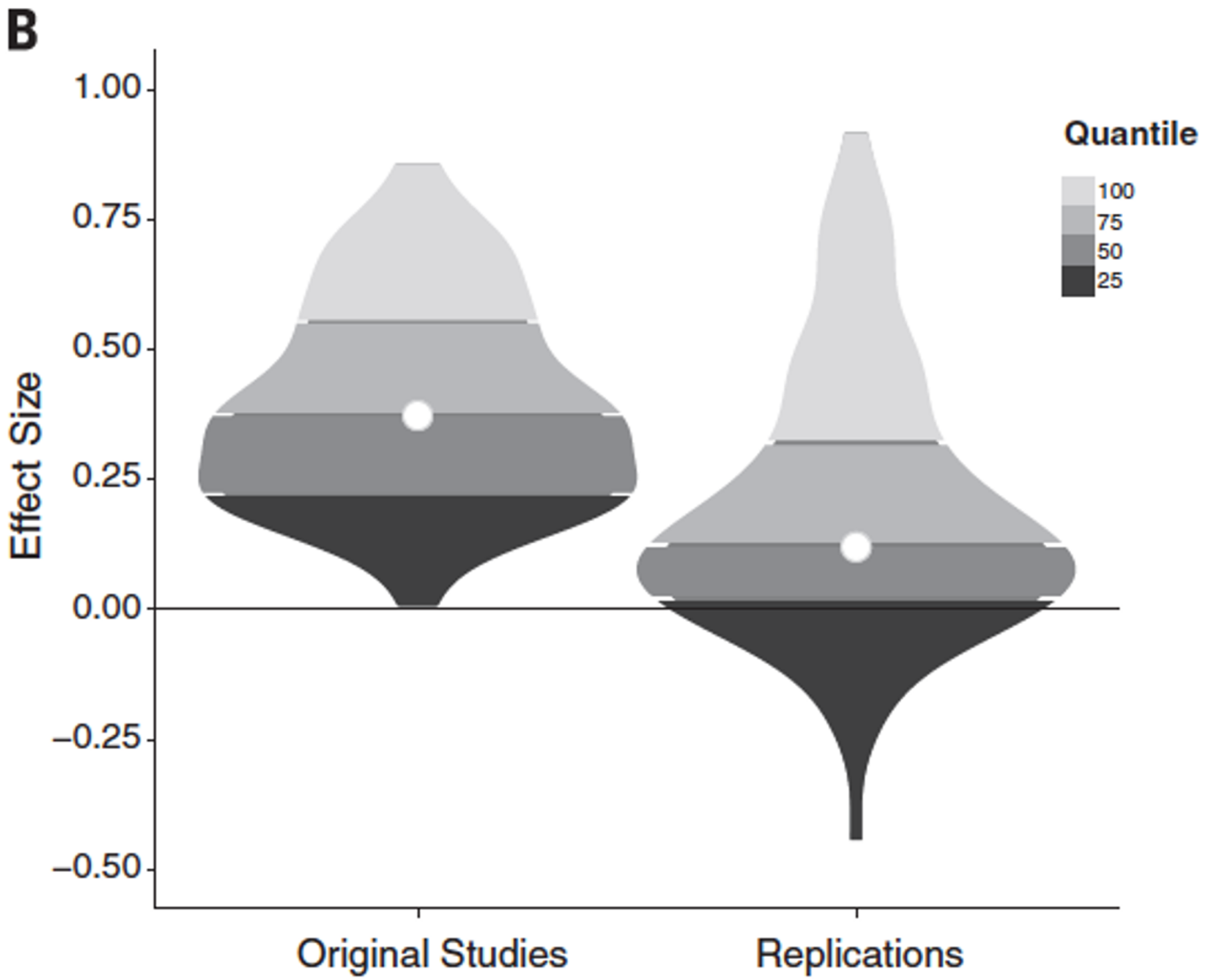
PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*†

Open Science Collaboration (2015). Science.

The effect sizes in replication studies were much lower, some of them were even in the opposite direction.

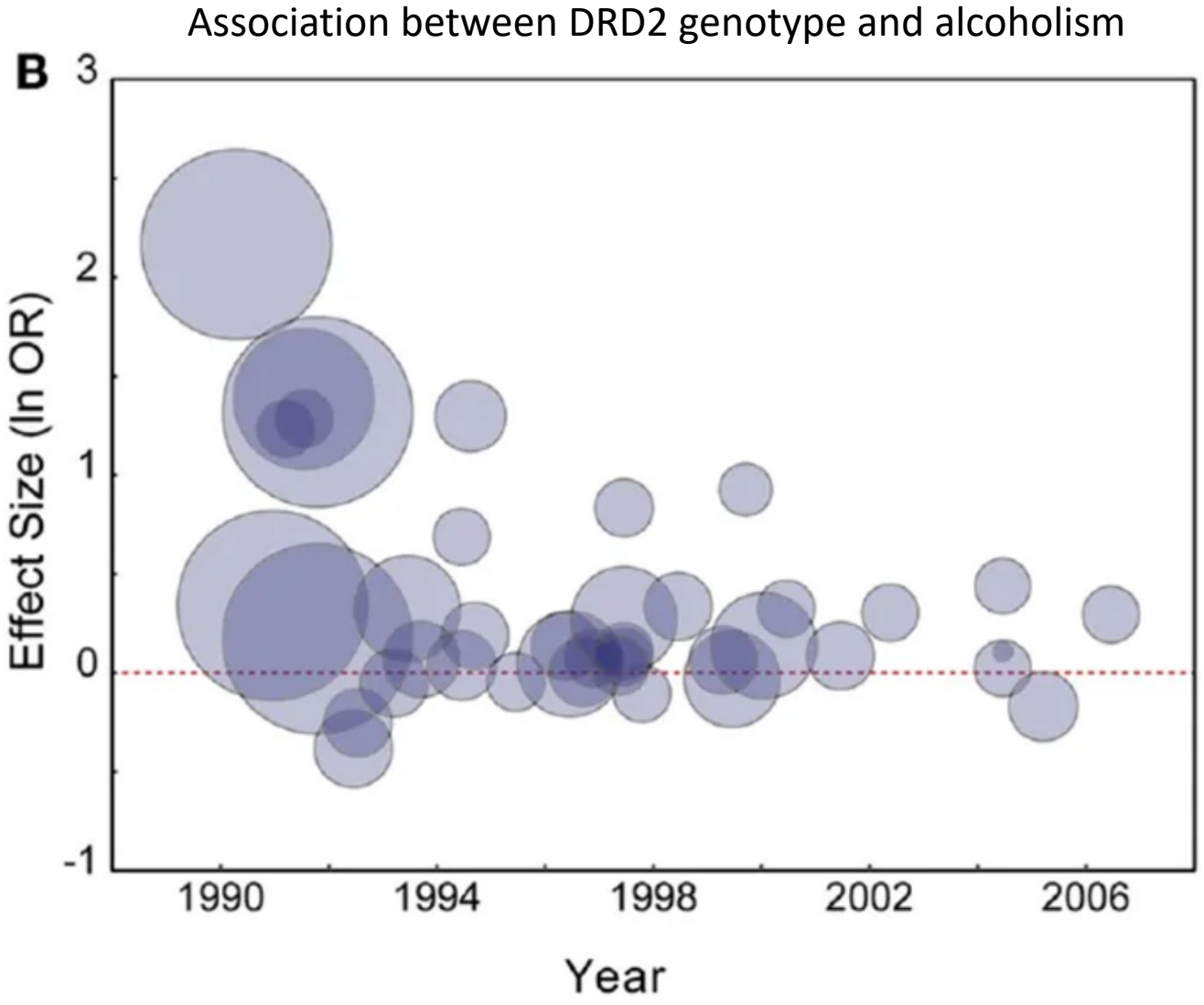


Brembs et al. (2013)

Effect size is negatively correlated with year of publication.

Note: the bigger the circle, the higher the journal's impact factor.

Initial discoveries are "groundbreaking", but *somehow* the effect size (of the same phenomenon!) gets lower with replication.



ESSAY

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • https://doi.org/10.1371/journal.pmed.0020124

Abstract

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Publication bias in the social sciences: Unlocking the file drawer

Annie Franco[1], **Neil Malhotra**[2,*], **Gabor Simonovits**[1]

+ See all authors and affiliations

Science 19 Sep 2014:
Vol. 345, Issue 6203, pp. 1502-1505
DOI: 10.1126/science.1255484

The file drawer is full. Should we worry?

Experiments that produce null results face a higher barrier to publication than those that yield statistically significant differences. Whether this is a problem depends on how many null but otherwise valid results might be trapped in the file drawer. Franco *et al.* use a Time-sharing Experiments in the Social Sciences archive of nearly 250 peer-reviewed proposals of social science experiments conducted on nationally representative samples. They find that only 10 out of 48 null results were published, whereas 56 out of 91 studies with strongly significant results made it into a journal.

Science, this issue p. **1502**

Abstract

We studied publication bias in the social sciences by analyzing a known population of conducted studies—221 in total—in which there is a full accounting of what is published and unpublished. We leveraged Time-sharing Experiments in the Social Sciences (TESS), a National Science Foundation–sponsored program in which researchers propose survey-based experiments to be run on representative samples of American adults. Because TESS proposals undergo rigorous peer review, the studies in the sample all exceed a substantial quality threshold. Strong results are 40 percentage points more likely to be published than are null results and 60 percentage points more likely to be written up. We provide direct evidence of publication bias and identify the stage of research production at which publication bias occurs: Authors do not write up and submit null findings.

nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-019-0787-z

Comparing meta-analyses and preregistered multiple-laboratory replication projects

Amanda Kvarven[1,3], Eirik Strømland[1,3] and Magnus Johannesson[2*]

Many researchers rely on meta-analysis to summarize research evidence. However, there is a concern that publication bias and selective reporting may lead to biased meta-analytic effect sizes. We compare the results of meta-analyses to large-scale preregistered replications in psychology carried out at multiple laboratories. The multiple-laboratory replications provide precisely estimated effect sizes that do not suffer from publication bias or selective reporting. We searched the literature and identified 15 meta-analyses on the same topics as multiple-laboratory replications. We find that meta-analytic effect sizes are significantly different from replication effect sizes for 12 out of the 15 meta-replication pairs. These differences are systematic and, on average, meta-analytic effect sizes are almost three times as large as replication effect sizes. We also implement three methods of correcting meta-analysis for bias, but these methods do not substantively improve the meta-analytic results.

NEWS | 12 October 2023

Reproducibility trial: 246 biologists get different results from same data sets

Wide distribution of findings shows how analytical choices drive conclusions.

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

CENTER FOR OPEN SCIENCE

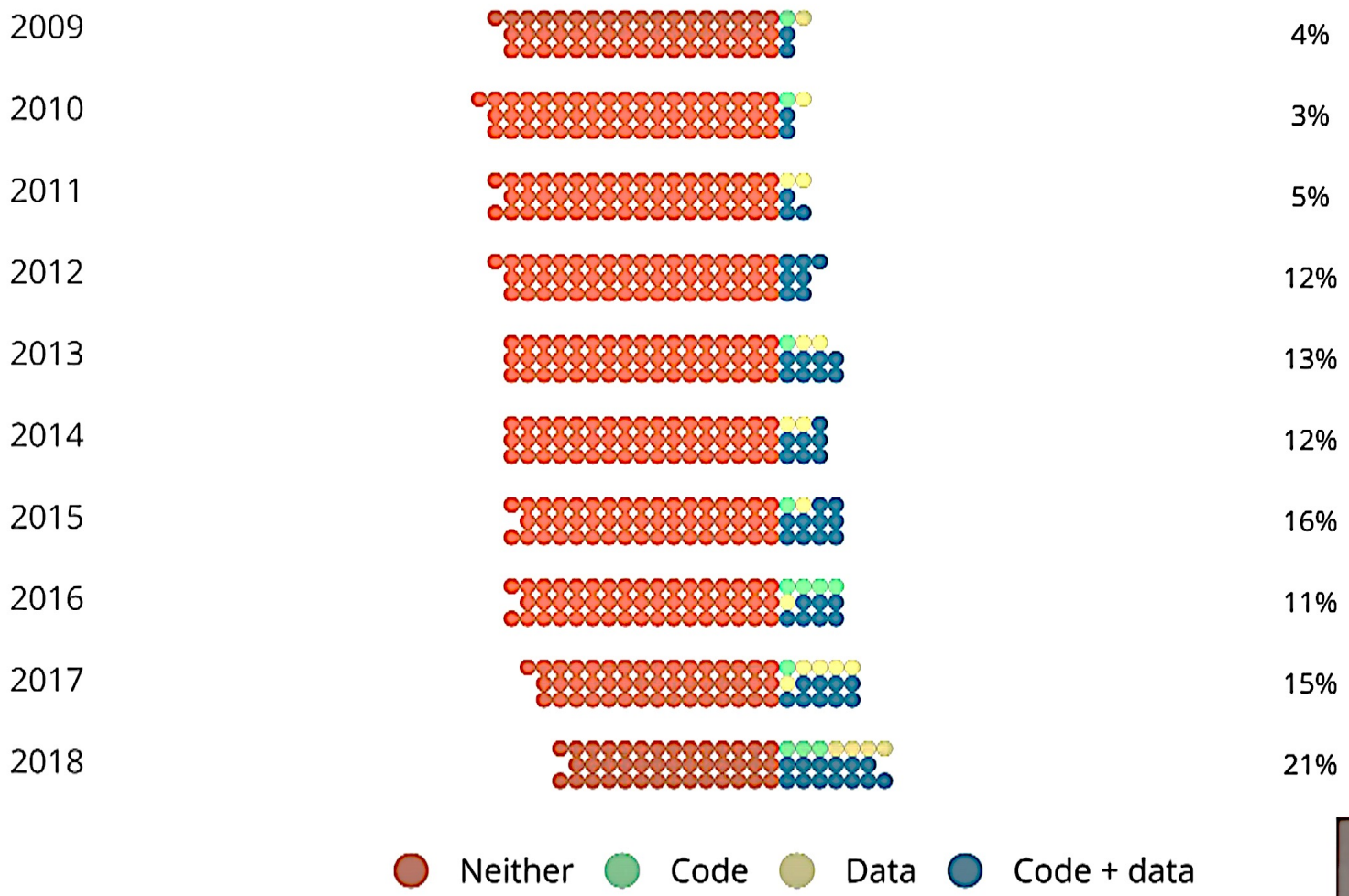
Systematizing Confidence in Open Research and Evidence (SCORE) (Archived)

The Department of Defense (DoD) often leverages social and behavioral science (SBS) research to design plans, guide investments, assess out and build models of human social systems and behaviors as they relate to national security challenges in the human domain. However, a numbe recent empirical studies and meta-analyses have revealed that many SBS results vary dramatically in terms of their ability to be independently reproduced or replicated, which could have real-world implications for DoD's plans, decisions, and models. To help address this situation, DARPA Systematizing Confidence in Open Research and Evidence (SCORE) program aims to develop and deploy automated tools to assign "confidenc scores" to different SBS research results and claims. Confidence scores are quantitative measures that should enable a DoD consumer of SBS research to understand the degree to which a particular claim or result is likely to be reproducible or replicable. These tools will assign explainabl confidence scores with a reliability that is equal to, or better than, the best current human expert methods. If successful, SCORE will enable DoD personnel to quickly calibrate the level of confidence they should have in the reproducibility and replicability of a given SBS result or claim, and th increase the effective use of SBS literature and research to address important human domain challenges, such as enhancing deterrence, enablir stability, and reducing extremism.

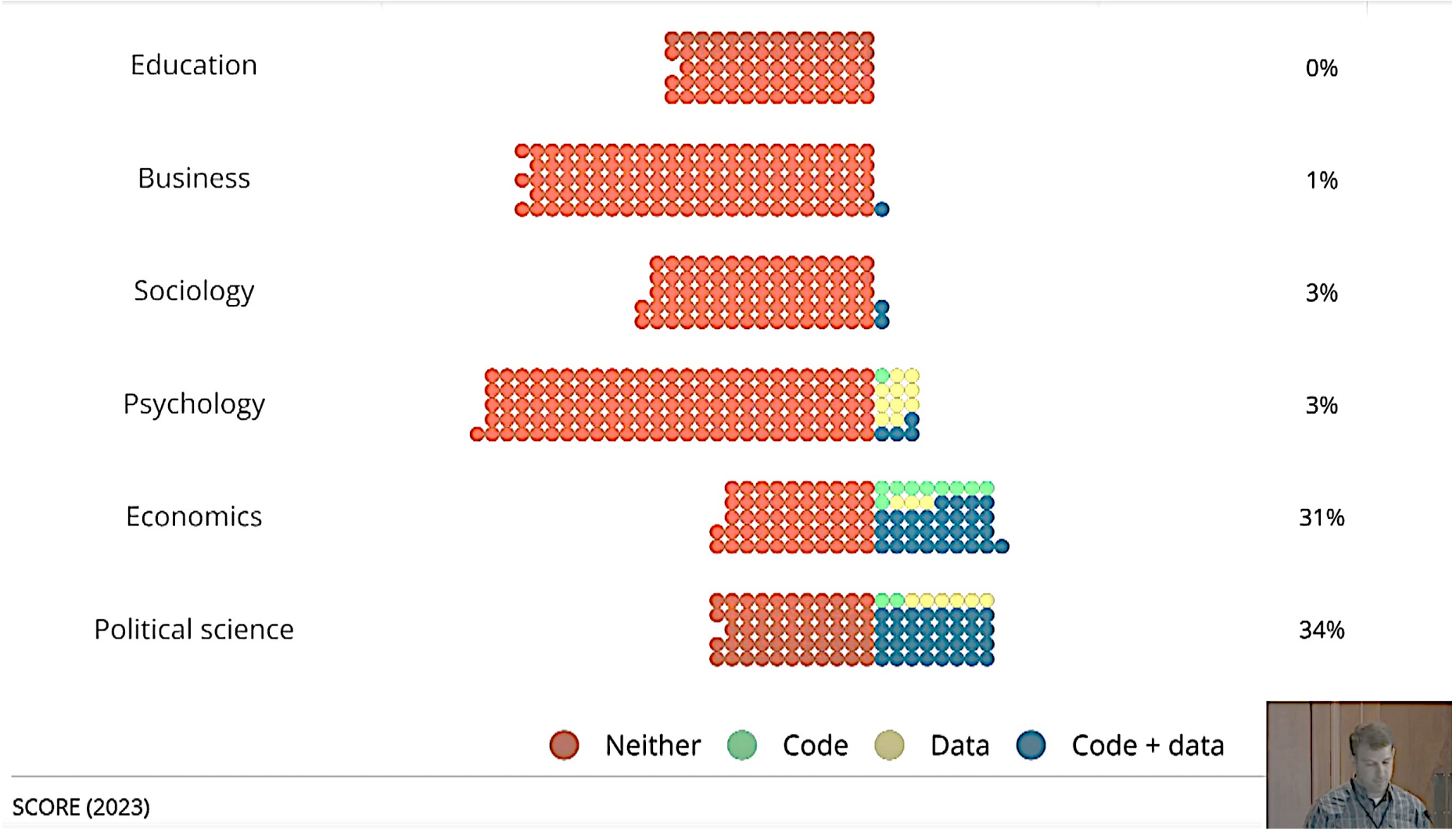
SCORE

- Začiatok v 2019
- Pravdepodobne najväčší meta-výskumný projekt
- Vytvorenie automatizovaného nástroja na rýchle a presné hodnotenie dôveryhodnosti výsledkov výskumu
- Skúmali sa:
  - Reprodukovateľnosť (rovnaké výsledky za použitia rovnakých dát a analýzy)
  - Robustnosť (rovnaké výsledky za použitia rovnakých dát, ale rozdielnych analýz)
  - Replikovateľnosť (rovnaké výsledky za použitia nových dáť)
  - Predikcie expertov/iek
- Momentálne sa projekt finalizuje
  - Validácia/kontrola zistení

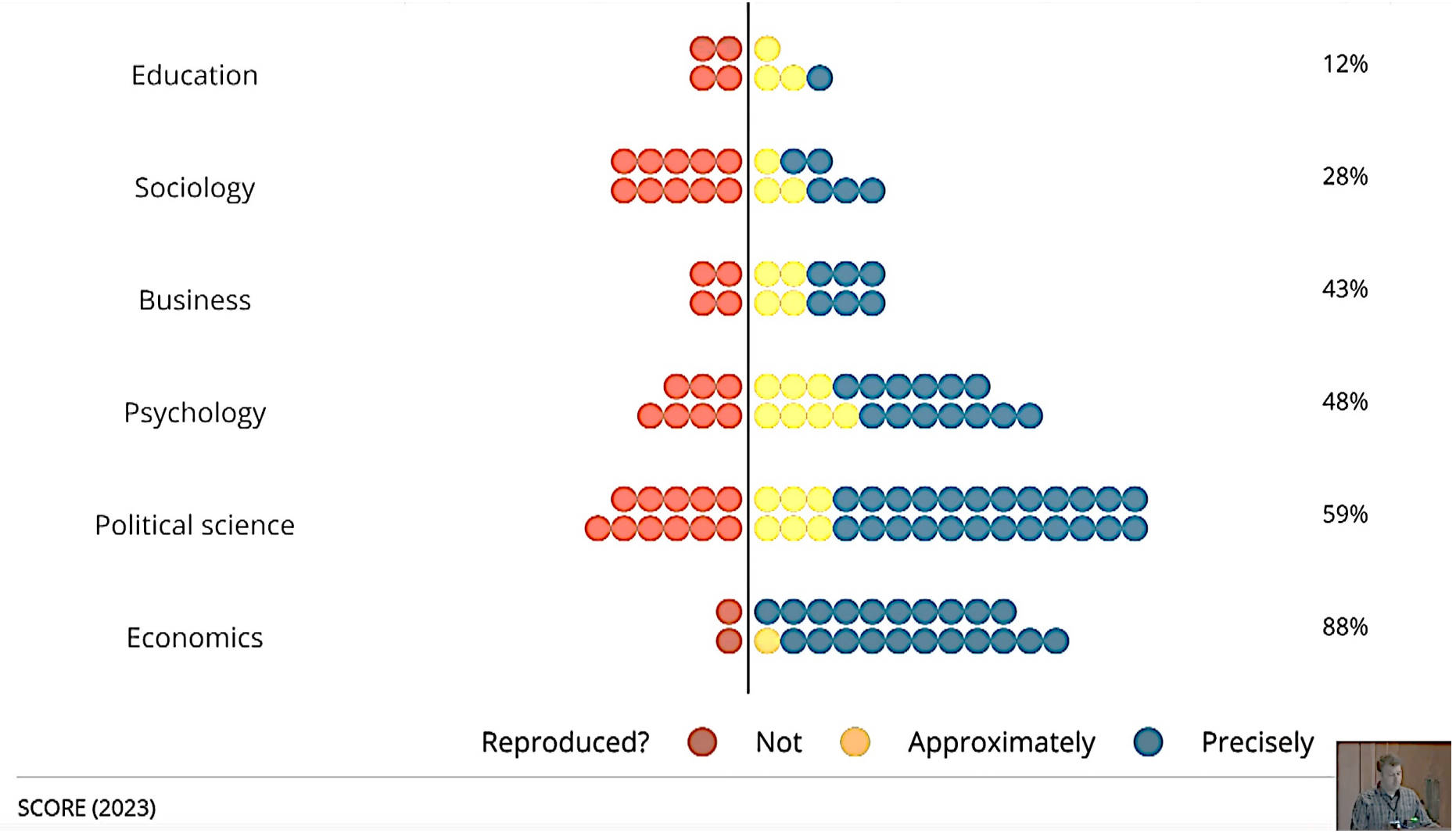
SCORE – zdieľanie dát (predbežné výsledky)



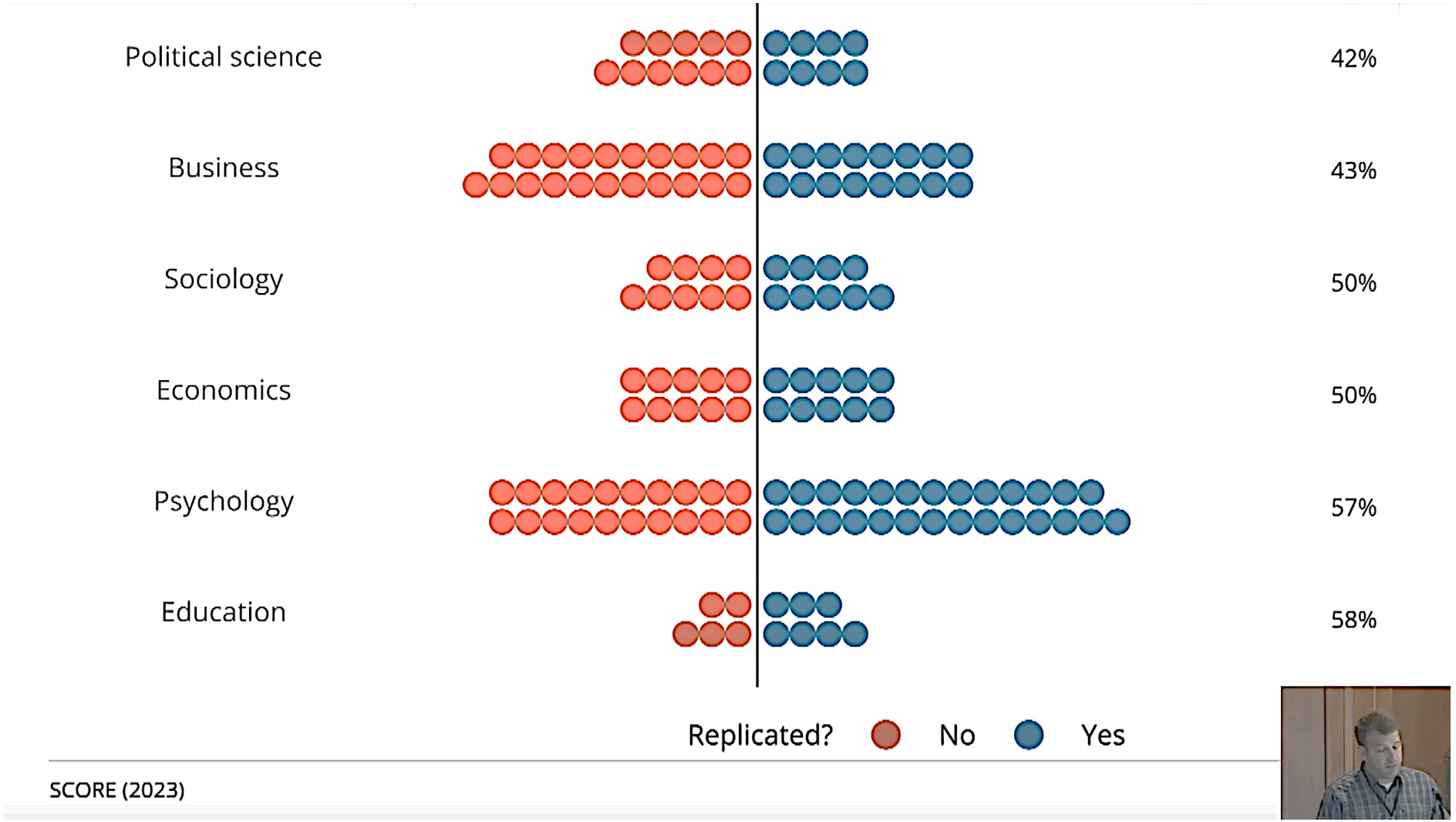
SCORE – zdieľanie dát (predbežné výsledky)

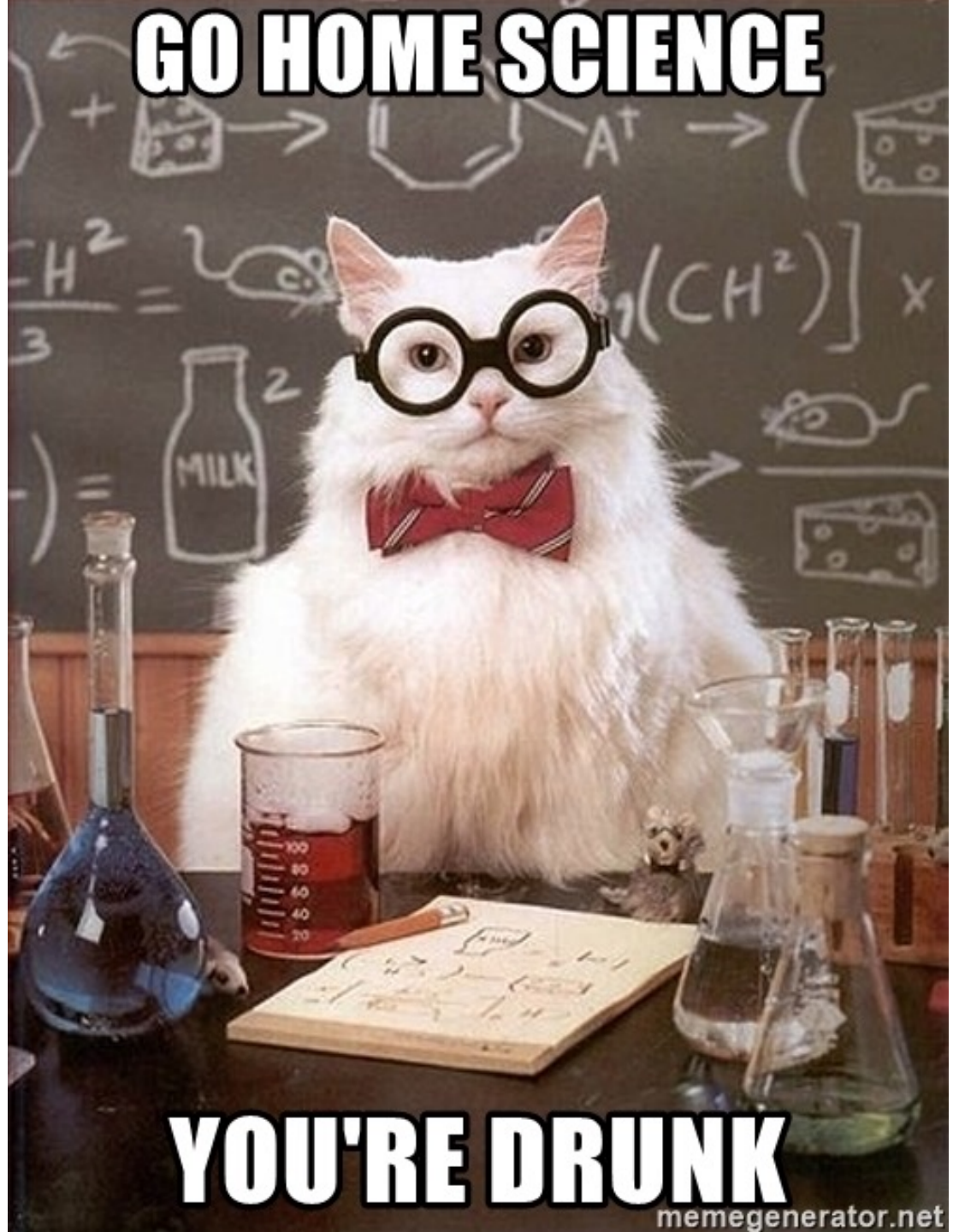


SCORE – reprodukovateľnosť (predbežné výsledky)



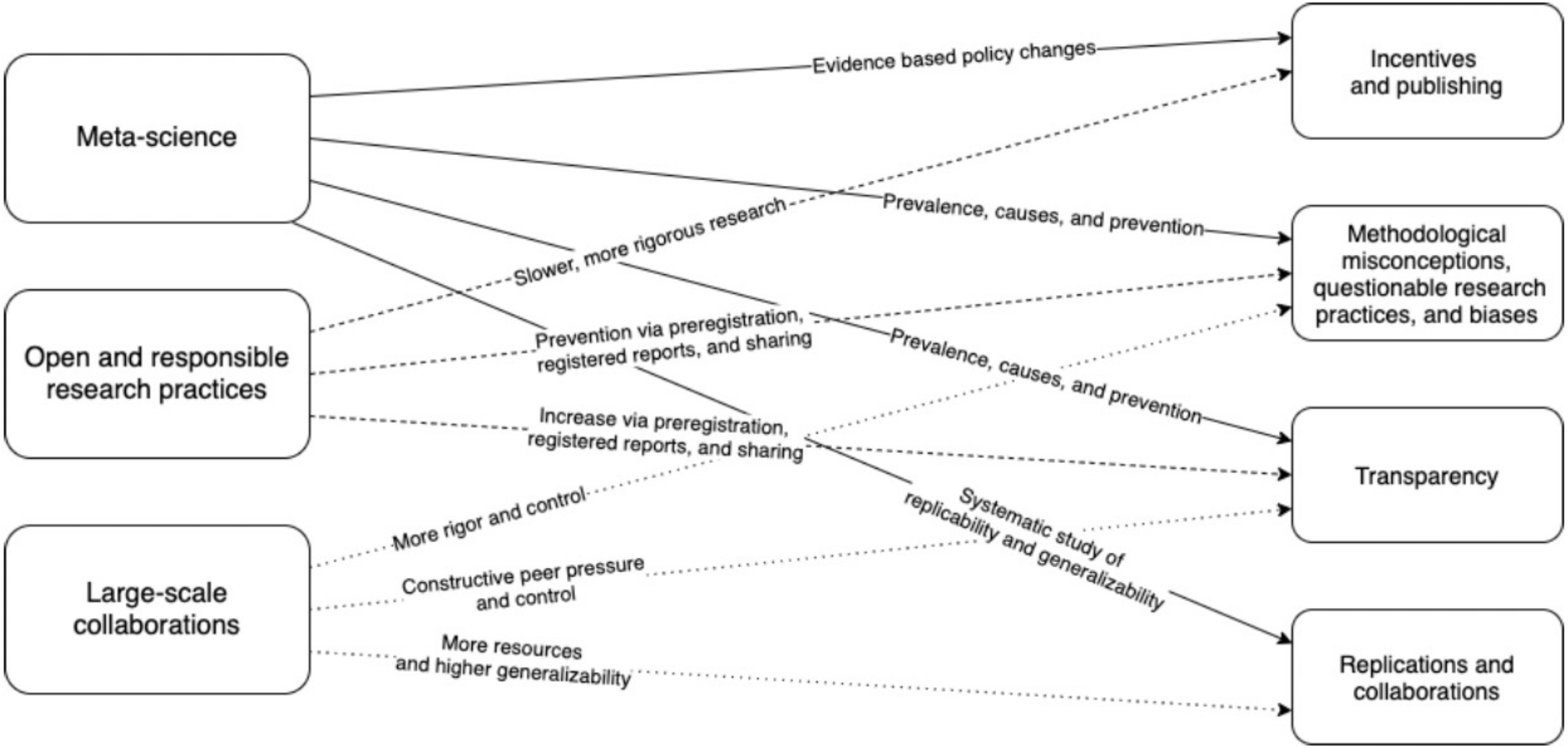
SCORE – replikovateľnosť (preliminary results)




GO HOME SCIENCE
MILK
YOU'RE DRUNK
memegenerator.net

Riešenia?

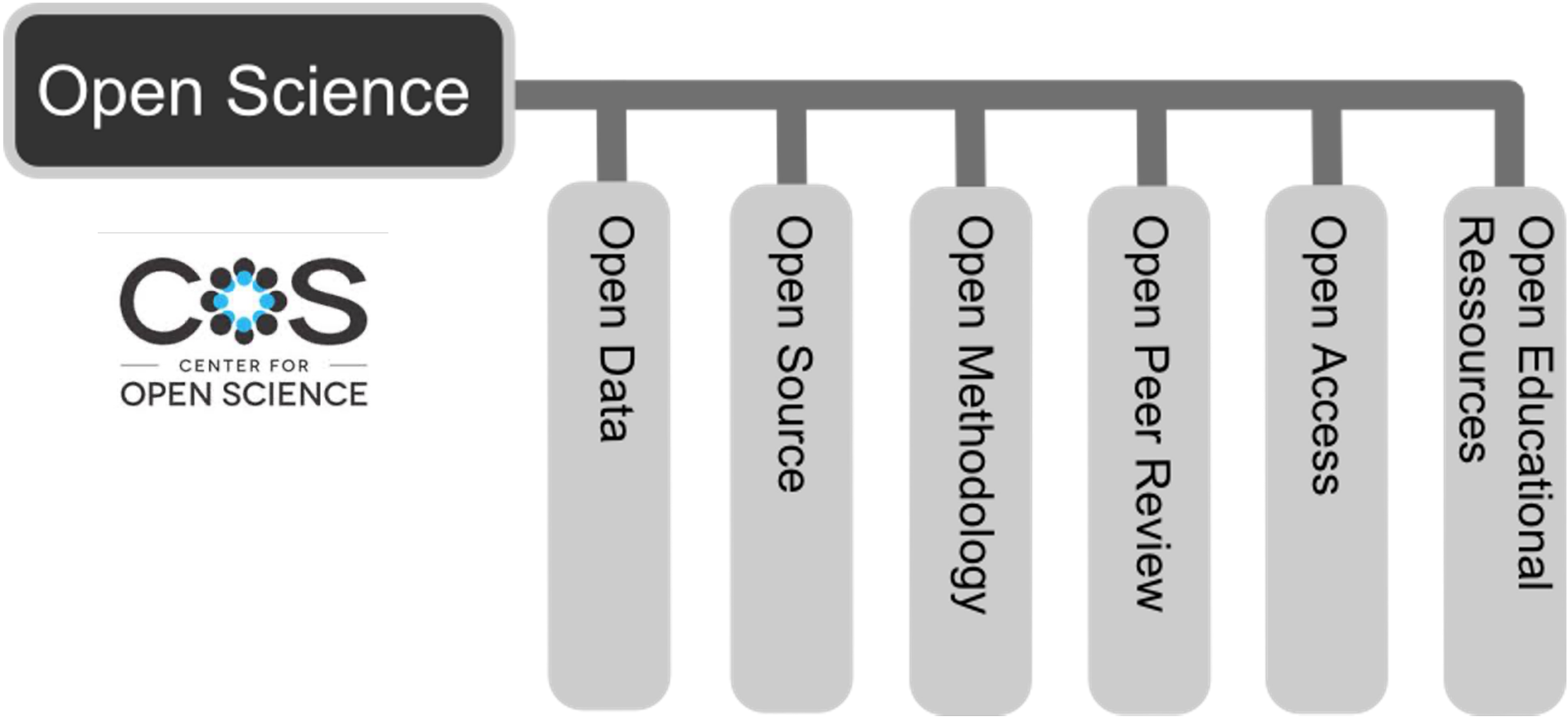
Piliere kredibilného sociálnovedného výskumu a ich efekt na problémové oblasti



Zodpovedné výskumné praktiky

- Transparentnosť, úprimnosť, zodpovednosť, otvorenosť voči kritike
- Vzdelávanie a tréning (špeciálne v metodológii)
- Udržiavanie zdravého balansu (zodpovedné praktiky sú *drahšie*)

- Zdieľanie dát, analytického kódu a materiálov
- Predregistrácia a registered reports (PCI Registered Reports)
- Preprinty a publikovanie open access
- Otvorené peer review a jeho história
- Replikačné a generalizačné štúdie



CENTER FOR
OPEN SCIENCE

Registered Reports

Peer Community In

Registered Reports

Free and transparent pre- and post-study recommendations across research fields

Zodpovedné výskumné praktiky

- Mali by byť samozrejmosťou
- Šetria zdroje
- Otvorená veda a zodpovedné výskumné praktiky nevyriešia všetko
- Kolektívna apatia

Kvalitatívny výskum?

Kvalitatívny výskum

- Postupy pre predregistrácie a registered reports už existujú
- Manažment dát je oveľa náročnejší a drahší
- Neskúmajú sa efekty – z podstaty teda nie je čo replikovať

Duplikácia (Duplication)

- Replikácia
  - Results-driven concept

- V kvalitatívnom výskume nie sú efekty
- Môže byť užitočné zbierať dáta v novom kontexte
  - Lepšie porozumenie skúmaného fenoménu
- Duplikácia
  - Methods-driven concept
  - Zopakovanie procedúry, ktorá viedla ku generácii dát (t.j. zopakovanie metódy)
  - Očakávania ohľadom výsledkov nie sú potrebné

Potenciálne spolupráce

- Ontologická rekonštrukcia hrania digitálnych hier
  - Kvalitatívny výskum
  - Hľadáme osoby, ktoré vyhľadali pomoc kvôli hraniu
- Longitudinálne dáta ohľadom duševného zdravia (2020 – 2023)
  - Komplexné dáta na reprezentatívnej vzorke
- Stav SVK literatúry v sociálnych vedách
  - Vrátane záverečných prác
  - Bibliometrická a hĺbková analýza
- Výskum kredibility poznatkov z psychológie a pedagogiky
  - Napr. reprodukovateľnosť experience sampling method štúdií


RN
SK
Slovak Reproducibility
Network

Ďakujem!

- Otázky/komentáre?

- matho.adamkovic@gmail.com | matus.m.adamkovic@jyu.fi