
REPLICATIONS
TATO PŘEDNÁŠKA
PRACTICAL RELEVANCE

placeholder

První průběžná příprava:

	2023	2024
podcast	22 (47 %)	15 (24 %)
opakování	19 (40 %)	43 (68 %)
jiné	4 (9 %)	0 (0 %)
neodpovědělo	2 (4 %)	5 (8 %)

Fisherův exaktní test $p(110) = 0{,}002$

MUNI FSS Katedra psychologie

Zobecnitelnost psychologického výzkumu

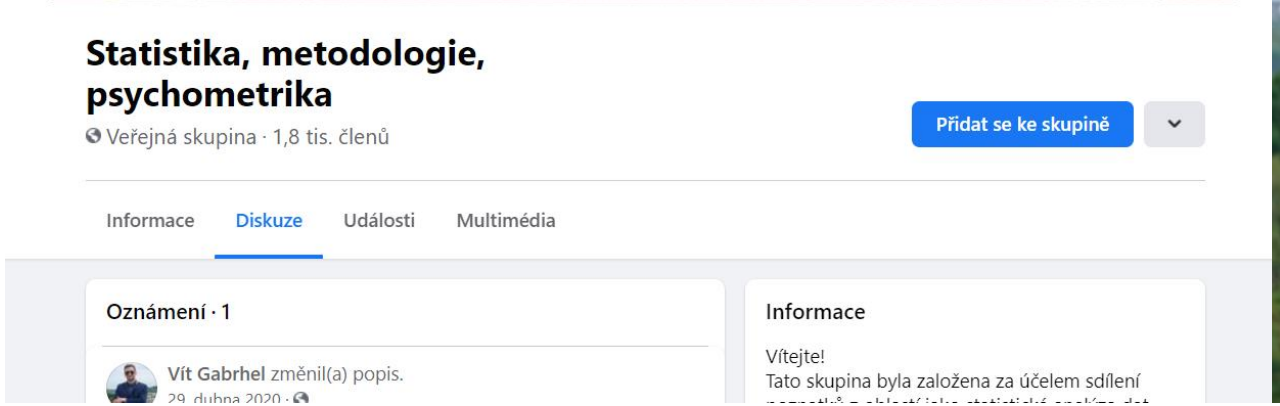
Přednáška 2 | 1. 10. 2024

PSYn4790 | **Psychometrika: Měření v psychologii**

Katedra psychologie, Fakulta sociálních studií MU

Hynek Cígler | cigler@fss.muni.cz

Acknowledgement: Děkuji Vítu Gabrhelovi, že dnešní předná[illegible] ROZDĚLĚNO NA DVĚ SETKÁNÍ!



https://www.facebook.com/groups/461796387316423

Testování statistických hypotéz

Žáky (N = 60) jsme rozdělili náhodně do dvou skupin.

- Skupina A (n = 30) byla vyučována tradičně, skupina B (n = 30) experimentálně.

Po ukončení experimentu byly znalosti žáků ve skupině B (M = 0,7; SD = 1) vyšší než ve skupině A (M = 0; SD = 1).

Rozdíl byl statisticky významný, $\boldsymbol{t(58) = 2{,}71,\ p = 0{,}009}$.

- 95% interval spolehlivosti pro rozdíl je $\boldsymbol{_{95\%}CI = [0{,}183—1{,}217]}$.

Jaká je korektní interpretace provedeného statistického testu?

Jaké jsou předpoklady tohoto závěru?

Do jaké míry lze výsledek zobecnit?

Testování statistických hypotéz

Žáky (N = 60) jsme rozdělili náhodně do dvou skupin.

- Skupina A (n = 30) byla vyučována tradičně, skupina B (n = 30) experimentálně.

Po ukončení experimentu byly znalosti žáků ve skupině B (M = 0,7; SD = 1) vyšší než ve skupině A (M = 0; SD = 1).

Velikost vzorku byla a-priori odhadnuta pomocí power analýzy:

- Pro α = 0,05, **$1-\beta$ = 0,8** a očekávanou velikost efektu d = 0,75.
- Pozorovaná síla testu je **$1-\beta$ = 0,76.**

Jaká je pozorovaná velikost účinku?

Co to je ta síla testu?

V čem se liší pozorovaná a a-priori odhadnutá síla testu?

Null hypothesis significance testing
Testování statistických hypotéz: NHST

P-hodnota: $p = P(D|H_0)$

- Pravděpodobnost pozorování stejných nebo extrémnějších dat D (resp. testové statistiky), pokud by platila nulová hypotéza H_0.
- Pravděpodobnost chybného zamítnutí nulové hypotézy.

Hladina spolehlivosti: α

- Slouží ke kategorickým soudům o p-hodnotě: $(p < \alpha) \rightarrow H_1$, $(p \geq \alpha) \rightarrow H_0$.
- Pravděpodobnost chyby I. typu (falešně pozitivní závěr): $\alpha = P(\mathrm{H}_1|\mathrm{H}_0)$

Síla testu: $1 - \beta = P(p < \alpha|N, r, \dots)$

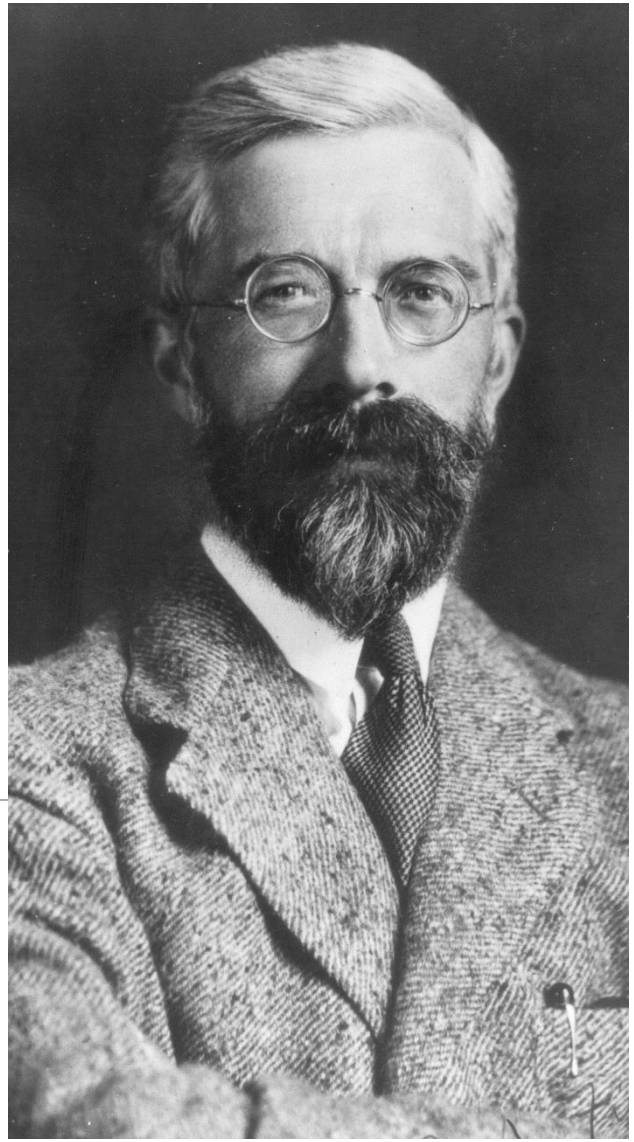
- Pravděpodobnost (správného) zamítnutí nulové hypotézy.
- β je pravděpodobnost chyby II. typu (falešně negativní závěr).

Velikost účinku: *d*, *r*, *B*, β, R^2, η^2...

Doporučuji: https://rpsychologist.com/d3/nhst/

„No isolated experiment, however significant in itself, can suffice for the experimental demonstration of any natural phenomenon.“

FISHER (1971, s. 13)



Fisher, R. (1971). *The Design of Experiments*. Hafner Publishing Company. 1st ed. 1935.

Vývoj vědeckého poznání

Tradiční model poznání: **kumulativní vývoj**.

- Kumulace pozorování a deduktivních úsudků.

Thomas S. Kuhn (1962): **paradigmatické pojetí**.

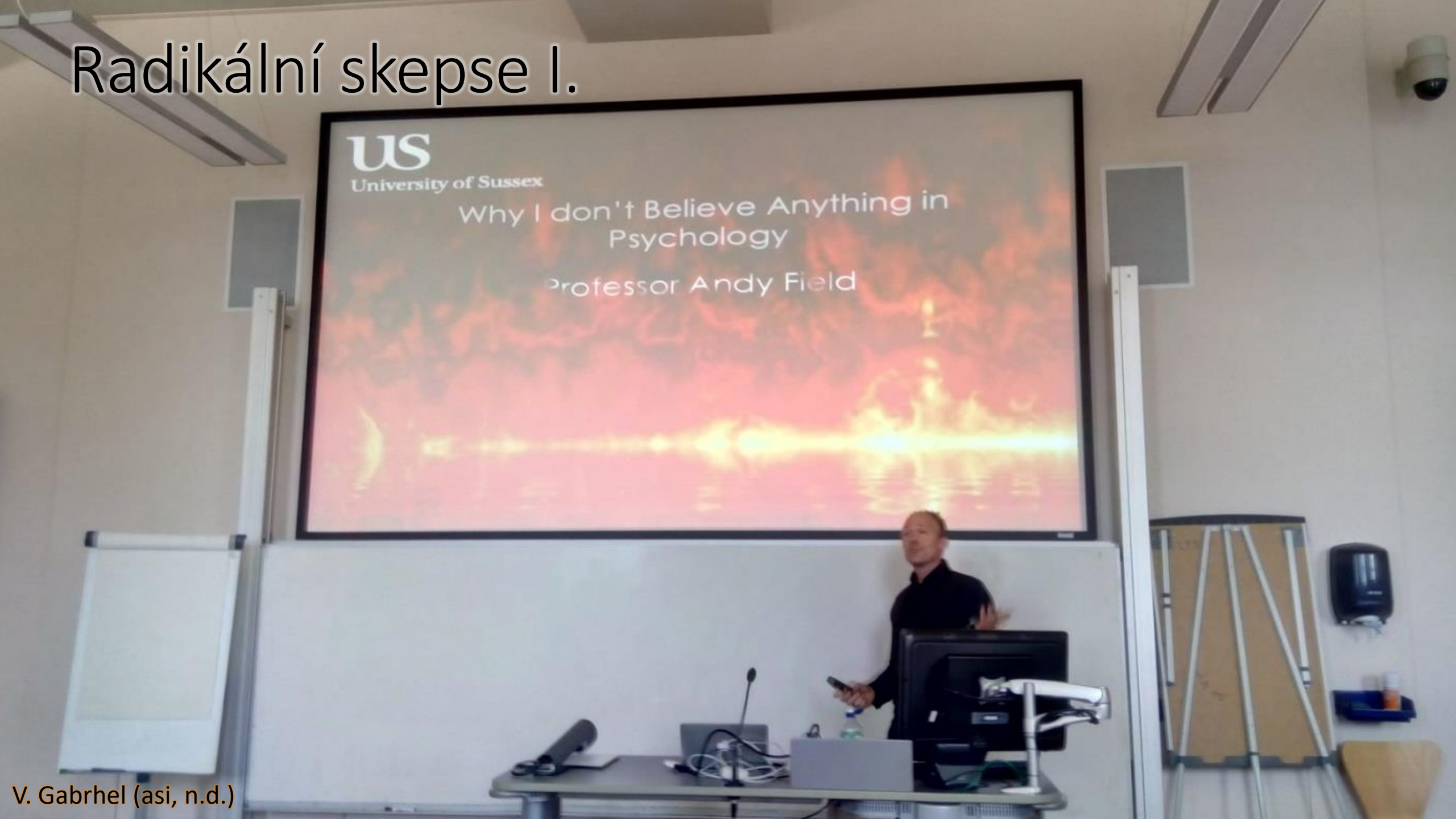
- Vědecké poznání prochází evolucí prokládanou „vědeckými revolucemi“.
- Paradigma – koherentní výkladový rámec přijímaný (drtivou) většinou odborníků.

V každém případě ale předpokládáme, že na předchozích poznatcích (datech) lze stavět.

- Interpretace mohou být mylné, vysvětlení chybná, ale pozorování zůstávají.

Opravdu můžeme dřívějším zjištěním věřit?

Radikální skepse I.



V. Gabrhel (asi, n.d.)

Začátek „krize“: 2011–2012

Daryl Bem (2011)
Feeling the Future

Diederik Stapel
(58 retrakcí 2011–2019)

John Bargh (2010)
priming stářím (5.000 citací)

Radikální skepse II:
Estimating the reproducibility of psychological science

„We conducted a large-scale, collaborative effort to obtain an initial estimate of the reproducibility of psychological science.“

100 studií a výsledky jejich replikace

- Psychological Science
- Journal of Personality and Social Psychology
- Journal of Experimental Psychology: Learning, Memory, and Cognition

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Alexander A. Aarts, Joanna E. Anderson, Christopher J. Anderson, Peter R. Attridge, Angela Attwood, Jordan Axt, Molly Babel, **Štěpán Bahník**, Erica Baranski, Michael Barnett-Cowan,Elizabeth Bartmess, Jennifer Beer, Raoul Bell, Heather Bentley, Leah Beyan, Grace Binion, Denny Borsboom, Annick Bosch, Frank A. Bosco, Sara D. Bowman, Mark J. Brandt, Erin Braswell, Hilmar Brohmer, Benjamin T. Brown, Kristina Brown, Jovita Brüning, Ann Calhoun-Sauls, Shannon P. Callahan, Elizabeth Chagnon, Jesse Chandler, Christopher R. Chartier, Felix Cheung, Cody D. Christopherson, Linda Cillessen, Russ Clay, Hayley Cleary, Mark D. Cloud, Michael Cohn, Johanna Cohoon,Simon Columbus, Andreas Cordes, Giulio Costantini, Leslie D. Cramblet Alvarez, Ed Cremata, Jan Crusius, Jamie DeCoster, Michelle A. DeGaetano, Nicolás Della Penna, Bobby den Bezemer, Marie K. Deserno, Olivia Devitt, Laura Dewitte, David G. Dobolyi, Geneva T. Dodson, M. Brent Donnellan, Ryan Donohue, Rebecca A. Dore, Angela Dorrough, Anna Dreber, Michelle Dugas, Elizabeth W. Dunn, Kayleigh Easey, Sylvia Eboigbe, Casey Eggleston, Jo Embley, Sacha Epskamp, Timothy M. Errington, Vivien Estel, Frank J. Farach, Jenelle Feather, Anna Fedor, Belén Fernández-Castilla, Susann Fiedler, James G. Field, Stanka A. Fitneva, Taru Flagan, Amanda L. Forest, Eskil Forsell, Joshua D. Foster, Michael C. Frank, Rebecca S. Frazier, Heather Fuchs, Philip Gable, Jeff Galak,Elisa Maria Galliani, Anup Gampa, Sara Garcia, Douglas Gazarian, Elizabeth Gilbert, Roger Giner-Sorolla, Andreas Glöckner, Lars Goellner, Jin X. Goh, Rebecca Goldberg, Patrick T. Goodbourn, Shauna Gordon-McKeon, Bryan Gorges, Jessie Gorges, Justin Goss, Jesse Graham, James A. Grange, Jeremy Gray, Chris Hartgerink, Joshua Hartshorne, Fred Hasselman, Timothy Hayes, Emma Heikensten, Felix Henninger, John Hodsoll,Taylor Holubar, Gea Hoogendoorn, Denise J. Humphries, Cathy O.-Y. Hung, Nathali Immelman, Vanessa C. Irsik, Georg Jahn, Frank Jäkel, Marc Jekel, Magnus Johannesson, Larissa G. Johnson, David J. Johnson, Kate M. Johnson, William J. Johnston, Kai Jonas, Jennifer A. Joy-Gaba, Heather Barry Kappes, Kim Kelso, Mallory C. Kidwell, Seung Kyung Kim, Matthew Kirkhart, Bennett Kleinberg, Goran Knežević,Franziska Maria Kolorz, Jolanda J. Kossakowski, Robert Wilhelm Krause, Job Krijnen, Tim Kuhlmann, Yoram K. Kunkels, Megan M. Kyc, Calvin K. Lai, Aamir Laique, Daniël Lakens,Kristin A. Lane, Bethany Lassetter, Ljiljana B. Lazarević, Etienne P. LeBel, Key Jung Lee,Minha Lee, Kristi Lemm, Carmel A. Levitan, Melissa Lewis, Lin Lin, Stephanie Lin,Matthias Lippold, Darren Loureiro, Ilse Luteijn, Sean Mackinnon, Heather N. Mainard,Denise C. Marigold, Daniel P. Martin, Tylar Martinez, E.J. Masicampo, Josh Matacotta,Maya Mathur, Michael May, Nicole Mechin, Pranjal Mehta, Johannes Meixner, Alissa Melinger, Jeremy K. Miller, Mallorie Miller, Katherine Moore, Marcus Möschl, Matt Motyl, Stephanie M. Müller, Marcus Munafo, Koen I. Neijenhuijs, Taylor Nervi, Gandalf Nicolas, Gustav Nilsonne, Brian A. Nosek, Michèle B. Nuijten, Catherine Olsson,Colleen Osborne, Lutz Ostkamp, Misha Pavel, Ian S. Penton-Voak, Olivia Perna, Cyril Pernet, Marco Perugini, R. Nathan Pipitone, Michael Pitts, Franziska Plessow, Jason M. Prenoveau, Rima-Maria Rahal, Kate A. Ratliff, David Reinhard, Frank Renkewitz,Ashley A. Ricker, Anastasia Rigney, Andrew M. Rivers, Mark Roebke, Abraham M. Rutchick, Robert S. Ryan, Onur Sahin, Anondah Saide, Gillian M. Sandstrom, David Santos, Rebecca Saxe, René Schlegelmilch, Kathleen Schmidt, Sabine Scholz,Larissa Seibel, Dylan Faulkner Selterman, Samuel Shaki, William B. Simpson, H. Colleen Sinclair, Jeanine L. M. Skorinko, Agnieszka Slowik, Joel S. Snyder, Courtney Soderberg,Carina Sonnleitner, Nick Spencer, Jeffrey R. Spies, Sara Steegen, Stefan Stieger, Nina Strohminger, Gavin B. Sullivan, Thomas Talhelm, Megan Tapia, Anniek te Dorsthorst,Manuela Thomae, Sarah L. Thomas, Pia Tio, Frits Traets, Steve Tsang, Francis Tuerlinckx, Paul Turchan, Milan Valášek, Anna E. van 't Veer, Robbie Van Aert, Marcel van Assen, Riet van Bork, Mathijs van de Ven, Don van den Bergh, Marije van der Hulst,Roel van Dooren, Johnny van Doorn, Daan R. van Renswoude, Hedderik van Rijn, Wolf Vanpaemel, Alejandro Vásquez Echeverría, Melissa Vazquez, Natalia Velez, Marieke Vermue, Mark Verschoor, Michelangelo Vianello, Martin Voracek, Gina Vuu, Eric-Jan Wagenmakers, Joanneke Weerdmeester, Ashlee Welsh, Erin C. Westgate, Joeri Wissink,Michael Wood, Andy Woods, Emily Wright, Sining Wu, Marcel Zeelenberg, Kellylynn Zuni

Radikální skepse II:
Estimating the reproducibility of psychological science

Původní velikost efektů:

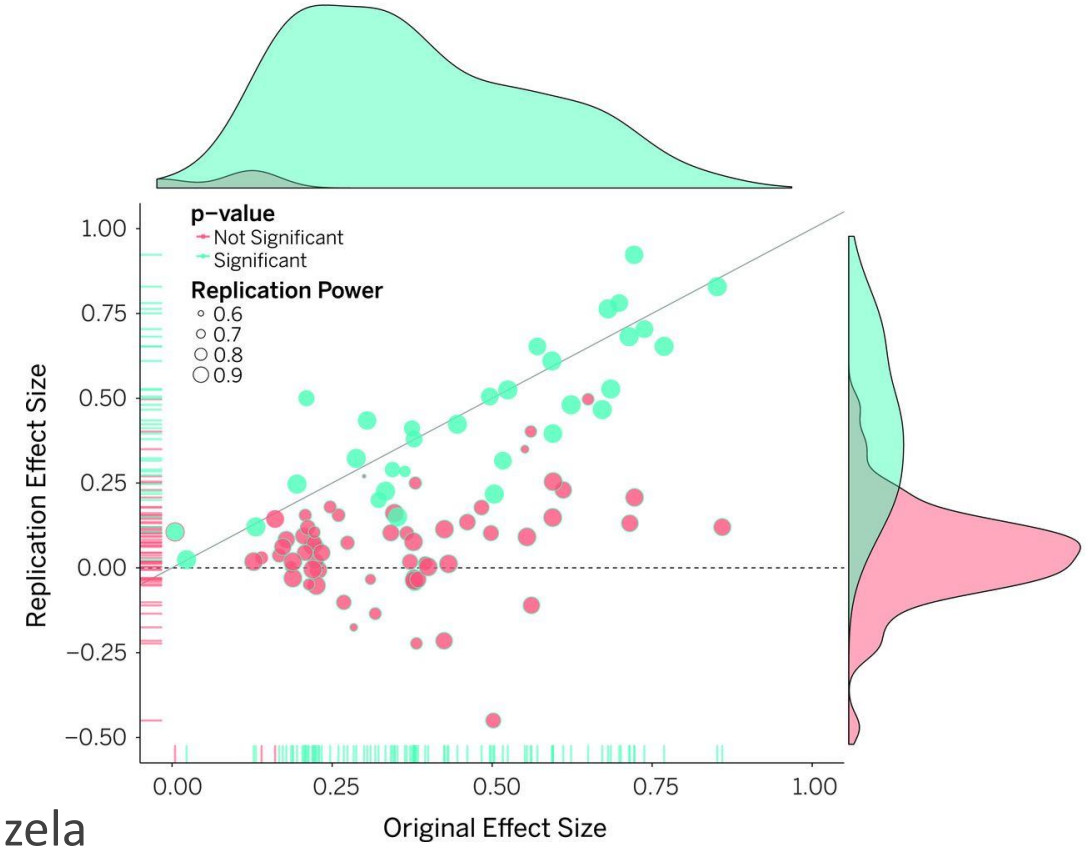
- Průměrná velikost účinku
  M_r = 0,403; *SD* = 0,188
- Statistická signifikance: 97 % studií $p < 0,05$

Design replikací: průměrná síla testu 1−β = 0,92.

- → 89 % replikací by mělo být signifikantní.
- Ale: průměrná síla testu originálních studií: 39 %.

Replikovaná velikost efektů:

- Průměrná velikost účinku
  M_r = 0,197; *SD* = 0,257
- Statistická signifikance: 36 % studií $p < 0,05$



Hodnota velikostí účinku z původních studií se nacházela v 95% intervalu spolehlivosti při replikaci v 47 % případů.

Příklady nereplikovatelny

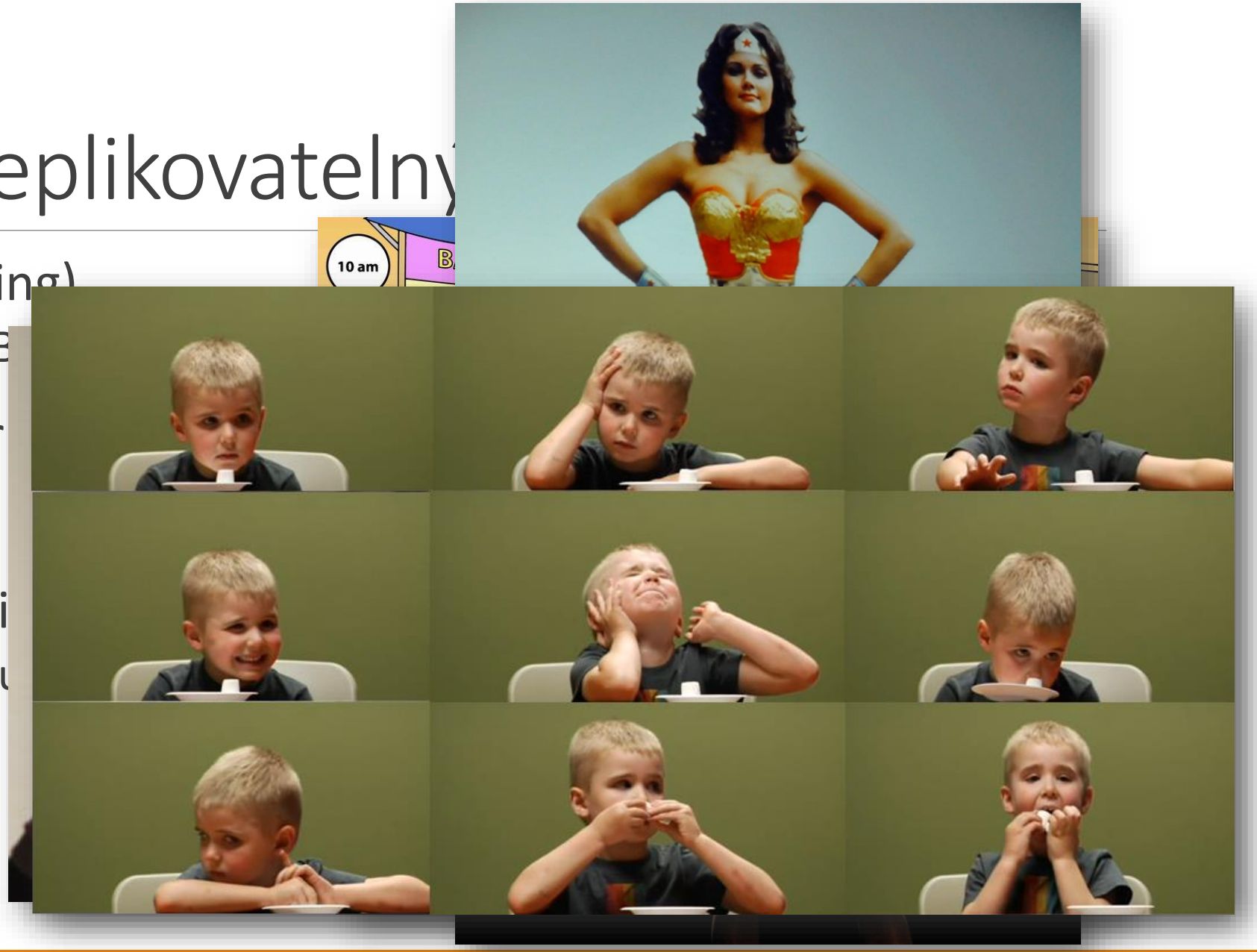
Priming (social priming)
- elderly priming, MacB

Ego deplation (vyčer

Power posing

Vybrané aspekty faci
- „smiling will make you

Marshmallow test



Příklady nereplikovatelných efektů

Priming (social priming).

- elderly priming, MacBeth effect, cleanliness priming, money priming...

Ego deplation (vyčerpání ega).

Power posing

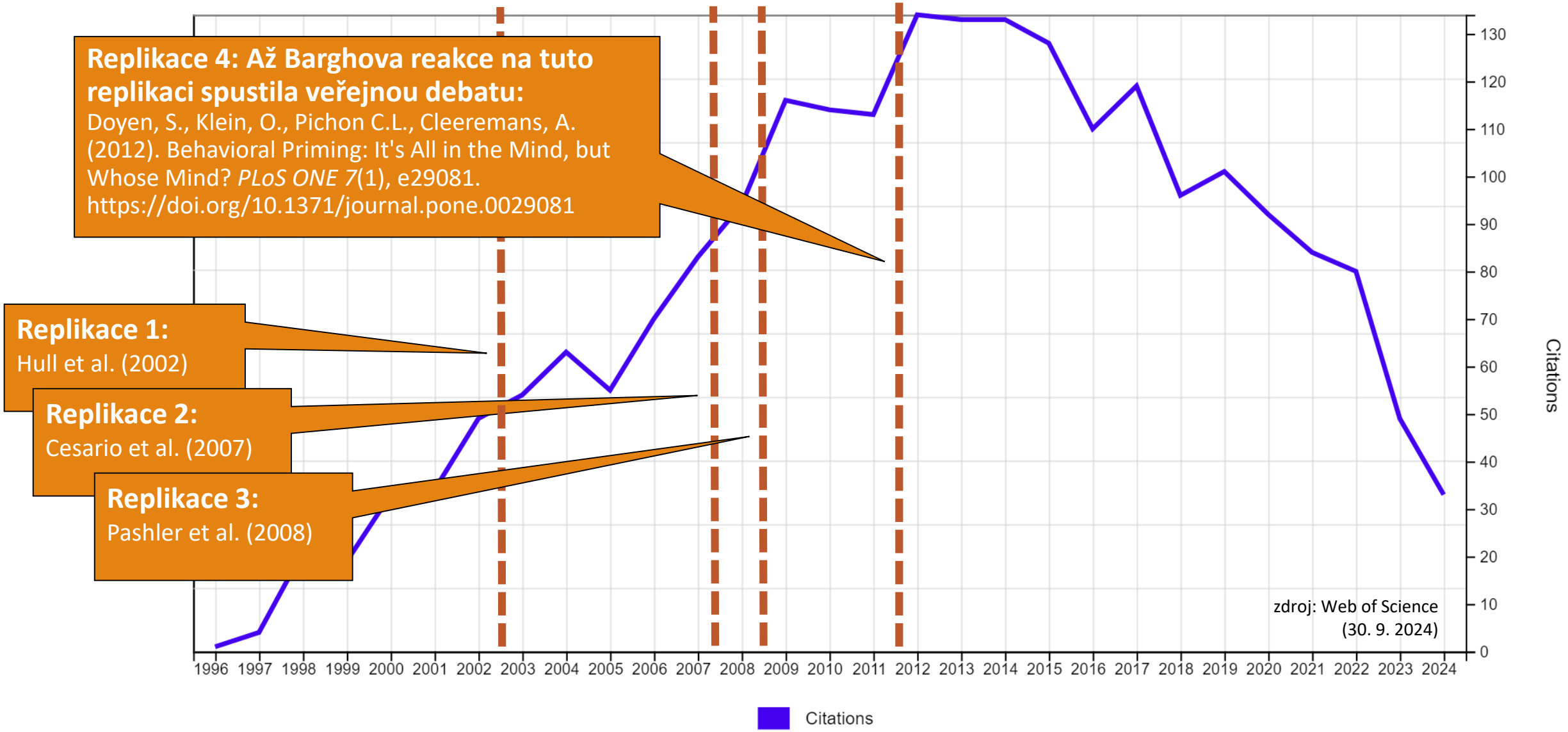
Vybrané aspekty facial-feedback hypothesis

- „smiling will make you feel happier“

Marshmallow test

Počet citací primingu stářím v čase

Bargh, J.A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71(2)*. https://doi.org/10.1037/0022-3514.71.2.230



Dishonest Report: vol 1 (2012)

Prominentní článek prominentních vědců Dana Arielyho a Francescy Gino.

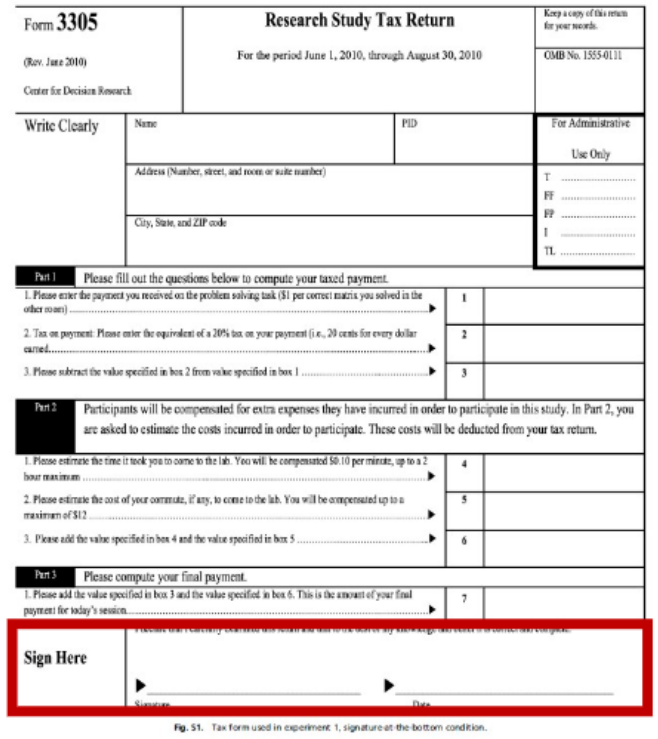
Shu, L.L., Mazar, N., Gino, F., Ariely, D., & Bazerman, Max. H. (2012). **Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end.** *PNAS, 109*(38), 15197-15200. https://doi.org/10.1073/pnas.1209746109

Článek obsahoval 3 studie:

- 1 a 2: Laboratorní experimenty.
- 3: Terénní experiment.

SIGN AT THE BOTTOM

Shu et al. 10.1073/pnas.1209746109



Form **3305** (Rev. June 2010) Center for Decision Research

Research Study Tax Return

For the period June 1, 2010, through August 30, 2010

Keep a copy of this return for your records. OMB No. 1555-0111

Write Clearly — Name | PID | For Administrative Use Only

Address (Number, street, and room or suite number)

City, State, and ZIP code

T FF FP I TL

Part 1 Please fill out the questions below to compute your taxed payment.

1. Please enter the payment you received on the problem solving task ($1 per correct matrix you solved in the other room) ▶ 1
2. Tax on payment: Please enter the equivalent of a 20% tax on your payment (i.e., 20 cents for every dollar earned) ▶ 2
3. Please subtract the value specified in box 2 from value specified in box 1 ▶ 3

Part 2 Participants will be compensated for extra expenses they have incurred in order to participate in this study. In Part 2, you are asked to estimate the costs incurred in order to participate. These costs will be deducted from your tax return.

1. Please estimate the time it took you to come to the lab. You will be compensated $0.10 per minute, up to a 2 hour maximum ▶ 4
2. Please estimate the cost of your commute, if any, to come to the lab. You will be compensated up to a maximum of $12 ▶ 5
3. Please add the value specified in box 4 and the value specified in box 5 ▶ 6

Part 3 Please compute your final payment.

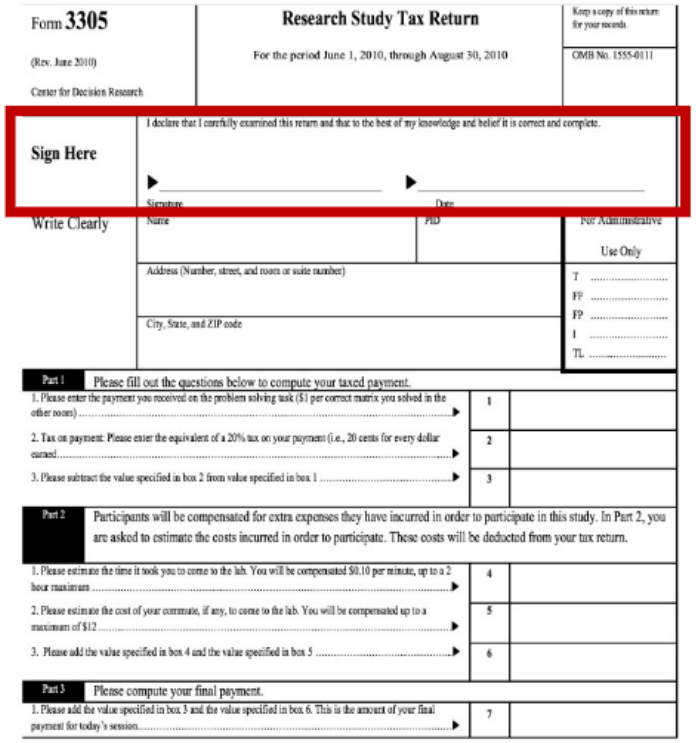
1. Please add the value specified in box 3 and the value specified in box 6. This is the amount of your final payment for today's session ▶ 7

Sign Here — I declare that I carefully examined this return and that to the best of my knowledge and belief it is correct and complete.

▶ Signature ▶ Date

Fig. S1. Tax form used in experiment 1, signature-at-the-bottom condition.

SIGN AT THE TOP



Form **3305** (Rev. June 2010) Center for Decision Research

Research Study Tax Return

For the period June 1, 2010, through August 30, 2010

Keep a copy of this return for your records. OMB No. 1555-0111

Sign Here — I declare that I carefully examined this return and that to the best of my knowledge and belief it is correct and complete.

▶ Signature ▶ Date

Write Clearly — Name | PID | For Administrative Use Only

Address (Number, street, and room or suite number)

City, State, and ZIP code

T FF FP I TL

Part 1 Please fill out the questions below to compute your taxed payment.

1. Please enter the payment you received on the problem solving task ($1 per correct matrix you solved in the other room) ▶ 1
2. Tax on payment: Please enter the equivalent of a 20% tax on your payment (i.e., 20 cents for every dollar earned) ▶ 2
3. Please subtract the value specified in box 2 from value specified in box 1 ▶ 3

Part 2 Participants will be compensated for extra expenses they have incurred in order to participate in this study. In Part 2, you are asked to estimate the costs incurred in order to participate. These costs will be deducted from your tax return.

1. Please estimate the time it took you to come to the lab. You will be compensated $0.10 per minute, up to a 2 hour maximum ▶ 4
2. Please estimate the cost of your commute, if any, to come to the lab. You will be compensated up to a maximum of $12 ▶ 5
3. Please add the value specified in box 4 and the value specified in box 5 ▶ 6

Part 3 Please compute your final payment.

1. Please add the value specified in box 3 and the value specified in box 6. This is the amount of your final payment for today's session ▶ 7

Dishonest Report: vol 2 (2021)

Simonsohn, U., Nelson, L., & Simmons, J. (August 17, 2021). Evidence of Fraud in an Influential Field Experiment About Dishonesty. *Data Colada* (98). https://datacolada.org/98

This is Table 1 in Kristal et al. (2020), reporting their re-analysis of Shu et al. (2012)

	Sign-at-the-bottom, means (SD)	Sign-at-the-top, means (SD)	Two-sided *t* test, values
Baseline odometer reading (*t*0)	75,034.50 (50,265.35)	59,692.71 (49,953.51)	$t_{(13,474)} = 17.78$, $P < 0.0001$
New odometer reading (*t*1)	98,705.14 (51,934.76)	85,791.10 (51,701.31)	$t_{(13,475)} = 14.47$, $P < 0.0001$
Difference in odometer readings; i.e., miles driven (*t*1–*t*0)*	23,670.64 (12,621.38)	26,098.40 (12,253.37)	$t_{(13,448)} = -11.331$, $P < 0.0001$

*This row was the outcome reported in the original paper.

Figure 1. Histogram of Miles Driven - Car #1 (N=13,488)

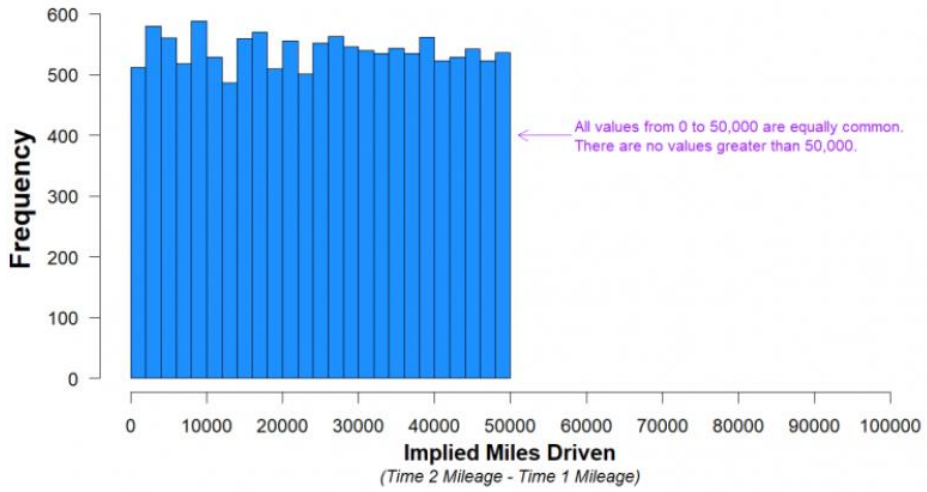
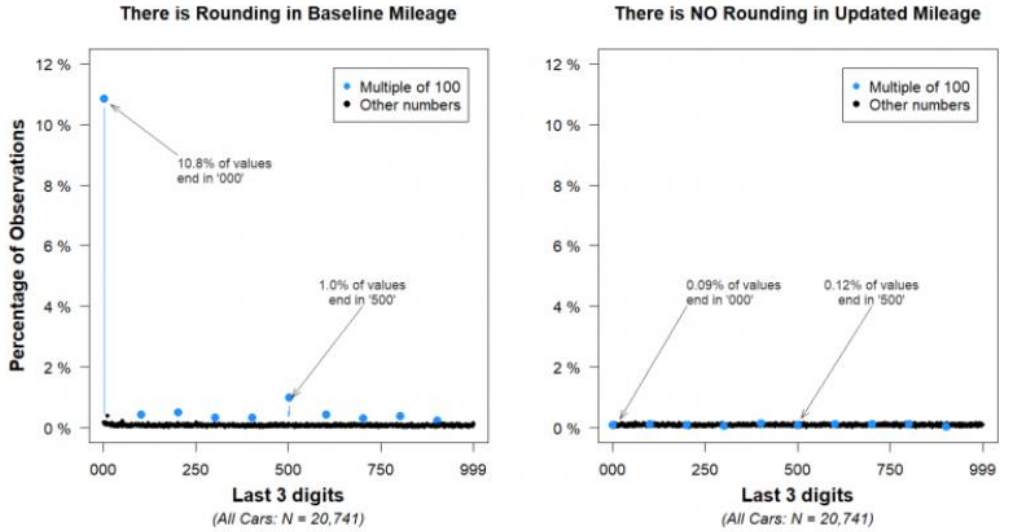


Figure 3. Last Three Digits at Baseline (Time 1) vs Updated (Time 2)



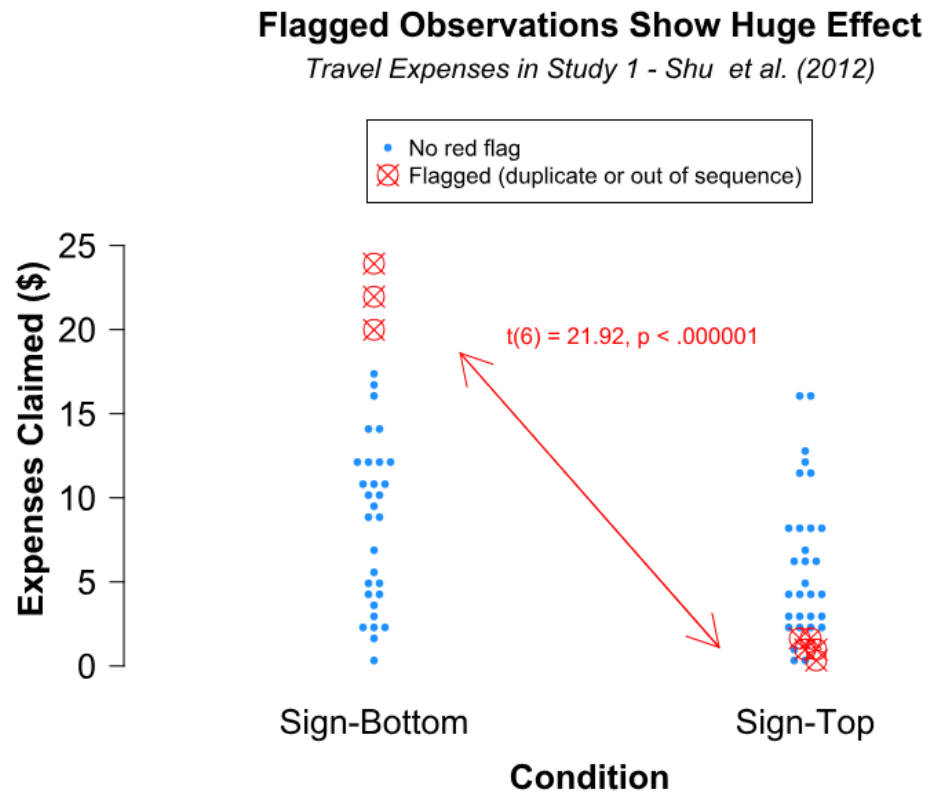
Dishonest Report: vol 3 (2023)

Simonsohn, U., Nelson, L., & Simmons, J. (June 17, 2023). Data Falsificada (Part 1): "Clusterfake". *Data Colada* (109). https://datacolada.org/98

- Harvard byl upozorněn už v roce 2021.

„It turns out that Study 1's data were also tampered with…but by a different person."

Francesca Gino navíc zfalšovala data v celé řadě dalších studií.



Dishonest Report: vol 4 (2023)

Články prof. Gino byly staženy z mnoha časopisů.

Gino na Harvard Business School už nepůsobí.

V srpnu 2023 nicméně prof. Gino zažalovala Simonsohna, Nelsona a Simmonse o 25 mil. USD.

- (Sbírka už byla ukončena, všichni mají dobré právníky.)

Během procesu a v žalobních důkazech byly nicméně publikovány výsledky interního vyšetřování Harvardu.

Data byla jednoznačně zmanipulovaná.

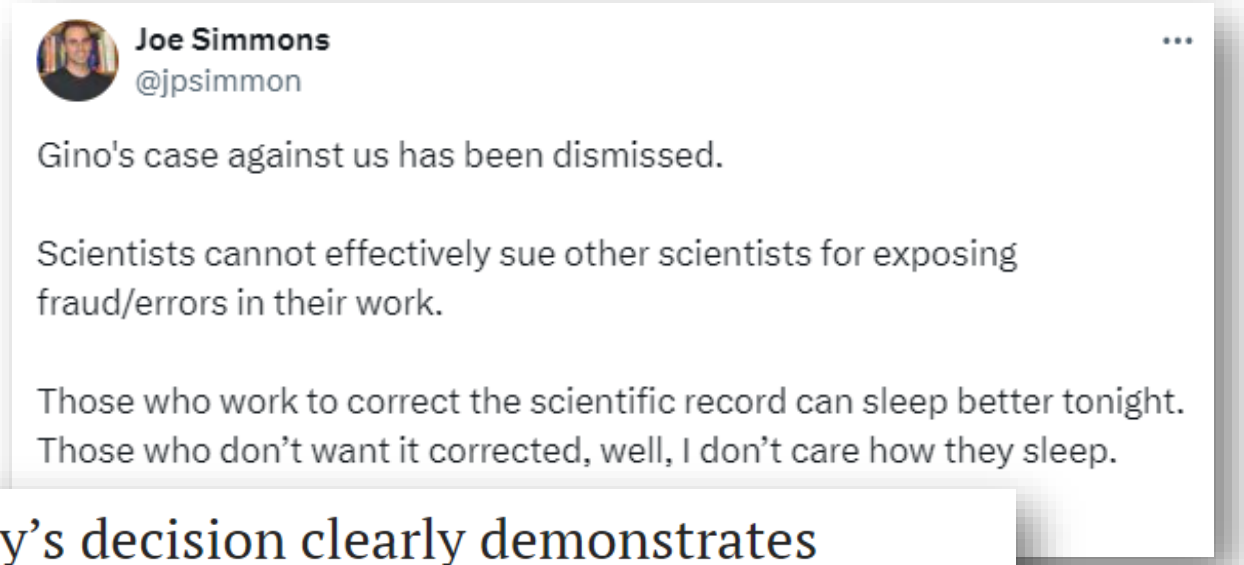
Dishonest Report: vol 5 (2024; rn)

Díky soudnímu sporu byl report Harvardu publikován: 1288 stran.

- Díky tomu bylo možné zrekonstruovat přesný popis falšování dat.

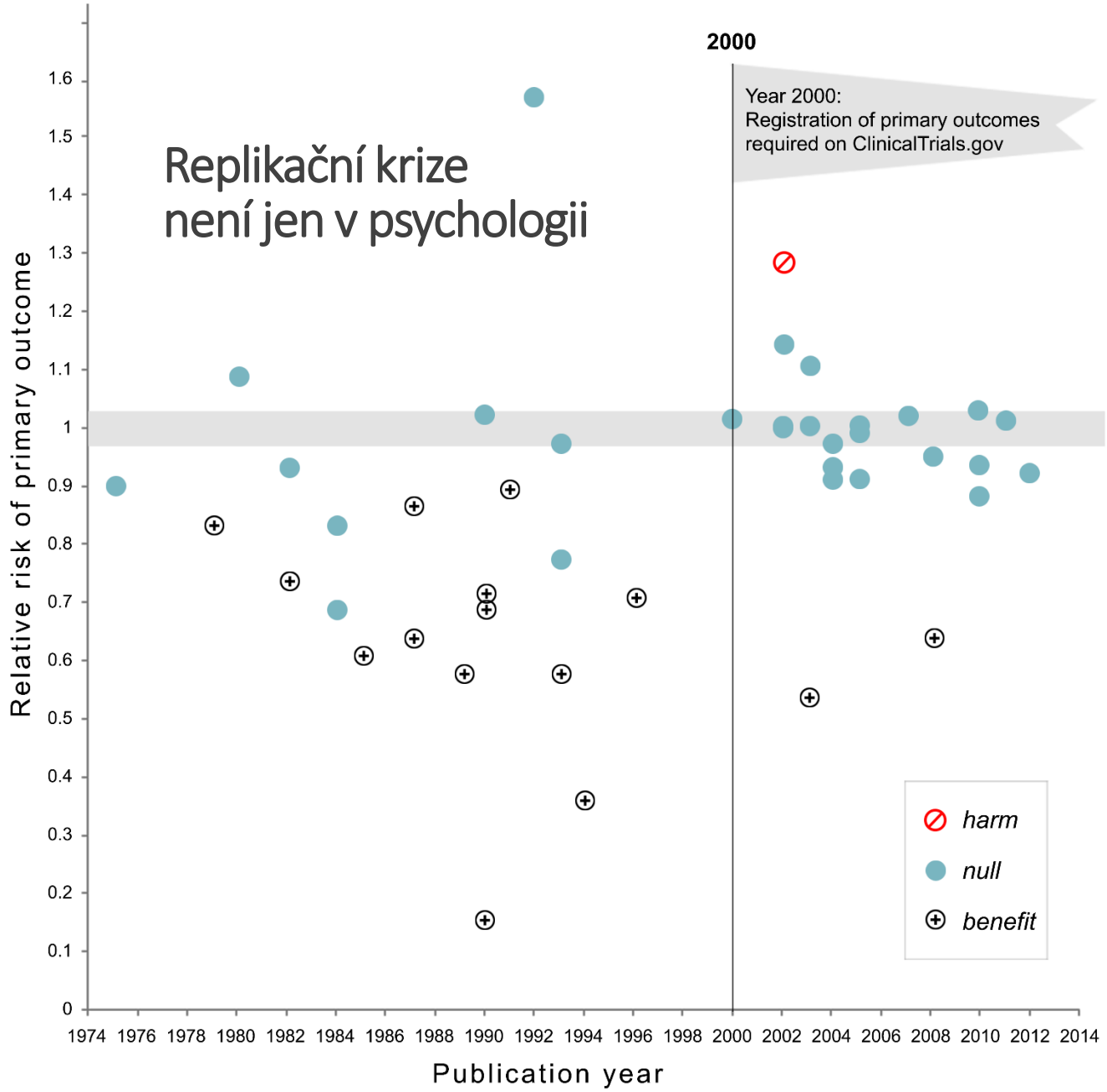
A nakonec happy end: soud žalobu zamítl (11. 9. 2024).

- V crowdfundingové kampani se nakonec vybralo 378.196 USD (k 30. 9. 2024).



"Today's decision clearly demonstrates Harvard treated Professor Gino differently from other misconduct investigations and their own stated policies," said Gino's attorney Andrew Miltenberg, in a statement to *Science*.

Replikační krize není jen v psychologii... a v zahraničí.



*„We identified all large NHLBI supported RCTs between 1970 and 2012 evaluating **drugs or dietary supplements for the treatment or prevention of cardiovascular disease**. Trials were **included if direct costs >$500,000/year**, participants were adult humans, and the **primary outcome was cardiovascular risk, disease or death**.* [...] *The number NHLBI trials reporting positive results declined after the year 2000. Prospective declaration of outcomes in RCTs, and the adoption of transparent reporting standards, as required by clinicaltrials.gov, may have contributed to the trend toward null findings.“*

Replikační krize nejen v psychologii.

- Kaplan, R.M., Irvin, V.L. (2015). Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time. PLoS ONE 10(8): e0132382.

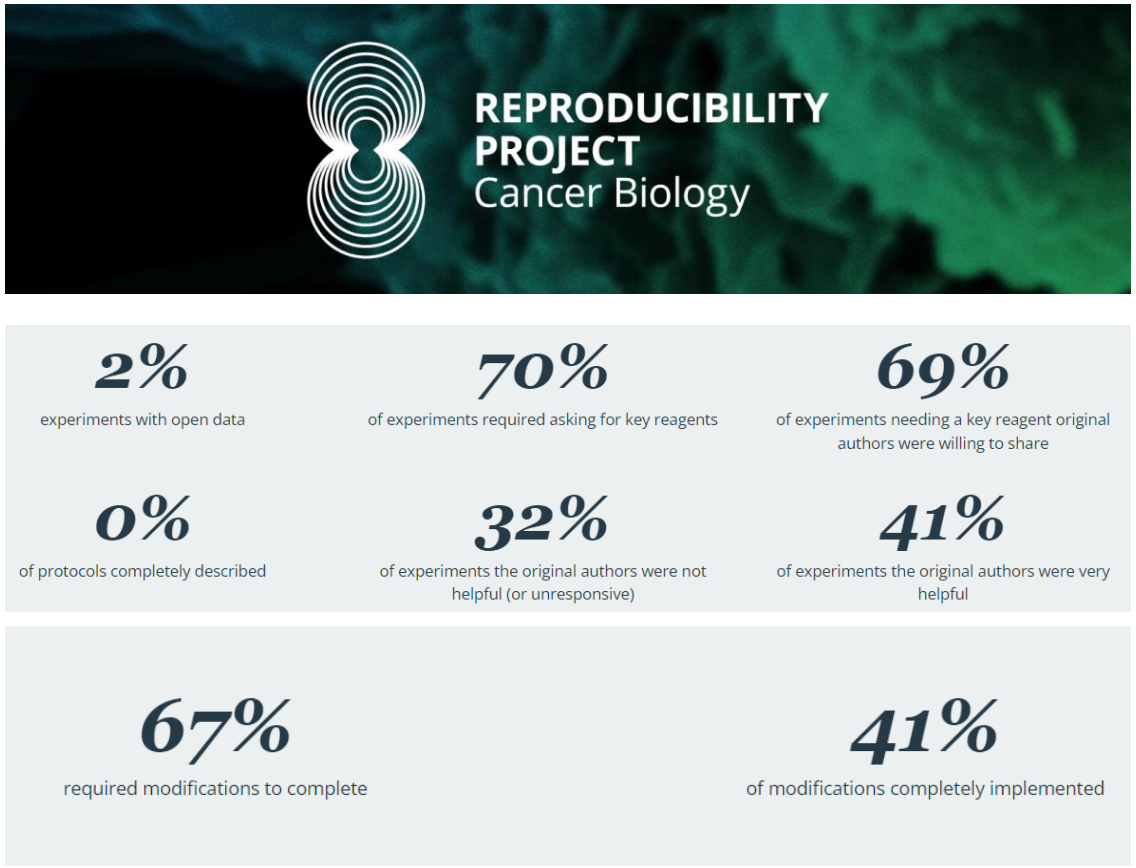
Replikační krize není jen v psychologii

Reproducibility project:
Cancer Biology (2021)

- https://www.cos.io/rpcb
- 193 navržených replikací celkem
  53 preklinických studií z let 2010–2012.

Výsledky:

- Realizace 50 replikací 23 článků
  (nedostatek informací, nespolupráce).
- Jen 46 % efektů bylo replikovaných.
- Velikost efektu o 85 % nižší.



Replikační krize není jen v psychologii

Populační genetika: analýza hlavních komponent (PCA) pro redukci informace z analýzy genomu jako „předkrok“ při analýzách (analyzovány jsou pak komponenty).

„Our findings raise concerns about the validity of results reported in the population genetics literature and related fields that place a disproportionate reliance upon PCA outcomes and the insights derived from them. We conclude that PCA may have a biasing role in genetic investigations and that ***32,000-216,000 genetic studies should be reevaluated****.“*

- Elhaik, E. (2022). Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports 12*(14683). https://doi.org/10.1038/s41598-022-14395-4
- Ale vůbec tomu nerozumím ☺ Twitter diskuze.

Nemusíme ale chodit do zahraničí…

Nemusíme ale chodit do zahraničí…

Originální článek: https://www.nature.com/articles/s41598-018-23637-3

Preprint komentáře: https://osf.io/j9xct/

Vlastně příklad dobré praxe: Byla k dispozici data, která umožnila revizi (byť ne replikovatelnost) výsledků.

Reproducibility, replicability, generalizability

Reproducibility (Reprodukovatelnost)

- *„Researcher B must have the following: (a) the **raw data**; (b) the **code book** (variable names and labels, value labels, and codes formissing data); and (c) knowledge of **the analyses** that were performed by Researcher A (e.g. the syntax of a statistics program).“*

Replicability (Replikovatelnost)

- *„The **finding can be obtained with other random samples** drawn from a multidimensional space that captures the most important facets of the research design. In psychology, the facets typically include the following: (a) **individuals** (or dyads or groups); (b) **situations** (natural or experimental); (c) **operationalizations** (experimental manipulations, methods, and measures); and (d) **time points**.“*

Generalizability (Zobecnitelnost)

- *„It does not depend on an originally unmeasured variable that has a systematic effect. In psychology, generalizability is often demonstrated by showing that a **potential moderator variable has no effect** on a group difference or correlation.“*

Asendorpf et al. (2013)

Změna paradigmatu

Pohled na celou „krizi“ se vyvíjí a dochází ke změně paradigmatu.

Ohrožení důvěry laické i odborné veřejnosti ve vědu jako takovou.

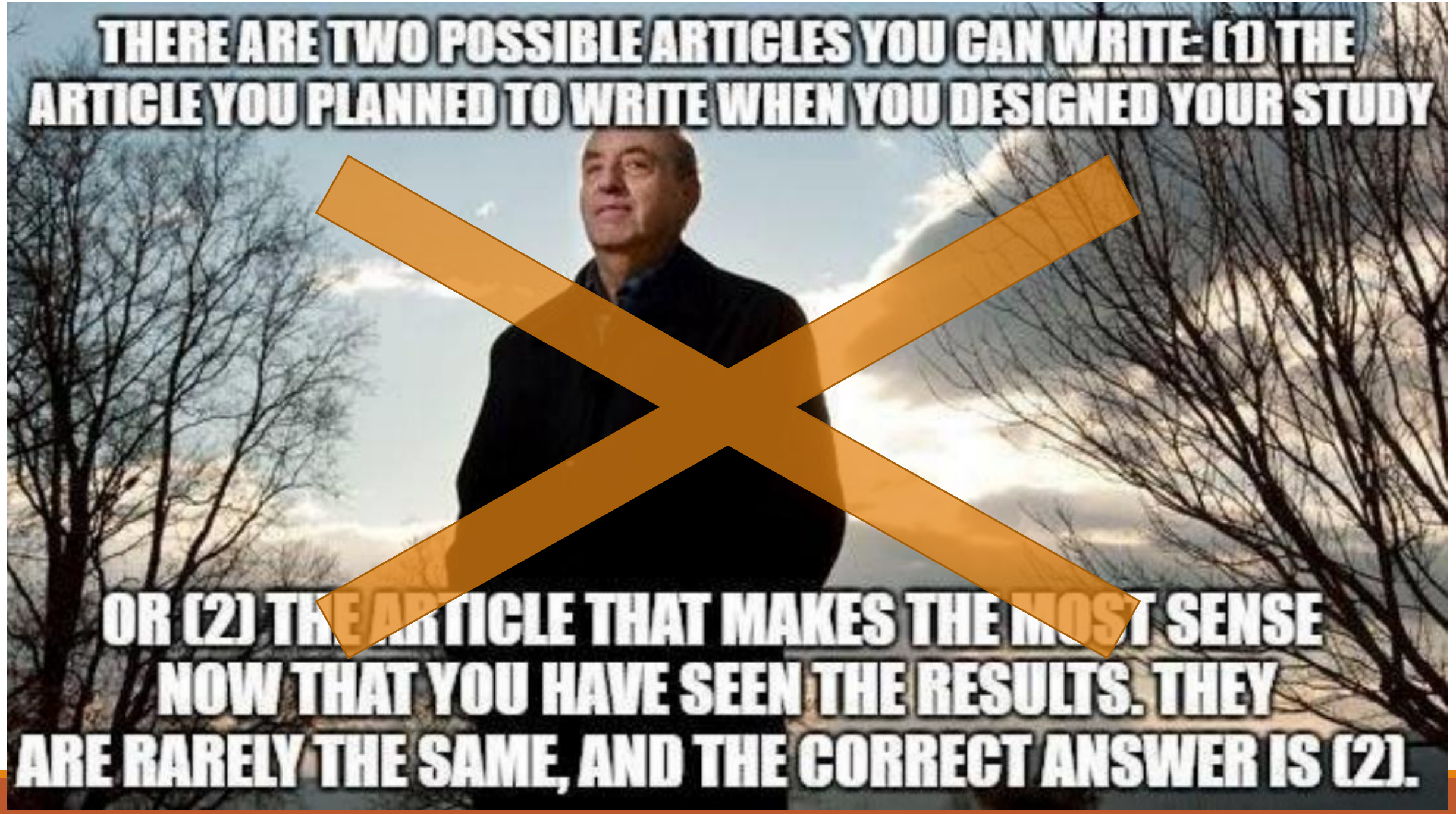
Replikační krize → krize důvěryhodnosti/zobecnitelnosti.

- replication crizis
- reproducibility crizis
- replicability crizis
- **generalizability crizis**
- **credibility crisis**

Jaké jsou podle vás příčiny?

Mimo evidentní a záměrný podvod?

Mimo obyčejnou chybu při analýze?


THERE ARE TWO POSSIBLE ARTICLES YOU CAN WRITE: (1) THE ARTICLE YOU PLANNED TO WRITE WHEN YOU DESIGNED YOUR STUDY
OR (2) THE ARTICLE THAT MAKES THE MOST SENSE NOW THAT YOU HAVE SEEN THE RESULTS. THEY ARE RARELY THE SAME, AND THE CORRECT ANSWER IS (2).

Pochybné praktiky ve výzkumu

„In a poll of more than 2000 psychologists, prevalences of 'Deciding whether to collect more data after looking to see whether the results were significant' and 'Stopping data collection earlier than planned because one found the result that one had been looking for' were subjectively estimated at 61% and 39%, respectively."

- John, Loewenstein, & Prelec, cit. dle Asendorpf et al., 2013

Questionable research practices.

Podvodné vs. pochybné jednání?

- *„Fraud is typically limited to cases in which researchers create false data."*
- *„In contrast, QRPs typically involve the exclusion of data that are inconsistent with a theoretical hypothesis. QRPs are treated differently than fraud because QRPs can sometimes be used for legitimate purposes."* (John, Loewenstein, & Prelec, 2012)

Kde je zakopaný pes?

Questionable Research Practices	OK
1. Not reporting "failed" studies.	83%
2. Not reporting DVs if not significant	92%
3. Not reporting Conditions that "did not work"	89%
4. Excluding data based on effect on p-value.	81%
5. Stopping data collection when significant.	89%
6. Reporting unexpected results "as predicted"	75%

(John, Loewenstein, & Prelec, 2012)

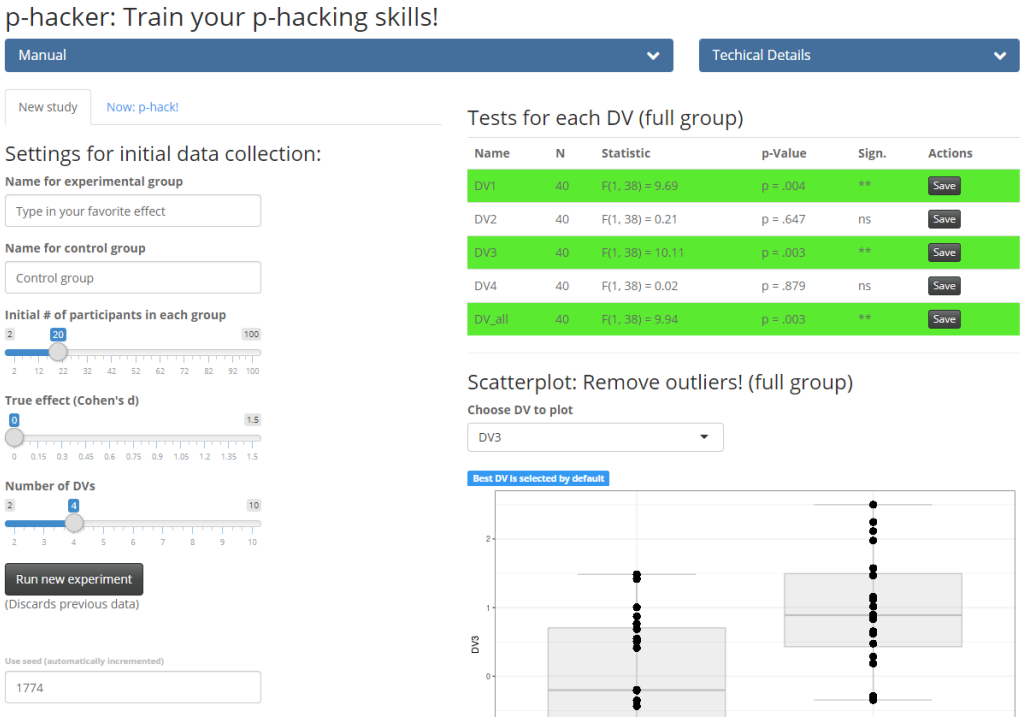
Table 1. Likelihood of Obtaining a False-Positive Result

	Significance level		
Researcher degrees of freedom	$p < .1$	$p < .05$	$p < .01$
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%

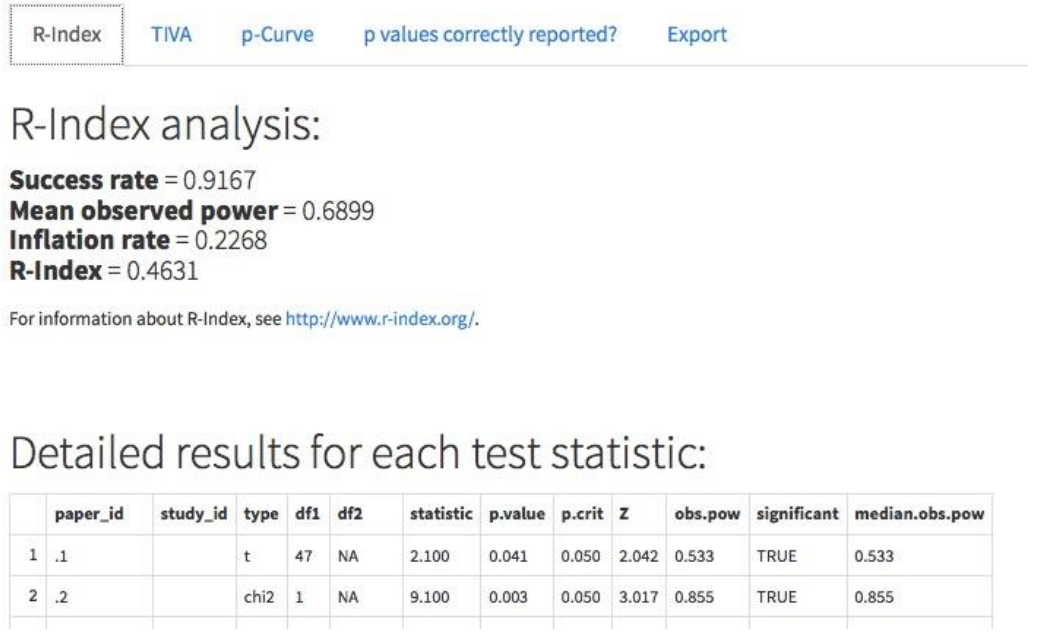
(Simmons, Nelson, & Simonsohn, 2011)

Kontrola předchozích zjištění

P-HACKER



P-CHECKER



	paper_id	study_id	type	df1	df2	statistic	p.value	p.crit	Z	obs.pow	significant	median.obs.pow
1	.1		t	47	NA	2.100	0.041	0.050	2.042	0.533	TRUE	0.533
2	.2		chi2	1	NA	9.100	0.003	0.050	3.017	0.855	TRUE	0.855

Preference pozitivních (signifikantních) výsledků

Každý vědec přirozeně chce na „něco přijít“, snaha „nalézt“ výsledek.

- Princip fungování NHST tomu nahrává.

Vědci jsou posuzovaní podle citačního „ohlasu“.

- Publikační tlak; „publish or perish“.

Studie se signifikantními výsledky citovány 1,6krát častěji (Duyx et al., 2017).

- Pokud autoři explicitně napíší, že nalezli podporu pro své hypotézy, tak dokonce 2,7krát.

Kvalita časopisu je posuzována podle citovanosti jeho studií.

Editoři proto preferují články, u kterých je vyšší pravděpodobnost citování.

To vše dohromady → **publikační zkreslení** (file-drawer effect; Rosenthal, 1972).

Důsledky publikačního zkreslení

Příklad A: ***Znáte*** skutečnou velikost efektu, d = 0,3.
Realizujete dvě studie, N_1 = 50 a N_2 = 500.

- Otázka A1: Ve které studii budete pravděpodobněji pozorovat statisticky významný efekt?
- Otázka A2: Ve které studii budete pravděpodobněji pozorovat větší velikost účinku?

Příklad B: ***Neznáte*** skutečnou velikost efektu.
Realizujete dvě studie, N_1 = 50 a N_2 = 500.

- Otázka B1: Ve které studii budete pravděpodobněji pozorovat statisticky významný efekt?
- Otázka B2: Ve které studii budete pravděpodobněji pozorovat větší velikost účinku?

Příklad C: ***Neznáte*** skutečnou velikost efektu.
V databázi naleznete dvě publikované studie, N_1 = 50 a N_2 = 500.

- Otázka C1: Ve které studii budete pravděpodobněji pozorovat statisticky významný efekt?
- Otázka C2: Ve které studii budete pravděpodobněji pozorovat větší velikost účinku?

Důsledky publikačního zkreslení

Nejen vyšší prevalence signifikantních zjištění, ale také:

- Vyšší pozorované efekty.
- Indukovaná souvislost velikosti vzorku a velikosti efektu.

Malý vzorek: QRP mají větší vliv na signifikanci výsledku.

- Např. vyřazení 2 respondentů má vliv na N = 50, nikoli na N = 500.

Malé studie jsou levné. Lze jich realizovat mnoho a publikovat ty, které „vyšly“.

False-positive rate je stejný u malých i velkých studií.
Pokud ale nastane, malá bude mít vyšší velikost efektu.

- Toto principu se využívá při identifikaci publikačního zkreslení nejen v meta-analýzách.

Nástroje k odhalení QRP

Egerův test (z-test) a funnel plot.

P-curve: Rozložení (resp. zešikmení) p-hodnot $p < 0,05$.

- Dobré rozložení: zprava zešikmené. QRP: zleva zešikmené (většina p-hodnot blízko cut-offu).

Z-curve: Srovnání pozorovaného „success-rate“ a mediánu statistické síly.

- R-index: Odhad podílu studií, které by bylo možné replikovat.

„Test of insufficient variance“ (TIVA):

- P-hodnoty převedené na z-skóry by měly být mít $SD \geq 1$.

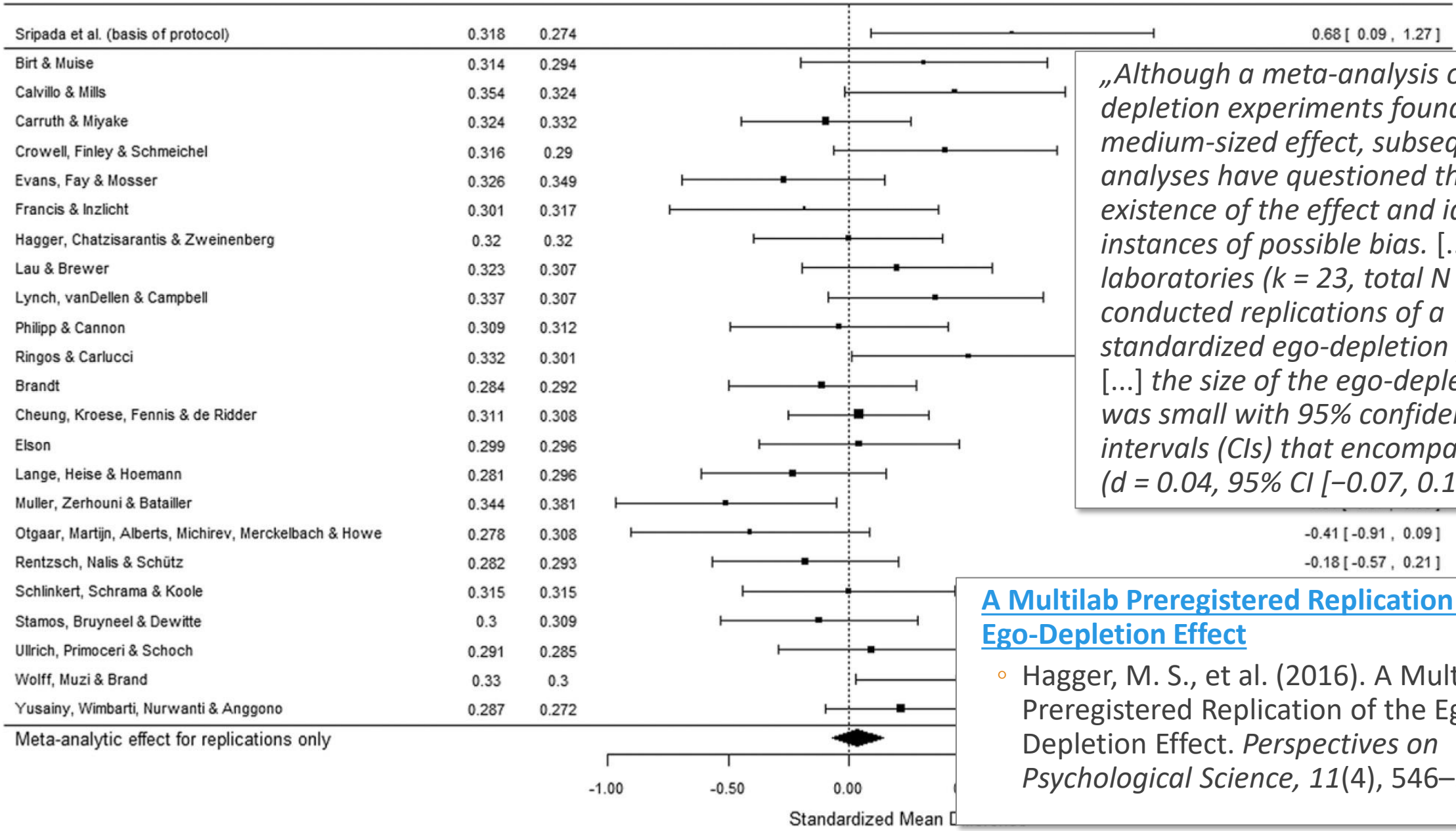
GRIM test: Detekce nemožných průměrů.

- Některé hodnoty desetinných míst nejsou přípustné v případě malých vzorků.
- http://www.prepubmed.org/grim_test/

P-checker: https://shinyapps.org/apps/p-checker/

Příklady replikačních pokusů Pokusy o další vysvětlení potíží

10+ let projektu Many Labs

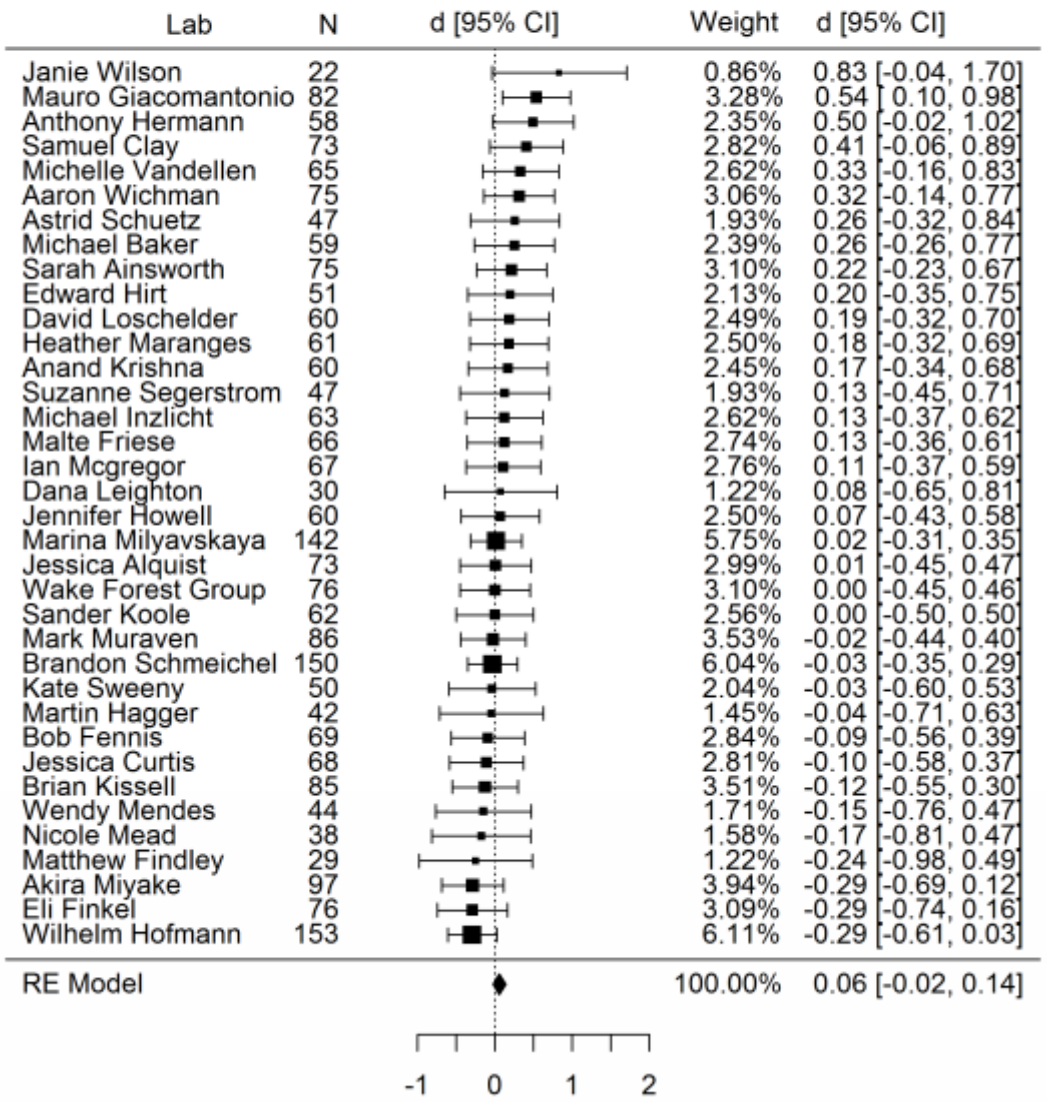


„Although a meta-analysis of ego-depletion experiments found a medium-sized effect, subsequent meta-analyses have questioned the size and existence of the effect and identified instances of possible bias. [...] *Multiple laboratories (k = 23, total N = 2,141) conducted replications of a standardized ego-depletion protocol* [...] *the size of the ego-depletion effect was small with 95% confidence intervals (CIs) that encompassed zero (d = 0.04, 95% CI [−0.07, 0.15].“*

A Multilab Preregistered Replication of the Ego-Depletion Effect

- Hagger, M. S., et al. (2016). A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspectives on Psychological Science, 11*(4), 546–573.

Figure 1. *Forest Plot of Performance Outcome by Laboratory*. The box plots and numerical values illustrate the same effect size estimates. For the plots, the size of the box represents its weighted contribution to the overall effect and its whiskers display 95% CIs. The dotted line represents a zero effect size. Numerical values show standardized mean differences between depletion and non-depletion conditions expressed in Cohen's *d* (with 95% CIs). The diamond is the overall meta-analytic effect derived from a random-effects model.



„We conducted a preregistered multi-laboratory project (k = 36; N = 3531) to assess the size and robustness of ego depletion effects using a novel replication method, termed the paradigmatic replication approach. [...] *non-significant result, d = 0.06. Confirmatory Bayesian meta-analyses using an informed prior hypothesis (δ = 0.30; SD = 0.15) found the data were four times more likely under the null than the alternative hypothesis. Hence, preregistered analyses did not find evidence for a depletion effect.“*

Vohs, K., et al. (2021). A multi-site preregistered paradigmatic test of the ego depletion effect. *Psychological Science.* https://doi.org/10.1177/0956797621989733

„This research tested variation in the replicability of 13 classic and contemporary effects across 36 independent samples totaling 6,344 participants [...] We compared whether the conditions such as lab versus online or US versus international sample predicted effect magnitudes. By and large they did not."

Many Labs 1

- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., . . . Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology, 45*(3), 142-152.

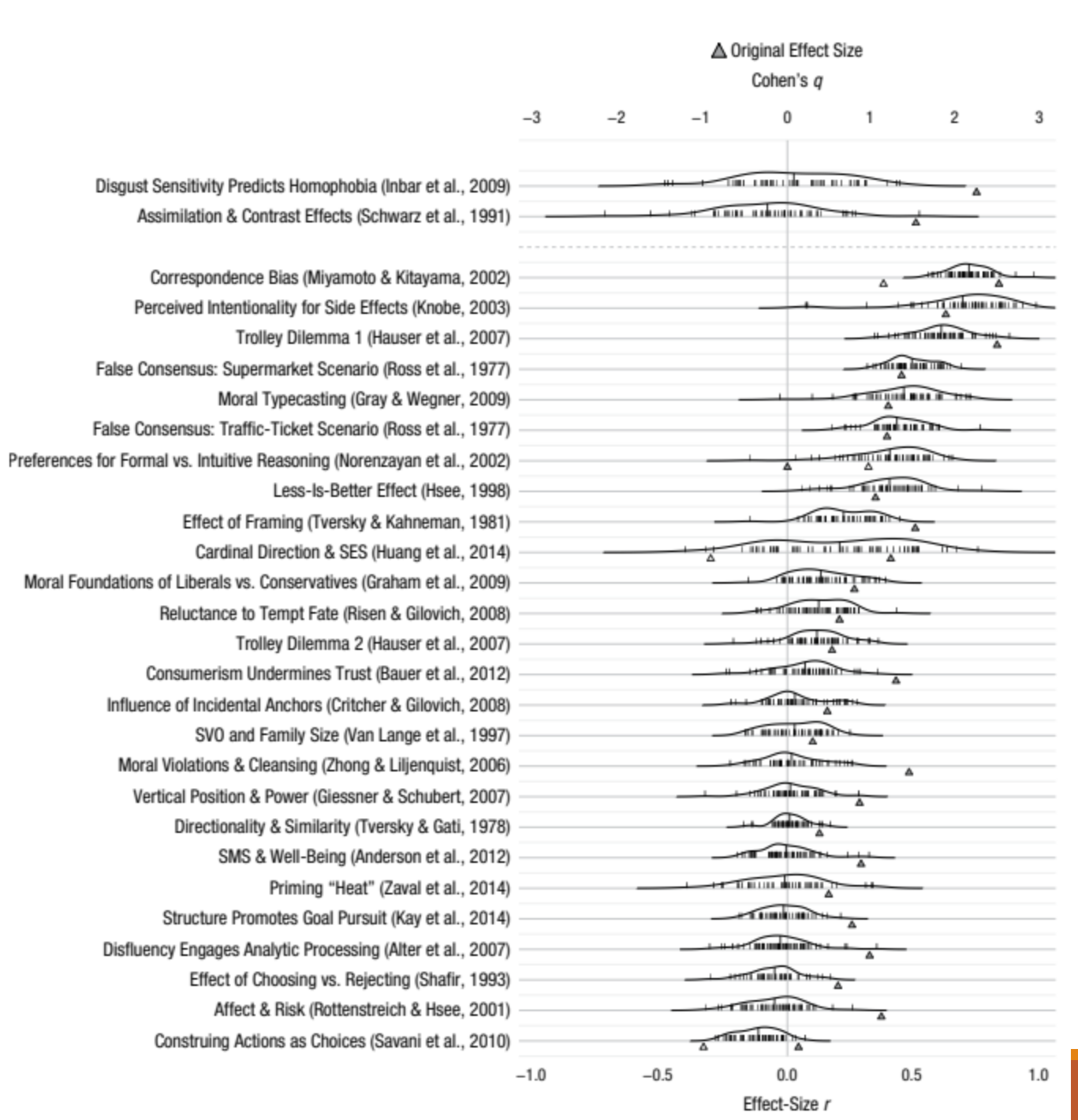
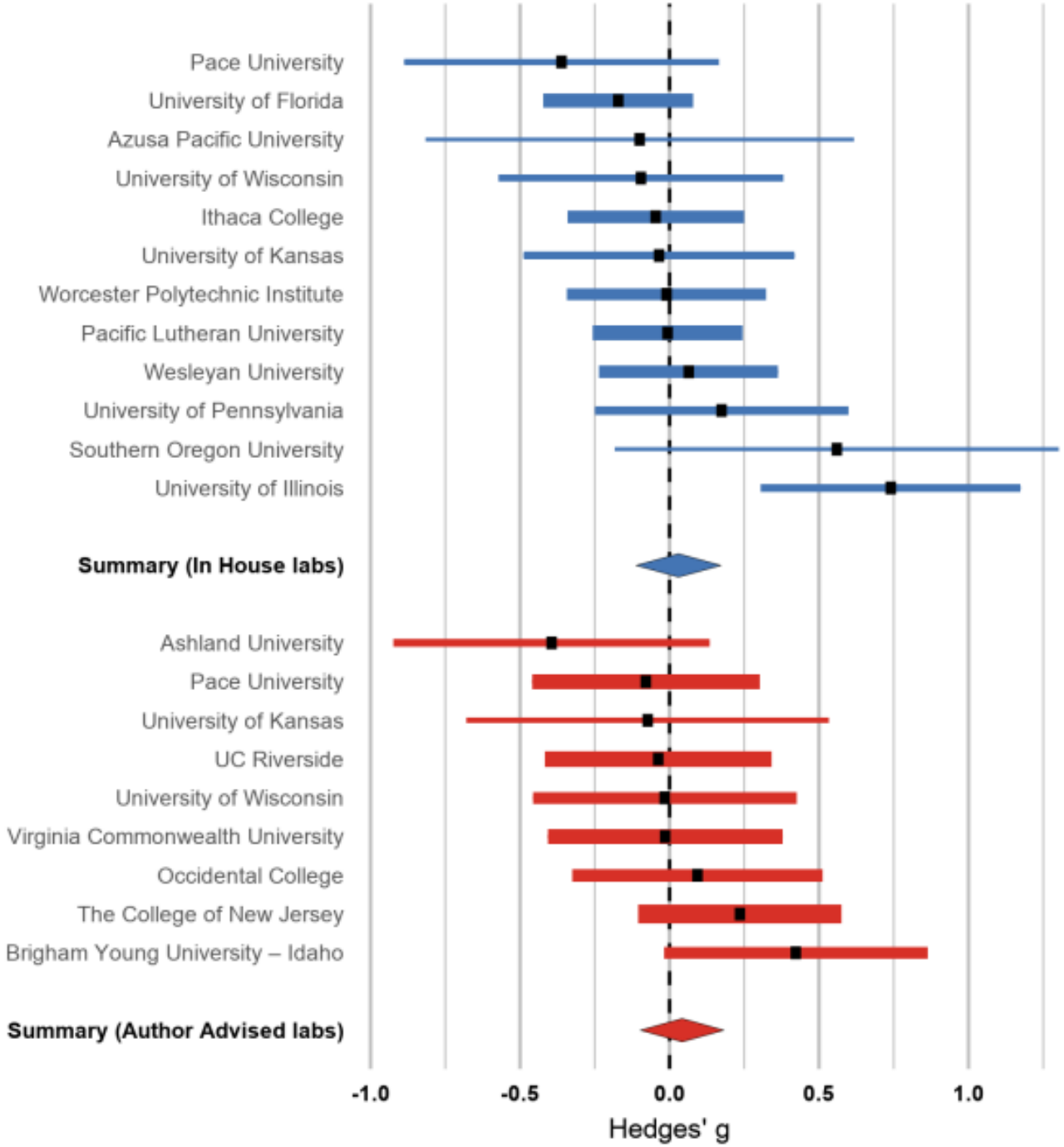


Fig. 2. Effect-size distributions for the 28 effects. The effect size for each replication sample is plotted as a short vertical line; the aggregate estimates are plotted as longer, thick vertical lines. Results for samples with fewer than 15 participants because of exclusions are not plotted, and some samples were excluded because of errors in administration. A detailed accounting of all exclusions is available at https://manylabsopenscience.github.io/ML2_data_cleaning. Positive effect sizes indicate effects consistent with the direction of the original findings in (

„Across settings, the Q statistic indicated significant heterogeneity in 11 (39%) of the replication effects, and most of those were among the findings with the largest overall effect sizes; only 1 effect that was near zero in the aggregate showed significant heterogeneity according to this measure. [...] Moderation tests indicated that very little heterogeneity was attributable to the order in which the tasks were performed or whether the tasks were administered in lab versus online. [...] Cumulatively, variability in the observed effect sizes was attributable more to the effect being studied than to the sample or setting in which it was studied."

Many Labs 2

- Klein, R. A., et al. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science, 1*(4), 443–490.



„We (N = 21 Labs and N = 2,220 participants) experimentally tested whether original author involvement improved replicability of a classic finding from Terror Management Theory (Greenberg et al., 1994). Our results were non-diagnostic of whether original author involvement improves replicability because we were unable to replicate the finding under any conditions. This suggests that the original finding was either a false positive or the conditions necessary to obtain it are not yet understood or no longer exist.“

Many Labs 4

- Klein, R. A., et al. (2019, December 11). Many Labs 4: Failure to Replicate Mortality Salience Effect With and Without Original Author Involvement.
- preprint

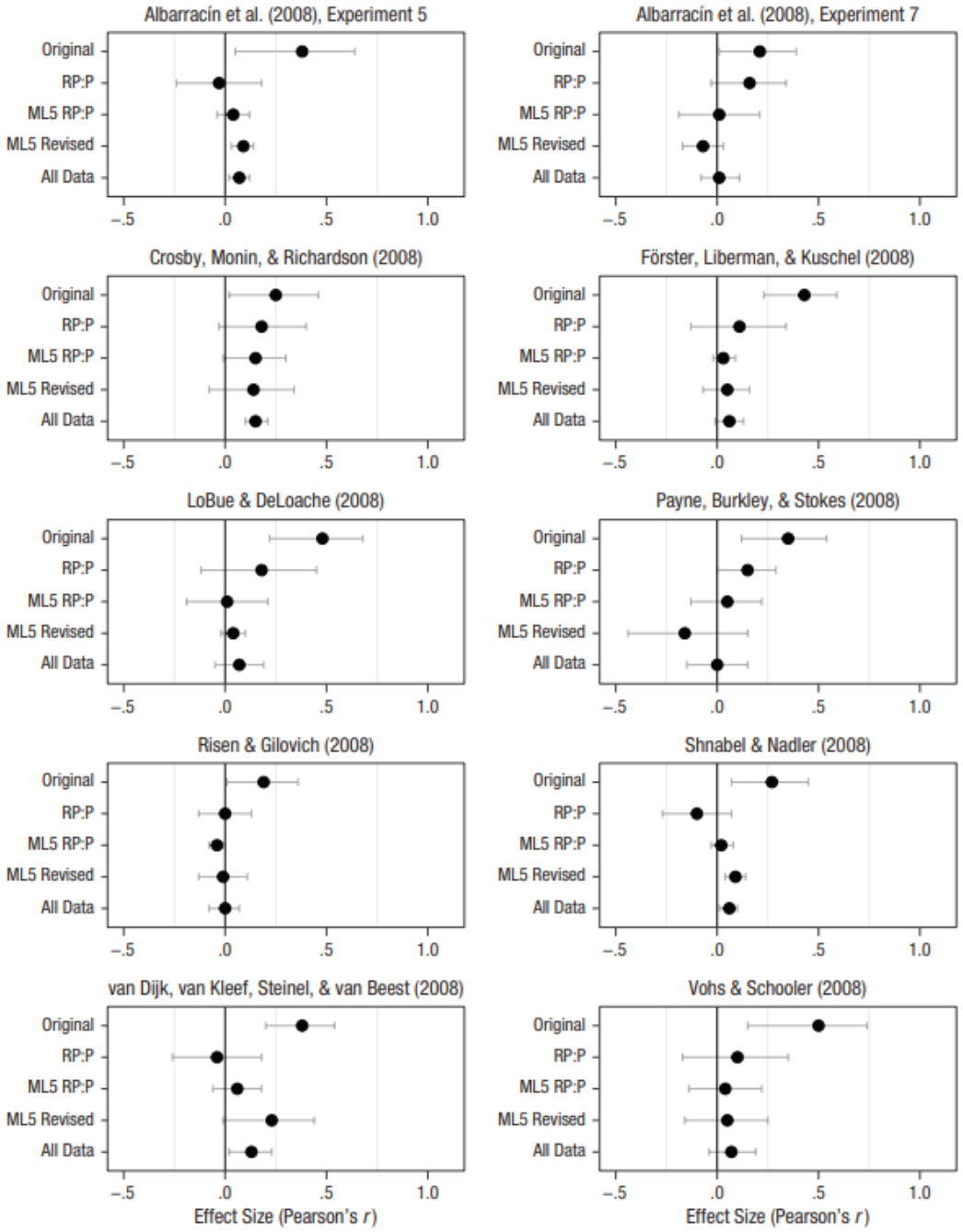


Fig. 2. Effect sizes from the 10 original studies and their replications in the Reproducibility Project: Psychology (RP:P; Open Science Collaboration, 2015) and the Many Labs 5 (ML5) protocols. The "All Data" results are estimates from random-effects meta-analyses including the original studies and their replications. Error bars represent 95% confidence intervals.

„If these [replication] *studies use methods that are unfaithful to the original study or ineffective in eliciting the phenomenon of interest, then a failure to replicate may be a failure of the protocol rather than a challenge to the original finding. Formal pre-data-collection peer review by experts may address shortcomings and increase replicability rates.* [...] *Overall, following the preregistered analysis plan, we found that the revised protocols produced effect sizes similar to those of the RP:P protocols (Δr = .002 or .014, depending on analytic approach).“*

Many Labs 5

- Ebersole, C.R., et al. (2020). Many Labs 5: Testing Pre-Data-Collection Peer Review as an Intervention to Increase Replicability. Advances in Methods and Practices in Psychological Science, 3(3), 309–331.

Co tedy dnes víme?

Replikovatelnost efektu je důsledkem efektu, nikoli intervenujících proměnných.

- **Many Labs 1 (2014):** Jazyk či konkrétní laboratoř nemá vliv.
- **Many Labs 2 (2018):** Charakteristiky laboratoře nemají vliv, heterogenita efektů se však různí.
- **Many Labs 3 (2016):** Výsledky na studentských populacích vycházejí stejně v průběhu roku (např. semestr vs. zkouškové).
- **Many Labs 4 (2022):** Účast původního autora nemá vliv.
- **Many Labs 5 (2020):** Úpravy výzkumného protokolu nemají vliv.

„After 10 Years, ‘Many Labs’ Comes to an End – But Its Success Is Replicable“

- https://news.virginia.edu/content/after-10-years-many-labs-comes-end-its-success-replicable

Měření v psychologii a replikovatelnost

Lilienfeld, S. O., & Strother, A. N. (2020). **Psychological measurement and the replication crisis: Four sacred cows.** *Canadian Psychology, 61*(4), 281–288. https://doi.org/10.1037/cap0000236

„Questionable Measurement Practices“ (QMP)

Namísto „measurement“ často spíše *„schmeasurement“* (Flake & Fried, 2020).

Lilienfeld & Strother (2020): Nedostatečná kvalita měření...

- ... snižuje věrohodnost výzkumných zjištění a ohrožuje interní validitu výzkumu;
- ... snižuje a zkresluje velikosti pozorovaných efektů;
- ... a snižuje reprodukovatelnost a hlavně zobecnitelnost výzkumných zjištění.

QMP mohou být jednou z dílčích příčin krize zobecnitelnosti.

V důsledku pak nedostatky v *měření* snižují kvalitu *vědy*, protože měření v širším slova smyslu je základním nástrojem vědy.

„Posvátné krávy“ měření v psychologii

1. Obsahová validita a spoléhání se na „název“ škál.
   - Škály se stejným názvem nemusí měřit to stejné.
   - Pro připomenutí: klasická testová teorie a operacionalismus.

2. Ignorování chyby měření a reliability v laboratorních experimentech.
   - Přesvědčení, že pro výzkum postačuje nižší reliabilita (rovněž i Helmstadter).
   - Behaviorální pozorování (vysoce reliabilní) není totožné s měřeným rysem (vztah může být vágní).
   - A jaká je reliabilita experimentální manipulace?

4. Důraz na konvergentní, nikoli divergentní validitu.
   - Konstruktově irelevantní rozptyl, nedostatek diferenciální validity.
   - Potíže zejména při výzkumu silně korelovaných jevů.

(3. Náročnost sběru dat opravňuje malé velikosti vzorku.)

Lilienfeld, S. O., & Strother, A. N. (2020). **Psychological measurement and the replication crisis: Four sacred cows.** *Canadian Psychology, 61*(4), 281–288. https://doi.org/10.1037/cap0000236

Krize replikovatelnosti: jeden z příznaků **krize zobecnitelnosti**

Yarkoni, T. (2020). **The generalizability crisis.** *Behavioral and Brain Sciences* [preprint], 1–37. https://doi.org/10.1017/S0140525X20001685

Psychologický výzkum je příliš orientovaný na pozorované proměnné namísto na konstrukty.

- 1. Nedostatek konstruktové validity ve smyslu Cronbacha a Meehla.
- 2. Zanedbání hypotetických zdrojů variability výsledků.

Statistické modely jsou jen alternativním „jazykem“ k popisu skutečnosti.

- Při „překladu“ našich otázek do jazyka statistiky a výsledků zpět dochází k chybám.

Doporučuji Yarkoniho číst **až po** přednáškách o epistemologii a teorii zobecnitelnosti.

Klíčové příznaky krize zobecnitelnosti

#1: Psychologové zanedbávají, že různé stimuly, položky dotazníku, operacionalizace konstruktů apod. jsou pouze „vzorky“ z univerza/domény „přípustných“ vzorků.

- Při „překladu“ VO do statistického modelu nejsou operacionalizovány informace o tomto „náhodném“ výběru vzorku pozorování.
- Při překladu výsledků zpět nejsou brány v potaz limity vyplývající z operacionalizace.

#2: Ignorace náhodného výběru zkresluje odhady parametrů. Druhy efektů[1]:

- **Pevné (fixed) efekty:** zpravidla zkoumaný efekt. Není vybrán z domény, je specifický pro danou situaci. Výsledky *nechceme generalizovat* na jiné pevné efekty.
- **Náhodné (random) efekty:** kontrolují náhodu spjatou s výběrem prvků z domény. *Chceme zobecňovat* efekt i na jiné prvky/výběry z dané domény.

„**Fixed-effect fallacy**“: V psychologii bývá zpravidla kontrolovaná náhoda spjatá pouze s between-subject variabilitou (lidmi/subjekty).

- Méně často se situací, laboratoří, stimuly a podobně („stimulus-as-fixed effect fallacy“).

[1] Ve shodě s Yarkonim (2020) používám terminologii generalizovaného lineárního smíšeného modelu (GLMM).

Příklad 1: Stroopův efekt

Příklad: Stroopův efekt.

- Simulace: 20 simulovaných datasetů o 20 osobách.
- Osa X: pozorovaný efekt ve studii.
- Osay Y: číslo experimentu.

Vlevo: between-subject variabilita je ignorovaná.

- Heterogenní výsledky studií.
- Neumožňuje zobecňovat na lidi obecně, ale jen „uvnitř“ vzorku.

Vpravo: Rozdíl lidí byl do modelu vložen jako náhodný efekt.

- Homogenní výsledky studií.
- Lze zobecňovat na lidi obecně v dané populaci.

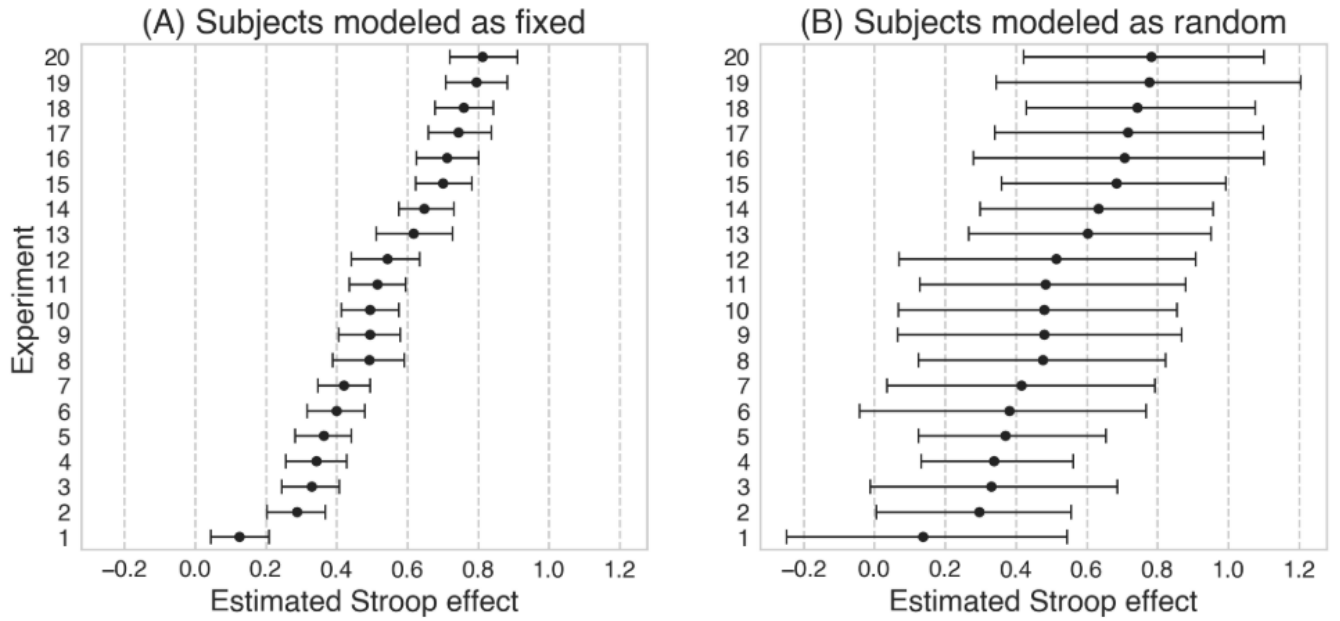


Figure 1: Co
a simulated St
ground truth f
posterior dens
ordered by the magnitude of the point estimate for visual cla
does not account for random subject sampling, and consequer
effect of interest. (B) The random-effects specification in Eq.
appropriately calibrated uncertainty estimates.

$$y_{ij} = \beta_0 + \beta_1 X_{ij} + e_{ij}$$
$$e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$$

del specific
subjects r
l experime
condition

$$y_{ij} = \beta_0 + \beta_1 X_{ij} + u_{0i} + u_{1i} X_{ij} + e_{ij}$$
$$u_{0i} \sim \mathcal{N}(0, \sigma_{u_0}^2)$$
$$u_{1i} \sim \mathcal{N}(0, \sigma_{u_1}^2)$$
$$e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$$

Příklad 1: Stroopův efekt

Yarkoni (2020, pp. 6):

- *„... it is the mismatch between our generalization intention and the model specification that introduces **an inflated risk of inferential error**, and not the model specification alone.“*
- *„Empirical studies in domains ranging from social psychology to functional MRI have demonstrated that test **statistic inflation of up to 300% is not uncommon**, and that, under realistic assumptions, **false positive rates in many studies could easily exceed 60%** (Judd et al., 2012; Westfall, Nichols, & Yarkoni, 2016; Wolsiefer, Westfall, & Judd, 2017).“*

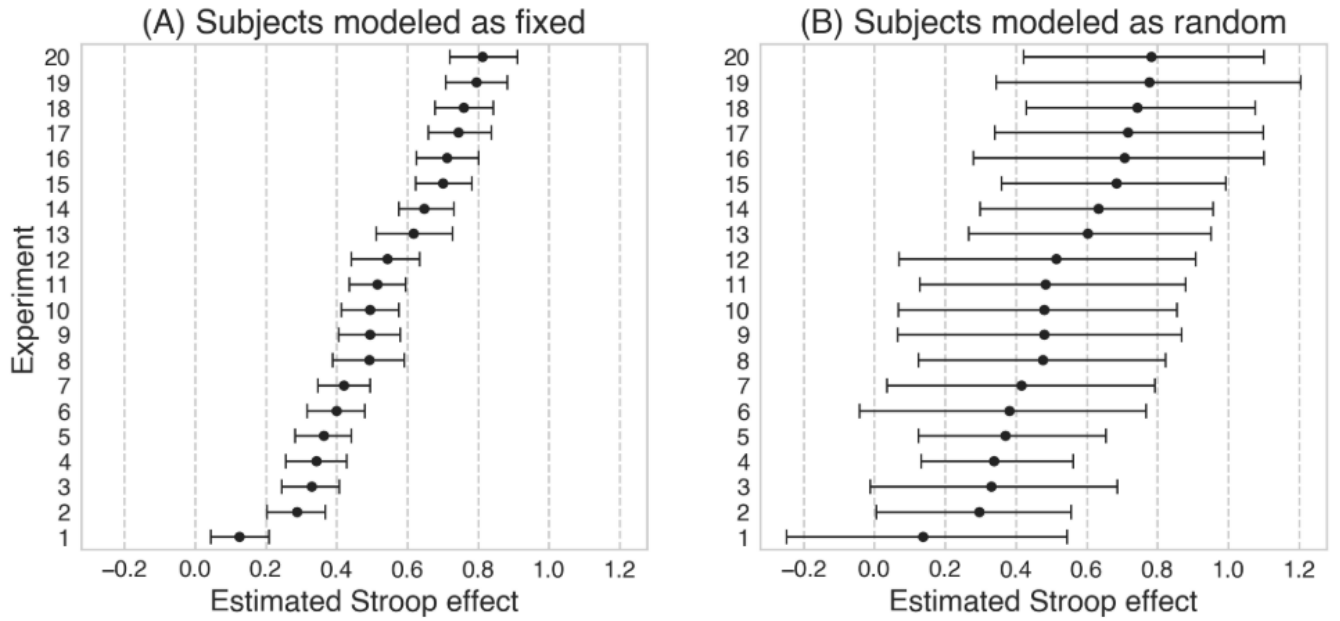


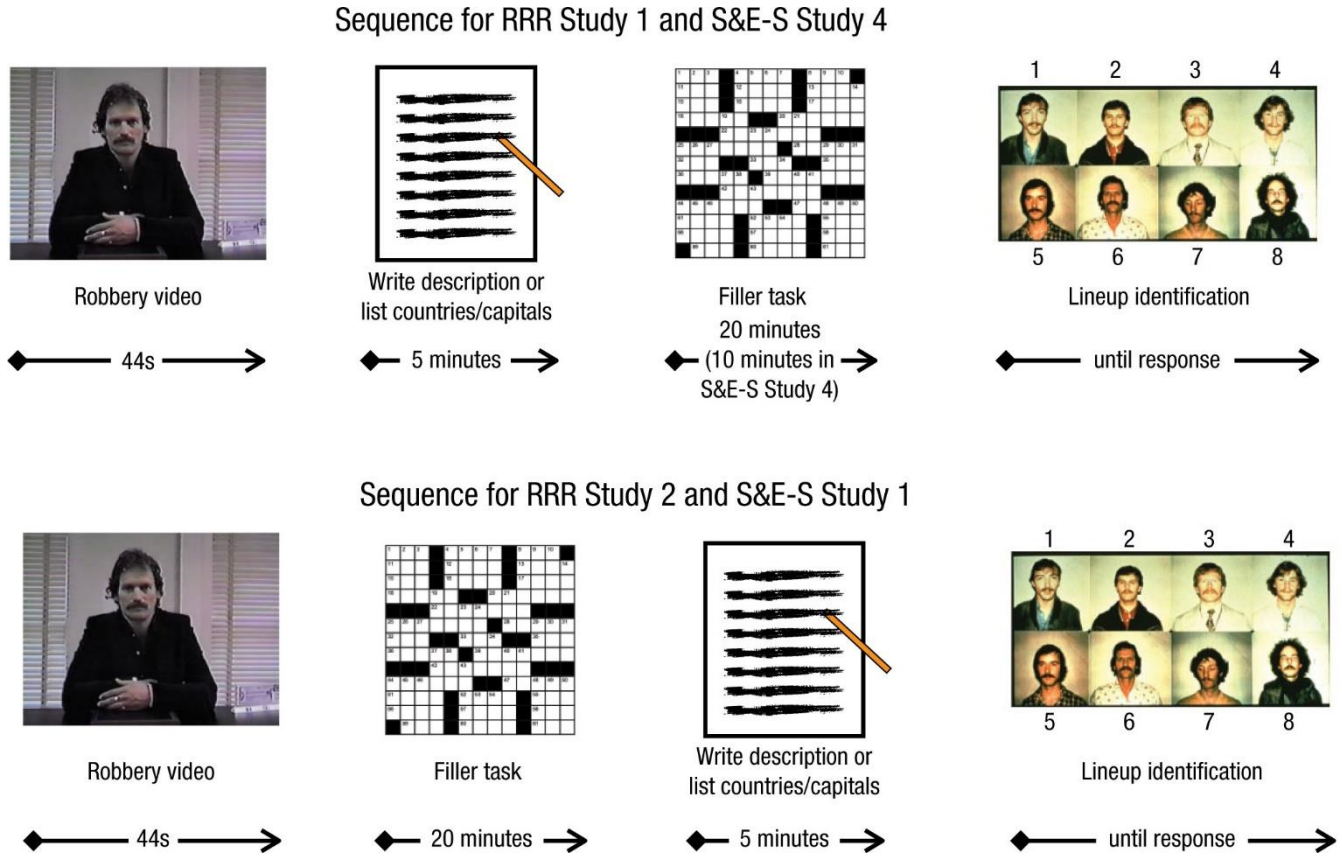
Figure 1: Consequences of mismatch between model specification and generalization intention. Each row represents a simulated Stroop experiment with $n = 20$ new subjects randomly drawn from the same global population (the ground truth for all parameters is constant over all experiments). Bars display the estimated Bayesian 95% highest posterior density (HPD) intervals for the (fixed) condition effect of interest in each experiment. Experiments are ordered by the magnitude of the point estimate for visual clarity. (A) The fixed-effects model specification in Eq. (1) does not account for random subject sampling, and consequently underestimates the uncertainty associated with the effect of interest. (B) The random-effects specification in Eq. (2) takes subject sampling into account, and produces appropriately calibrated uncertainty estimates.

Příklad 2: Verbal overshadowing

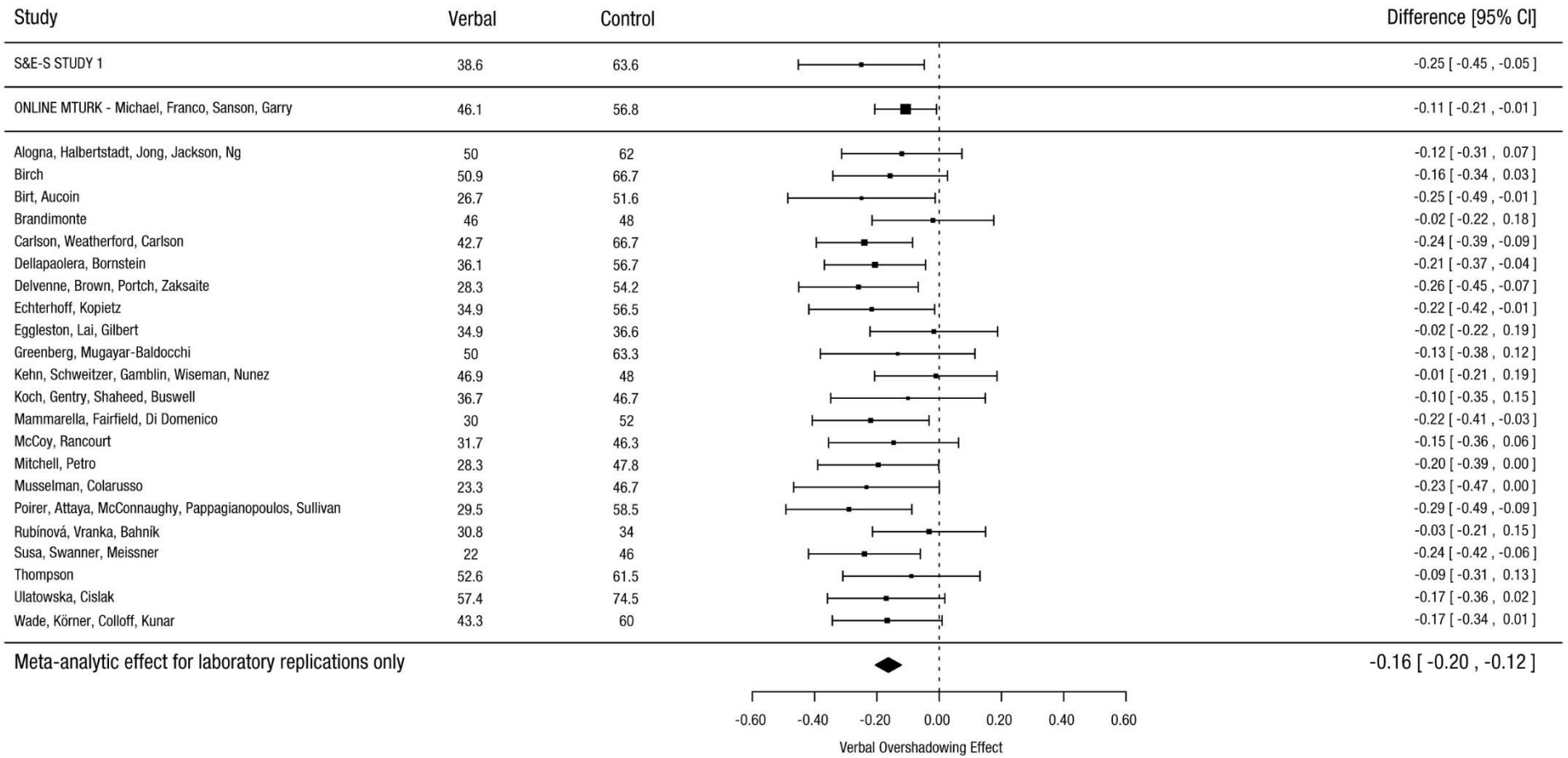
Velká replikační studie „verbálního zastínění“.

- Replikace: Alogna a kol. (2014).
- Originální studie:
  Schooler a Engstler-Schooler (1990)
- 31 laboratoří, N_{tot} > 2000.

„Original authors showed that participants who were asked to verbally describe the appearance of a perpetrator caught com- mitting a crime on video showed poorer recognition of the perpetrator following a delay than did partic- ipants assigned to a control task (naming as many countries and capitals as they could).“



Příklad 2: Verbal overshadowing



Příklad 2: Verbal overshadowing

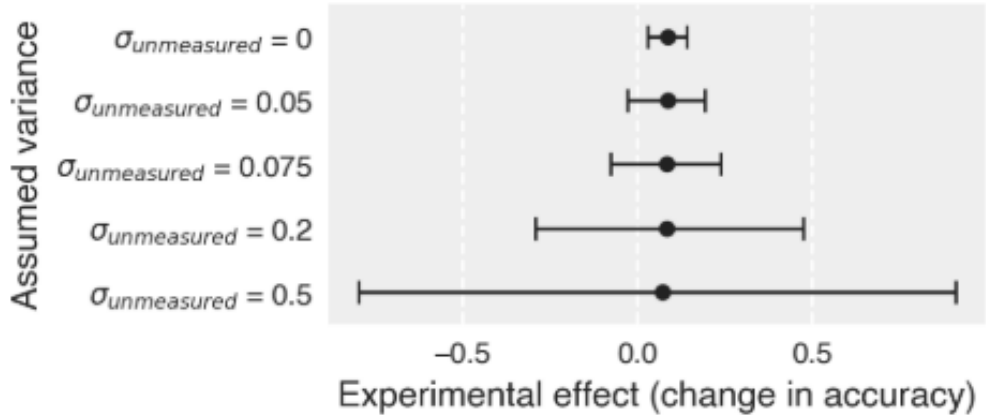
Silný důkaz pro existenci efektu. Sice nižší než originální, ale rostoucí v čase.

Nulová heterogenita výsledků napříč laboratořemi a to včetně MTurk, $I^2 = 0$.

Ale: **Ve shodě s originálními autory pouze jediná nahrávka a jediný line-up.**

- *„The strict conclusion* [...] *is that there is at least one particular video containing one particular face that, when followed by one particular lineup of faces, is more difficult for participants to identify if they previously verbally described the appearance of the target face than if they were asked to name countries and capitals. This narrow conclusion does not preclude the possibility that the observed effect is specific to this one particular stimulus, and that many other potential stimuli the authors could have used would have eliminated or even reversed the observed effect.“* (Yarkoni, 2020, pp. 8).

Pokud by nekontrolované rozdíly ve stimulech (tvářích) měly velmi malý vliv na pozorování SD=0,05 (ve srovnání se zvýšením přesnosti o cca 0,1), souhrnný efekt přestane být signifikantní.



Doporučení pro zvýšení replikovatelnosti psychologického výzkumu

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., van Aken, M. A. G., Weber, H., & Wicherts, J. M. (2013). **Recommendations for Increasing Replicability in Psychology.** *European Journal of Personality, 27*(2), 108–119. https://doi.org/10.1002/per.1919

Doporučení – shrnutí

Transparence.

- Sdílení analytických skriptů, dat, laboratorních protokolů apod.

Preregistrace a Registered report.

- Snižuje počet df výzkumníka.
- Zamezuje vydávání exploračních zjištění za konfirmační testy hypotéz.

Open Science Framework: www.osf.io

- Případně www.aspredicted.org

Kašlete na akademické frčky a dělejte dobrou vědu! ☺

- Vocaď pocaď.

A 21 Word Solution

Simmons, Joseph P. and Nelson, Leif D. and Simonsohn, Uri, A (2012). 21 Word Solution. *SSRN*. http://dx.doi.org/10.2139/ssrn.2160588

Choir: There is no need to wait for everyone to catch-up with your desire for a more transparent science. If *you* did not *p*-hack a finding, *say it*, and your results will be evaluated with the greater confidence they deserve.

If you determined sample size in advance, *say it.*
If you did not drop any variables, *say it.*
If you did not drop any conditions, *say it.*

These 21 words in a Methods section can *say it* succinctly:

"We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study."

When needed, supplemental materials can be used to ensure the 21 words are accurate.
When sample size is not determined in advance, one could write:

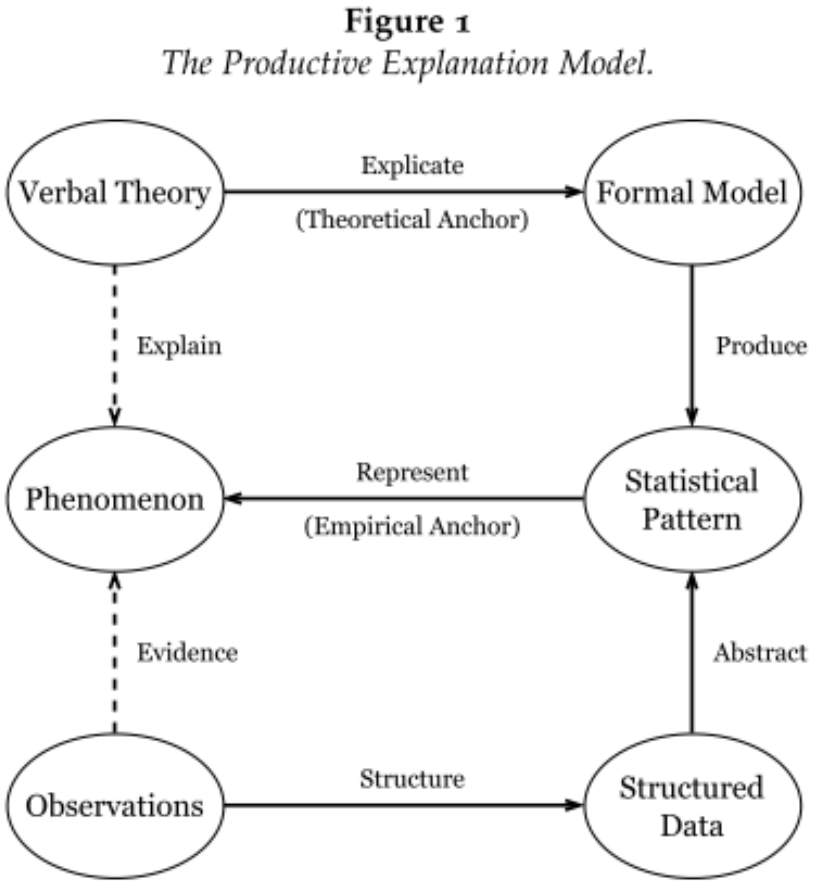
"We added 50 observations after analyzing the first 100".

Formalizované teorie v psychologii

Popis teorie pomocí formálního (matematického) jazyka.

Průlomový článek:

- van Dongen, N., van Bork, R., Finnemann, A., Haslbeck, J. M. B., van der Maas, H. L. J., Robinaugh, D. J., de Ron, J., Sprenger, J., & Borsboom, D. (2024). Productive explanation: A framework for evaluating explanations in psychological science. *Psychological review*, 10.1037/rev0000479. Advance online publication. https://doi.org/10.1037/rev0000479



Formalizované teorie v psychologii

Popis teorie pomocí formálního (matematického) jazyka.

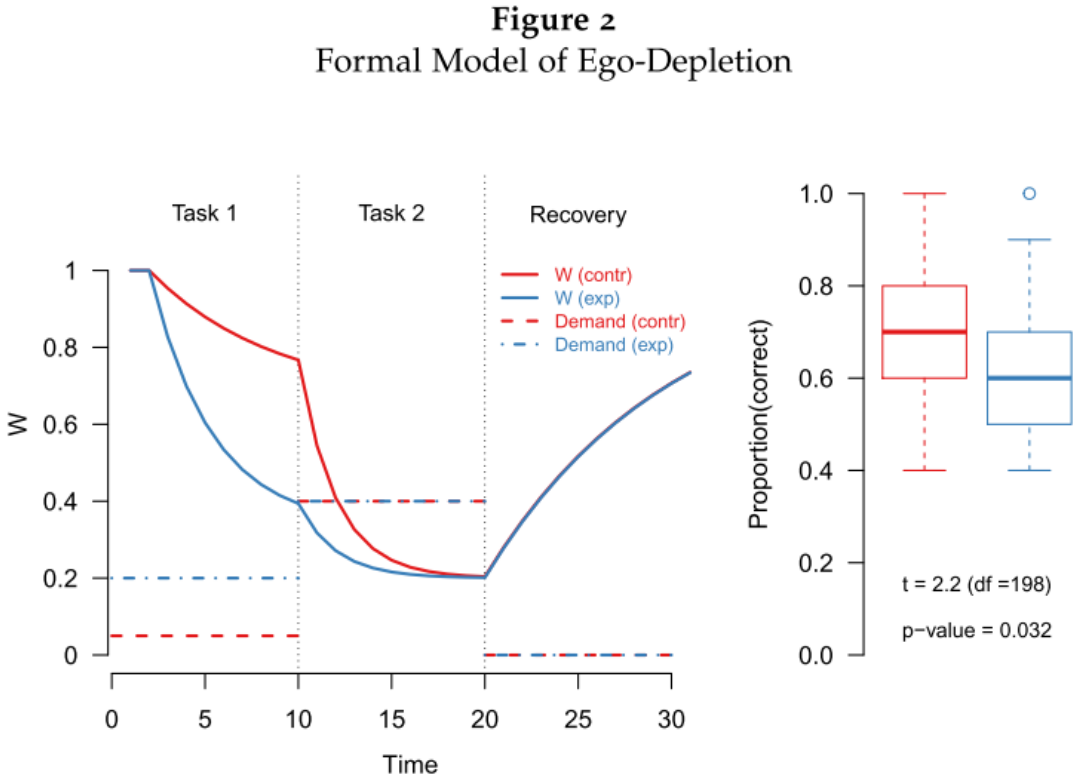
Ukazuje se totiž, že řada psychologických teorií je nejenže neplatných, ale přímo **netestovatelných**.

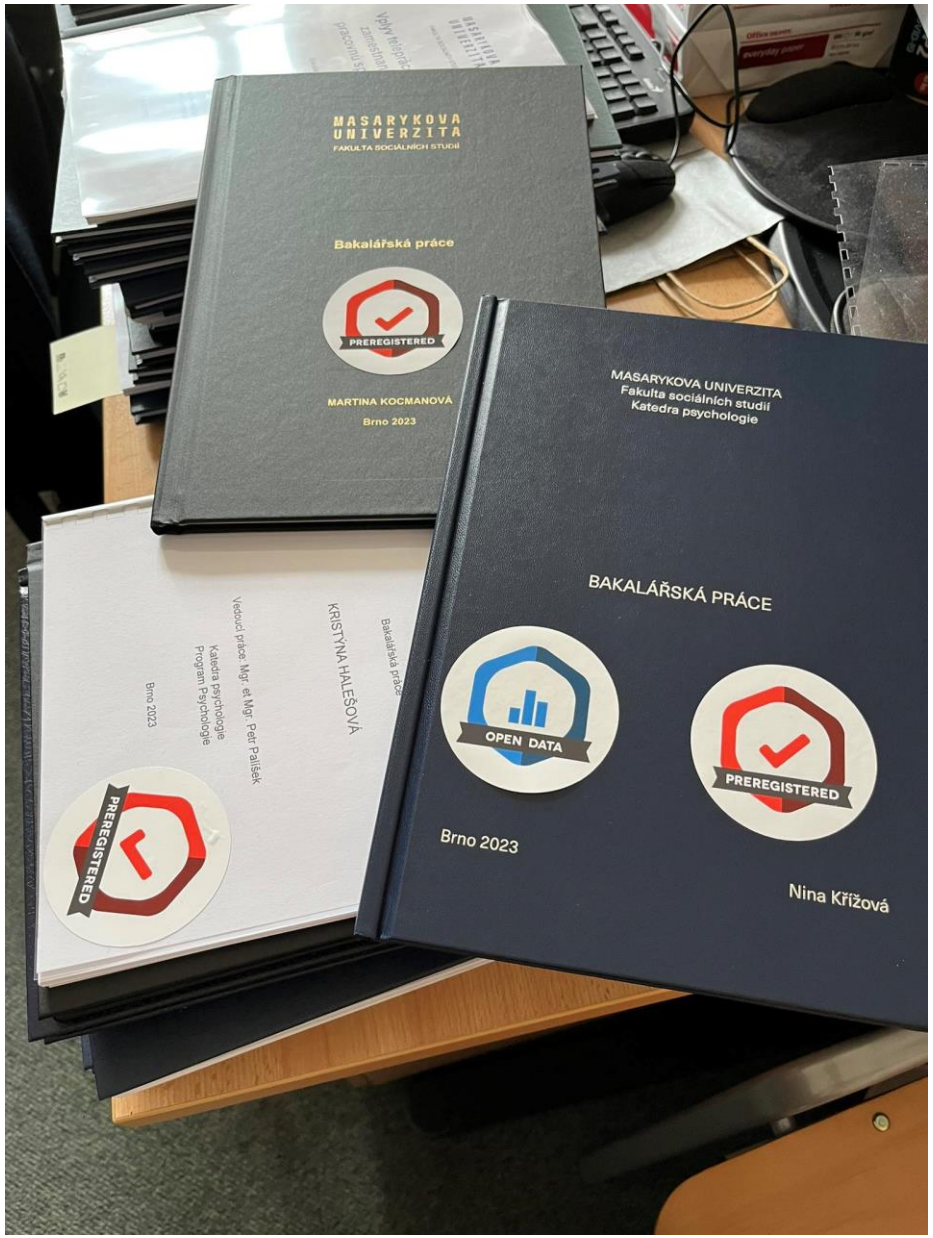
- Popper: Nefalzifikovatelné → nevědecké.

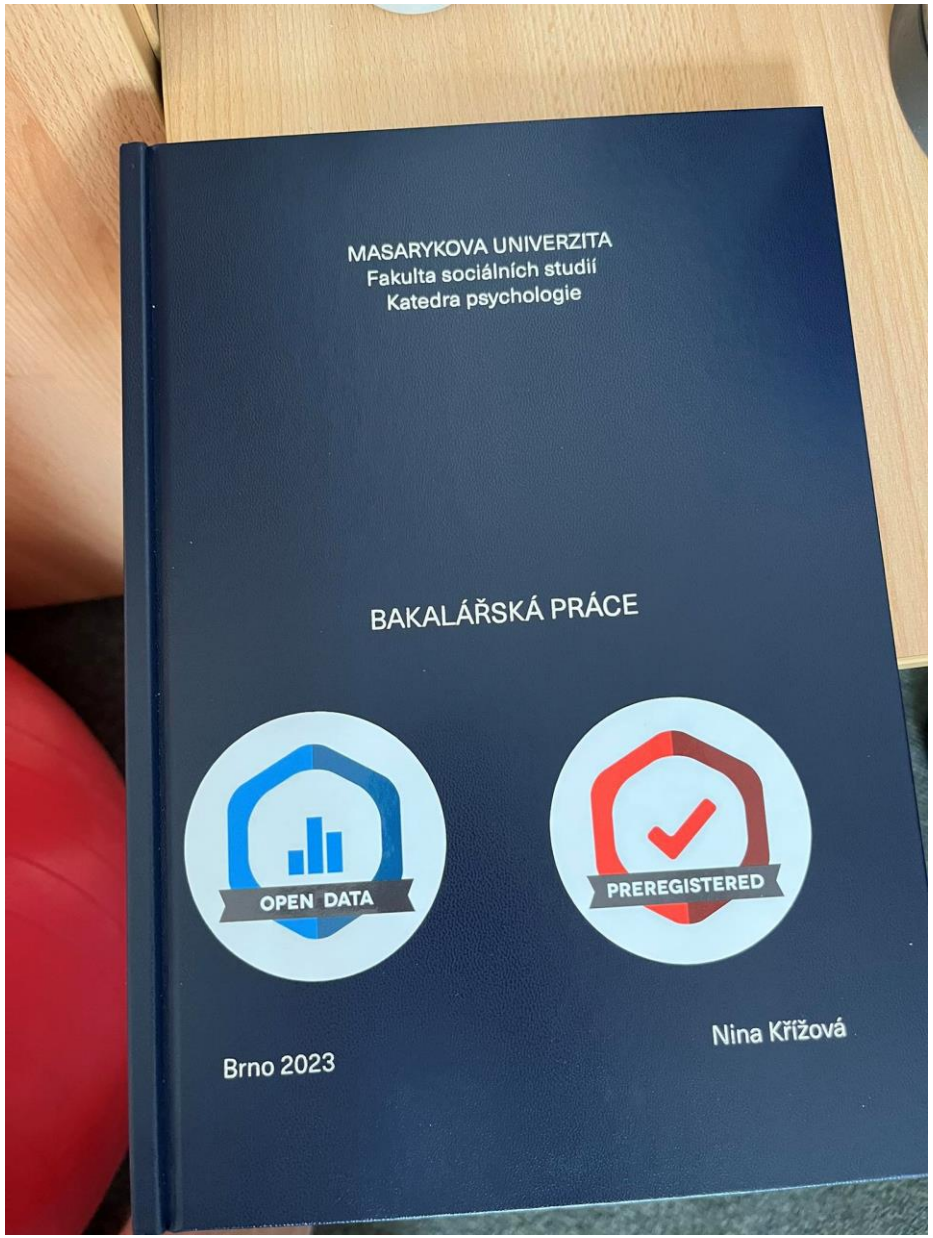
Případová studie: neúspěšné replikace ego-depletion ani nemohly být úspěšné – chybný design (Hagger et al., 2016, a Vohs, et al., 2021).

Postup navržený van Dongen et al. (2024):

- Zajišťuje testovatelnost teorie.
- Zamezuje vágnosti ve formulaci.




MASARYKOVA UNIVERZITA
FAKULTA SOCIÁLNÍCH STUDIÍ
Bakalářská práce
PREREGISTERED
MARTINA KOCMANOVÁ
Brno 2023
MASARYKOVA UNIVERZITA
Fakulta sociálních studií
Katedra psychologie
BAKALÁŘSKÁ PRÁCE
OPEN DATA
PREREGISTERED
Brno 2023
Nina Křížová
KRISTYNA HALEŠOVÁ
Vedoucí práce: Mgr. et Mgr. Petr Paillšek
Katedra psychologie
Program Psychologie
Brno 2023
PREREGISTERED


MASARYKOVA UNIVERZITA
Fakulta sociálních studií
Katedra psychologie
BAKALÁŘSKÁ PRÁCE
OPEN DATA
PREREGISTERED
Brno 2023
Nina Křížová

Doporučení: Design a analýza

Zmenšit chybu měření

- ... zvýšením velikosti vzorku;
- ... zvýšením statistické síly;
- ... zvýšením reliability měřícího nástroje;
- ... korektním užíváním korekcí pro vícenásobná srovnání,
  - Užívání postupů typu Bonferroniho korekce snižuje statistickou sílu

Od "p < 0,05" k...

- ... reportování skutečné velikosti p-hodnoty;
- ... důrazu na ukazatele velikosti účinku;
- ... důrazu na intervaly spolehlivosti apod.

Doporučení: Publikační proces

Autoři studií, výzkumníci: **transparence**.

- Literature review ve vztahu k dosavadnímu stavu replikace.
  - Existují dřívější replikační studie? Podařilo se původní výsledek replikovat? Apod.
- Zdůvodnění volby velikosti vzorku
- Zveřejnění dat, postupů analýz, work-in-progress, pre-registrací
- Provádění replikací, účast na diskuzích odborné veřejnosti atd.

Žurnály, recenzenti, editoři: **Podpora dobrých výzkumných praktik.**

- Publikování replikací a podpora autorů v této činnosti
- Ústup od konfirmačního zkreslení v publikačním procesu

Doporučení: Vyučující metodologie

Aneb: **Co mají studenti chtít po svých učitelích?**

Rigorózní výuka metodologie, statistické analýzy dat apod.

- Statistická síla, velikost účinku, zobecnitelnost atd.
- Informace o replikovatelnosti efektů při výuce jiných kurzů.

Podpora **transparentnosti**.

- Publikování dat, skriptů apod., analýza takovýchto souborů.

Podpora **studentských replikací**.

- Přínos pro studenty i pro obor.

Podpora **kritického myšlení**.

- Obsahuje studie veškeré podstatné informace? Zvolili výzkumníci vhodnou proceduru pro ověření stanovené hypotézy? Jsou závěry korektně interpretovány?
- Na úrovni jednotlivých studií i v rámci meta-analýz

Doporučení: Instituce

Změna Publish or Perish politiky:

- Počet publikací a impact faktor jako rozhodující proměnná při přidělování grantů, přijetí do zaměstnání či kariérním postupu

Alternativy:

- Oceňování a podpora replikační činnosti
- Vynaložení části prostředků v rámci výzkumu na replikaci

Doporučení: Obor

Přesun od efektů k teoriím.

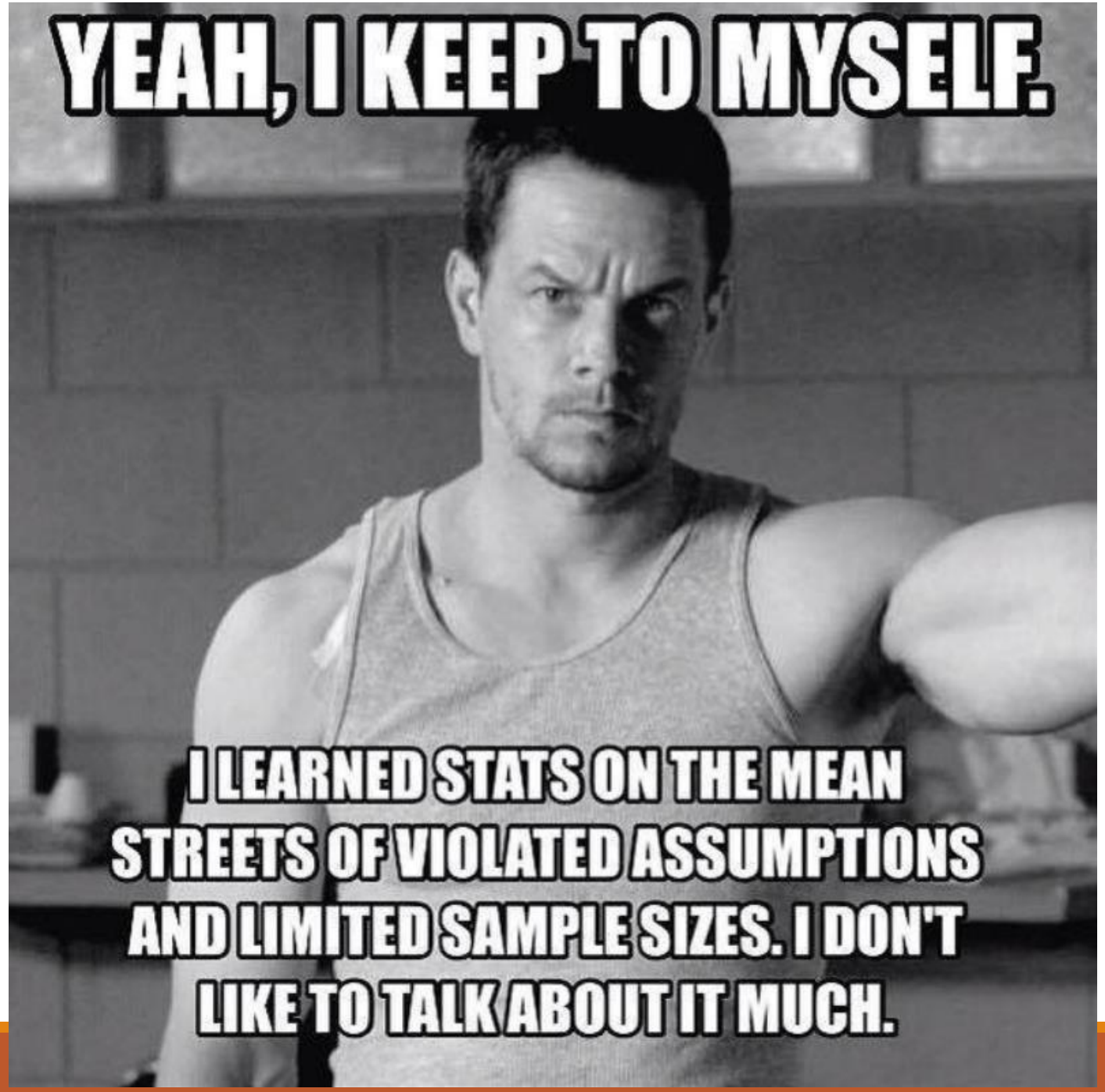
Přesun od dílčích studií k agregaci výzkumného poznání.

Větší důraz na způsob, kvalitu a podstatu měření.

- Vzhledem k měřenému atributu.

Větší míra standardizace výzkumných nástrojů.

Adekvátní statistické postupy.


YEAH, I KEEP TO MYSELF.
I LEARNED STATS ON THE MEAN STREETS OF VIOLATED ASSUMPTIONS AND LIMITED SAMPLE SIZES. I DON'T LIKE TO TALK ABOUT IT MUCH.