Reproducibility in Psychology

Brian Nosek
University of Virginia -- Center for Open Science
http://briannosek.com/ -- http://cos.io/

CORRESPONDENCE

Believe it or not: how much can we rely on published data on potential drug targets?

Florian Prinz, Thomas Schlange and Khusru Asadullah

Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. Glenn Begley and **Lee M. Ellis** propose how methods, publications and incentives must change if patients are to benefit.

Manufacturing beauty

Flexibility in Analysis

Selective Reporting

Presenting Exploratory as Confirmatory

Incentives for individual success are focused on getting it published, not getting it right

Nosek, Spies, & Motyl, 2012

Crowdsourcing Science

COLLECTION REPORTING

Standard Model

Vertical Integration

Resource Intensive

Exclusive

Produces lots of small science

Singular contribution model

Complementary Model

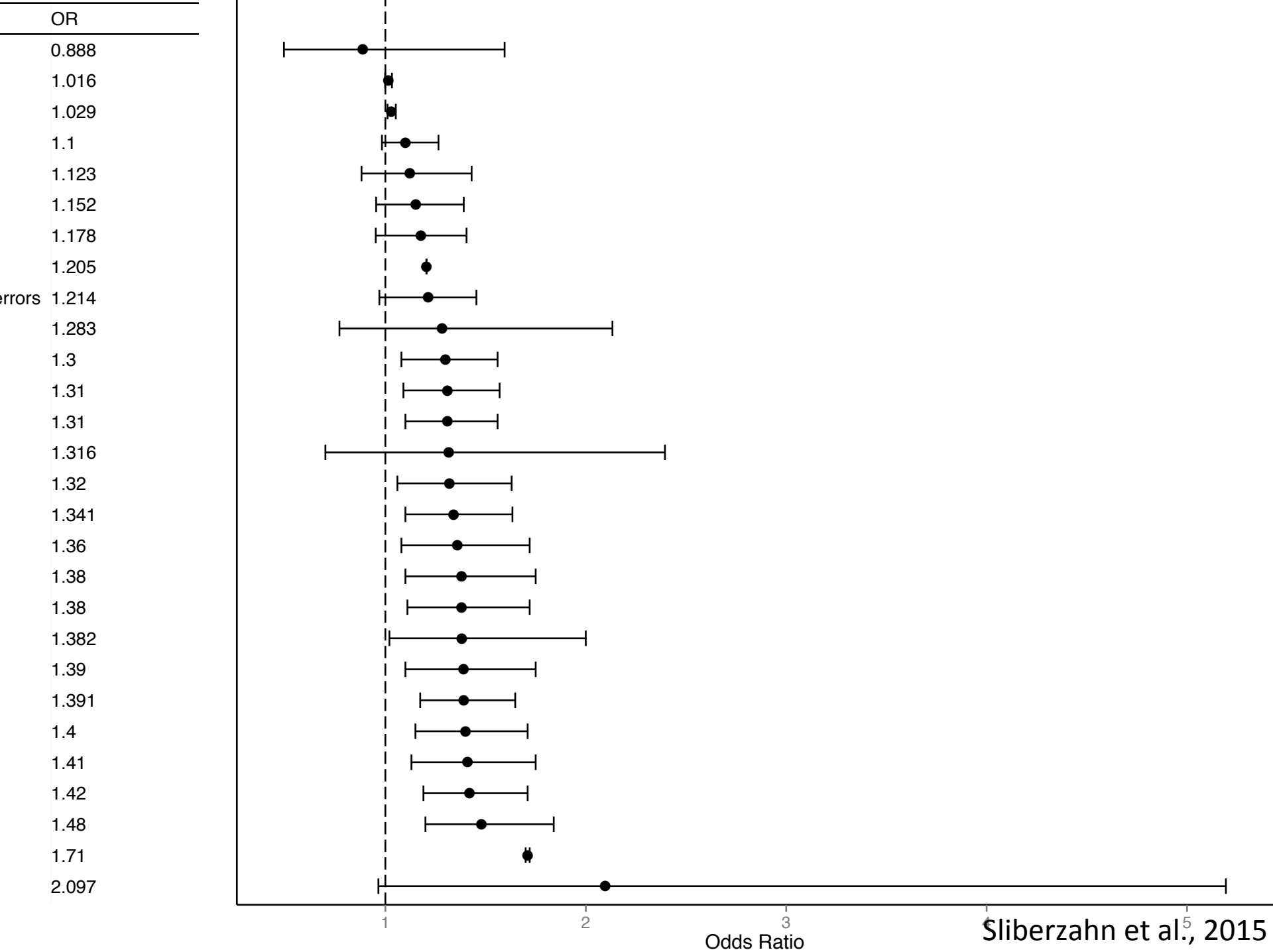
Horizontal Distribution

Resource Light

Inclusive

Enables big science

Diversifies contribution model



Sliberzahn et al., 2015

Reproducibility Project: Psychology

Sampling frame, 2008 issues of

PSCI

JEP:LMC

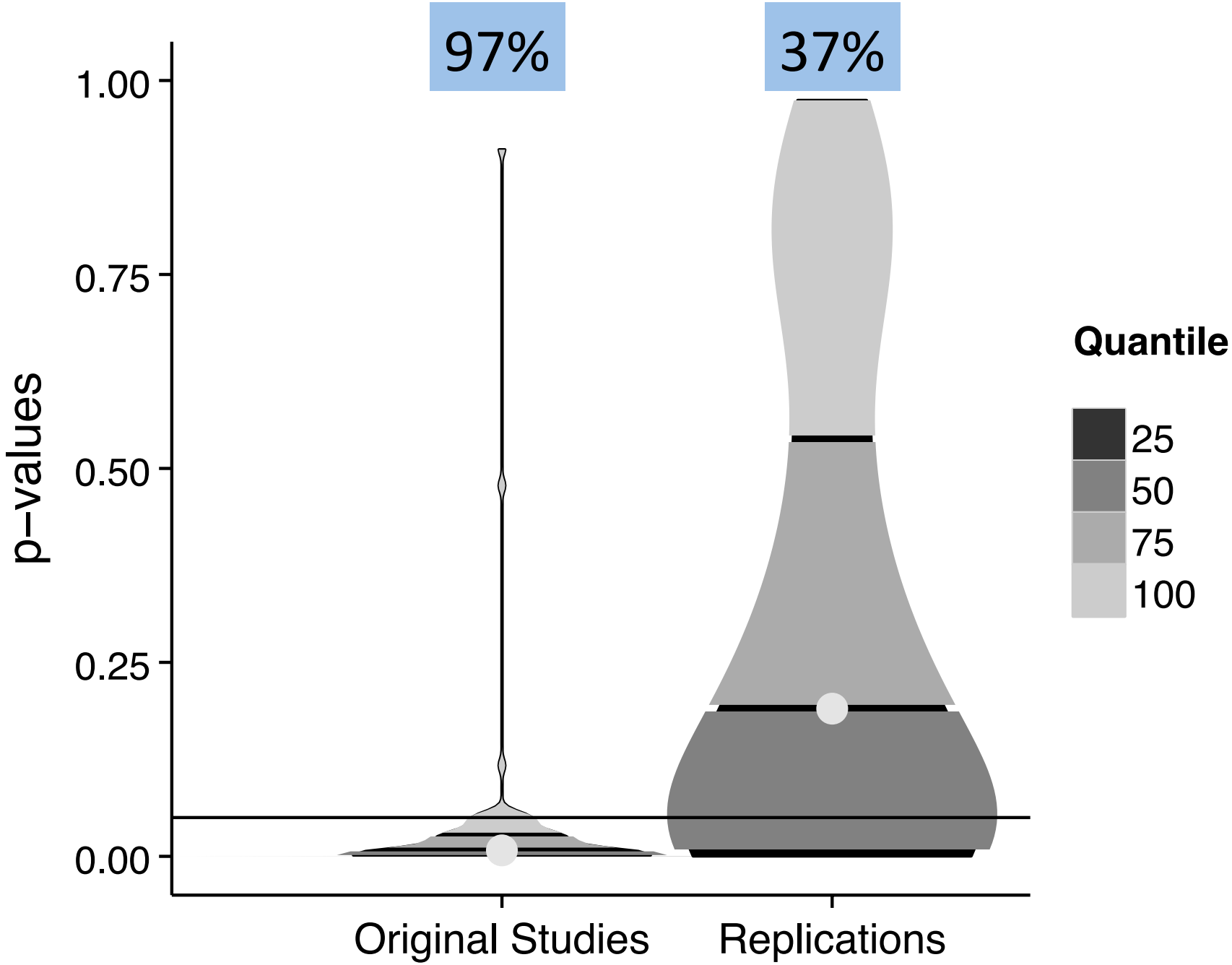
JPSP

Open Science Collaboration, 2015

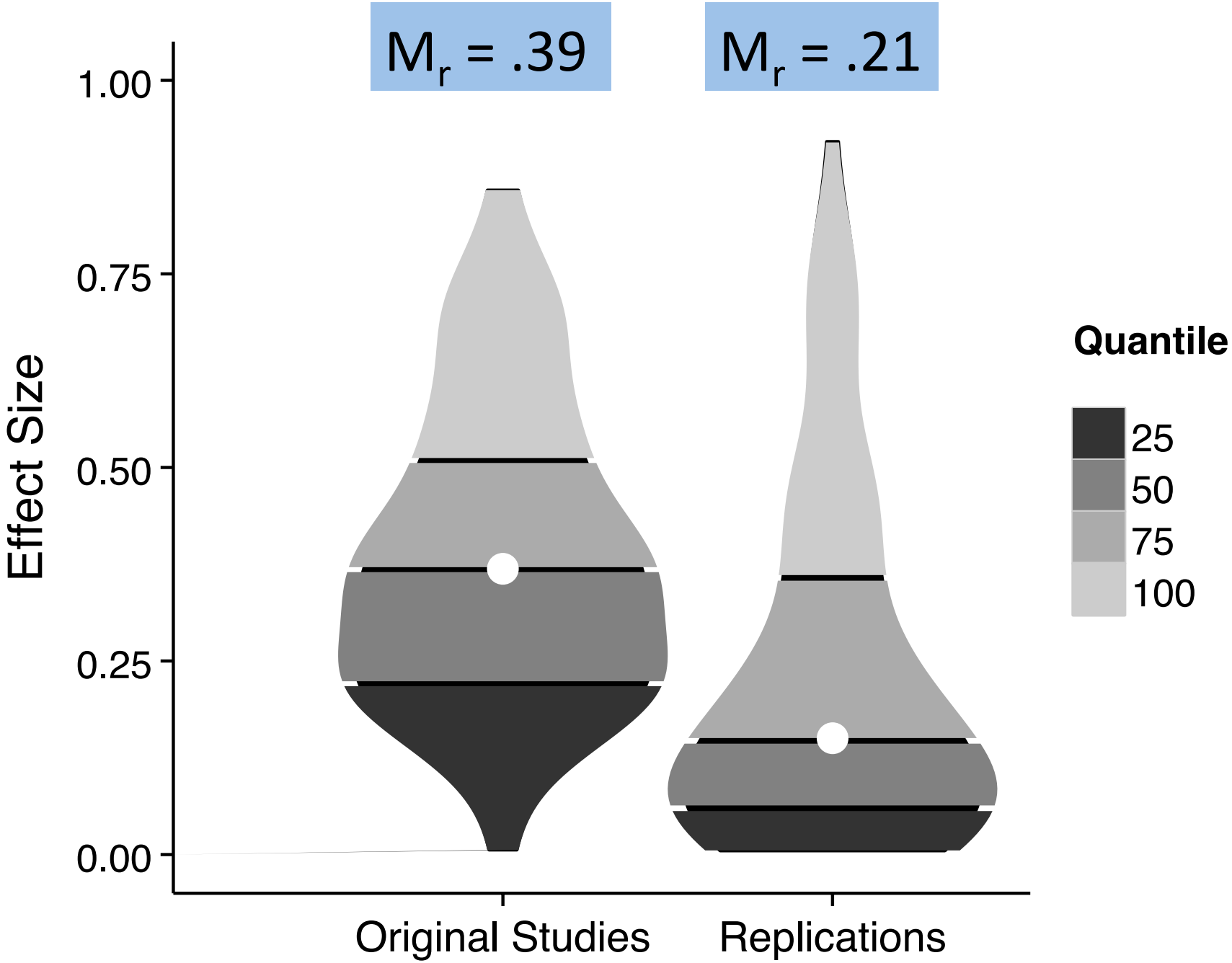
Reproducibility Project: Psychology

- 266 co-authors + 85 volunteers
- 100 replications
- Quality Control
  - High statistical power
  - Original materials
  - Standardized replication protocol
  - Vetting of replication process
  - Public reporting

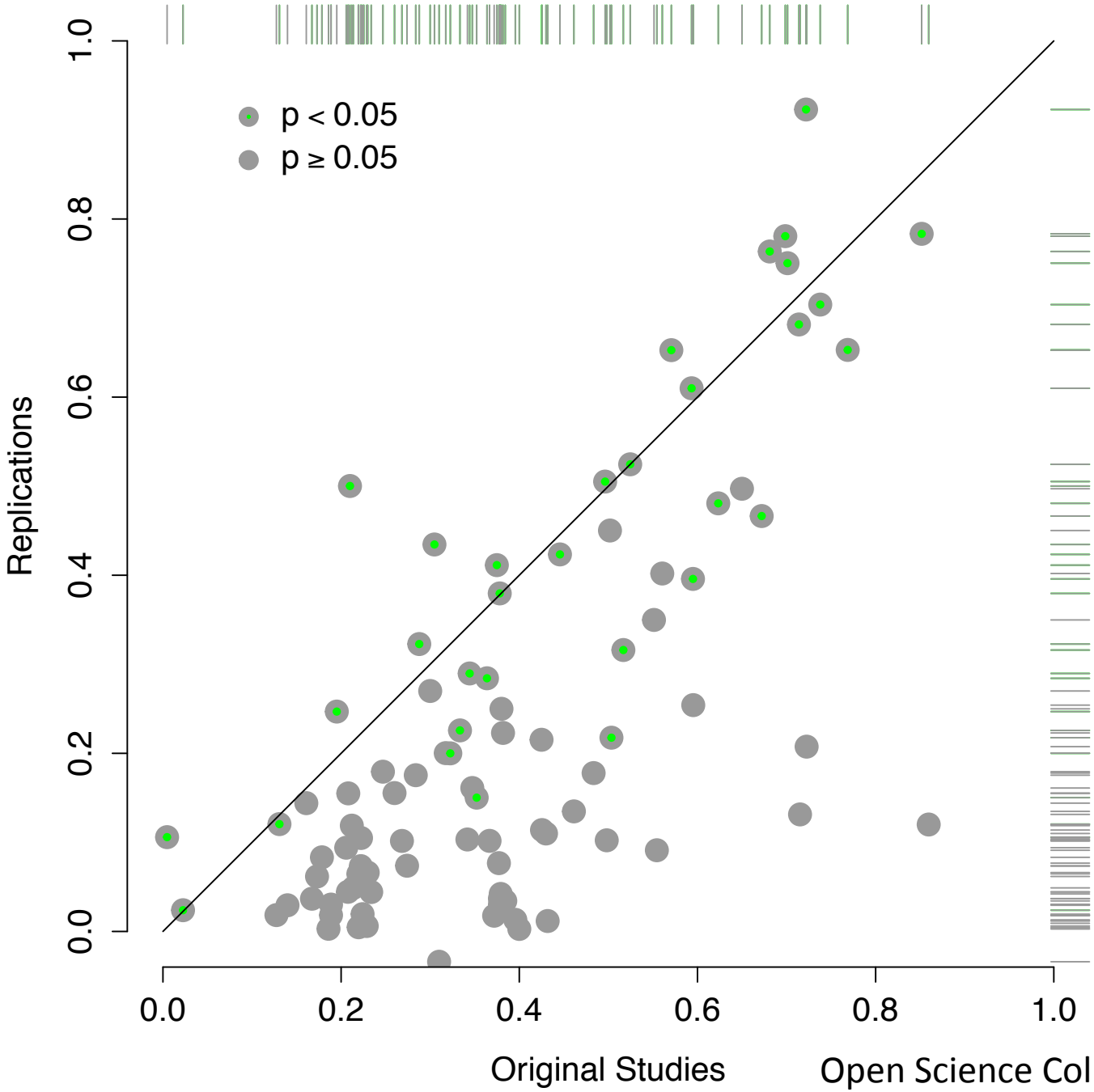
Open Science Collaboration, 2015



Open Science Collaboration, 2015



Open Science Collaboration, 2015



Open Science Collaboration, 2015

	$p < .05$	Original M (SD)	Replication M (SD)
Overall	35%	.39 (.19)	.21 (.24)
JPSP (n=29)	17%	.29 (.10)	.07 (.10)
JEP:LMC (n=26)	46%	.46 (.19)	.28 (.24)
PSCI – Soc (n=24)	29%	.37 (.21)	.22 (.28)
PSCI – Cog (n=9)	67%	.51 (.24)	.34 (.25)
PSCI – Oth (n=6)	17%	.52 (.13)	.29 (.34)

Open Science Collaboration, 2015

The Science Prediction Market Project

My Account (bwilson) Sign

HUUNU Engaging Wisd

TOPICS | MY ANSWERS | LEADERS | HOW TO PLAY | FAQ | REFER A FRIEND | CONTACT US | DASHBOARD

hypothesis_19
54.76 +4.76

In each of the below questions that you participate, you will bet on a binary outcome: whether or not the replication study finds a statistically significant effect in the same direction as the original study. By statistically significant we mean a p-value smaller than 0.05. By same direction we mean a coefficient that has the same sign as in the original study (i.e. positive or negative).

Networth	9,998
My Rank	35
Available Points	8,998

Hypotheses

Hypotheses

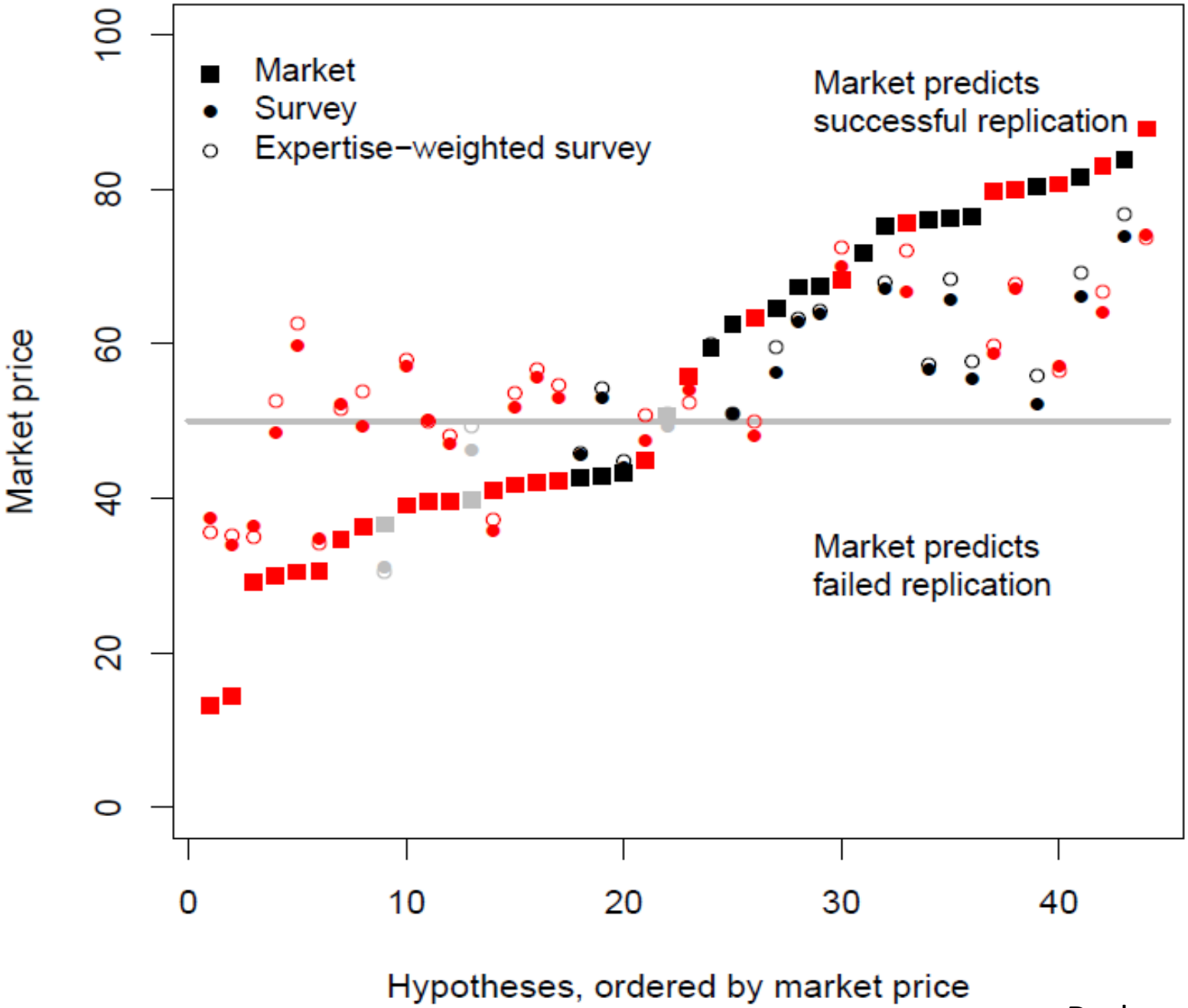
	SCORE	
Hypothesis 19, Vul et al., "Temporal selection is suppressed, delayed, and diffused during the attentional blink", Psychological Science (hypothesis_19)	54.76 +4.76	Adjust

TIPS AND TRICKS

A simple range is 500 to 2000 points for each question you wish to answer based on your level of confidence. Be sure to not run out of points.

ANOTHER TIP

Dreber et al., 2015

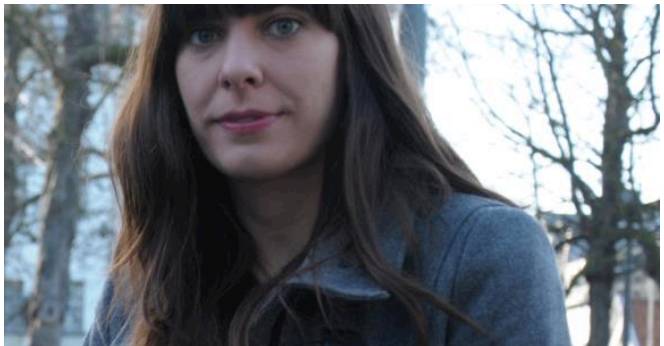


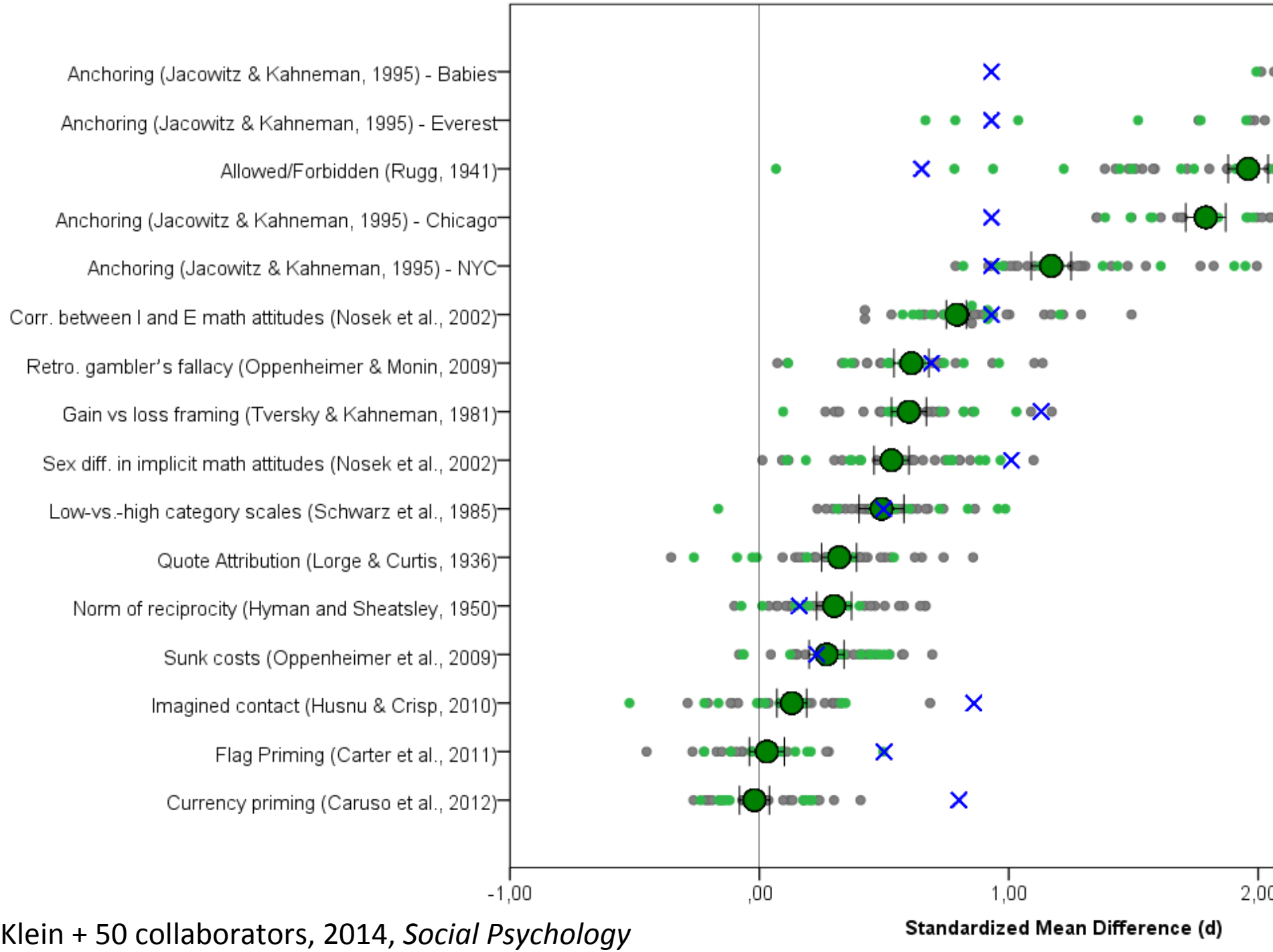
Dreber et al., 2015

Investigating Variation in Replicability

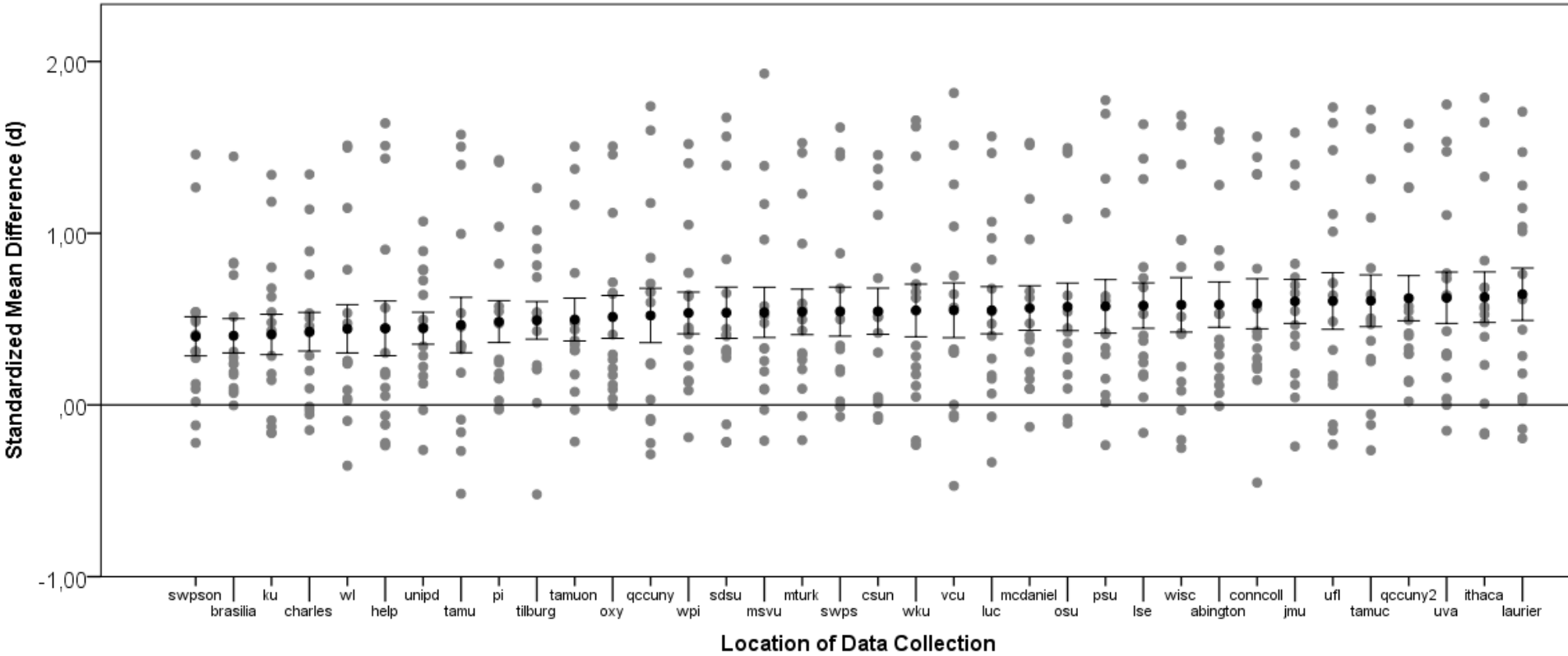
A "Many Labs" Replication Project

Richard A. Klein,[1] Kate A. Ratliff,[1] Michelangelo Vianello,[2] Reginald B. Adams Jr.,[3] Štěpán Bahník,[4] Michael J. Bernstein,[5] Konrad Bocian,[6] Mark J. Brandt,[7] Beach Brooks,[1] Claudia Chloe Brumbaugh,[8] Zeynep Cemalcilar,[9] Jesse Chandler,[10,36] Winnee Cheong,[11] William E. Davis,[12] Thierry Devos,[13] Matthew Eisner,[10] Natalia Frankowska,[6] David Furrow,[15] Elisa Maria Galliani,[2] Fred Hasselman,[16,37] Joshua A. Hicks,[12] James F. Hovermale,[17] S. Jane Hunt,[18] Jeffrey R. Huntsinger,[19] Hans IJzerman,[7] Melissa-Sue John,[20] Jennifer A. Joy-Gaba,[17] Heather Barry Kappes,[21] Lacy E. Krueger,[18] Jaime Kurtz,[22] Carmel A. Levitan,[23] Robyn K. Mallett,[19] Wendy L. Morris,[24] Anthony J. Nelson,[3] Jason A. Nier,[25] Grant Packard,[26] Ronaldo Pilati,[27] Abraham M. Rutchick,[28] Kathleen Schmidt,[29] Jeanine L. Skorinko,[20] Robert Smith,[14] Troy G. Steiner,[3] Justin Storbeck,[8] Lyn M. Van Swol,[30] Donna Thompson,[15] A. E. van 't Veer,[7] Leigh Ann Vaughn,[31] Marek Vranka,[32] Aaron L. Wichman,[33] Julie A. Woodzicka,[34] and Brian A. Nosek[29,35]





Klein + 50 collaborators, 2014, *Social Psychology*



Klein + 50 collaborators, 2014, *Social Psychology*

Many Labs 3: Participant Pool Edition

- Time of semester/quarter
- 10 Effects
- 20 pools; ~2,700 participants

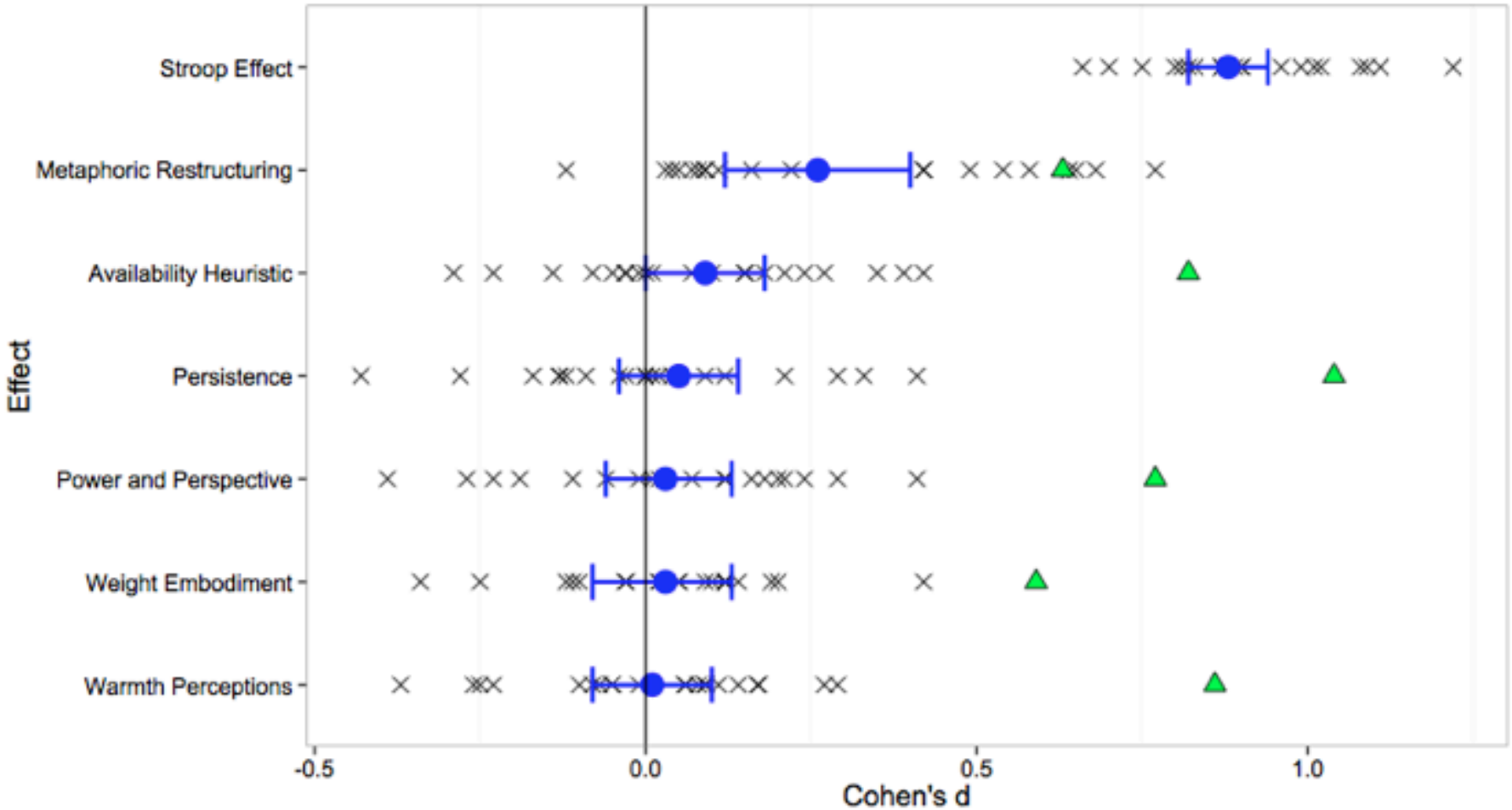
- Coordinated by:
  - Charlie Ebersole, UVA
  - Olivia Atherton, UC Davis
  - Aimee Belanger, Miami U - Ohio
  - Hayley Skulborstad, Miami U - Ohio

Ebersole et al., 2015

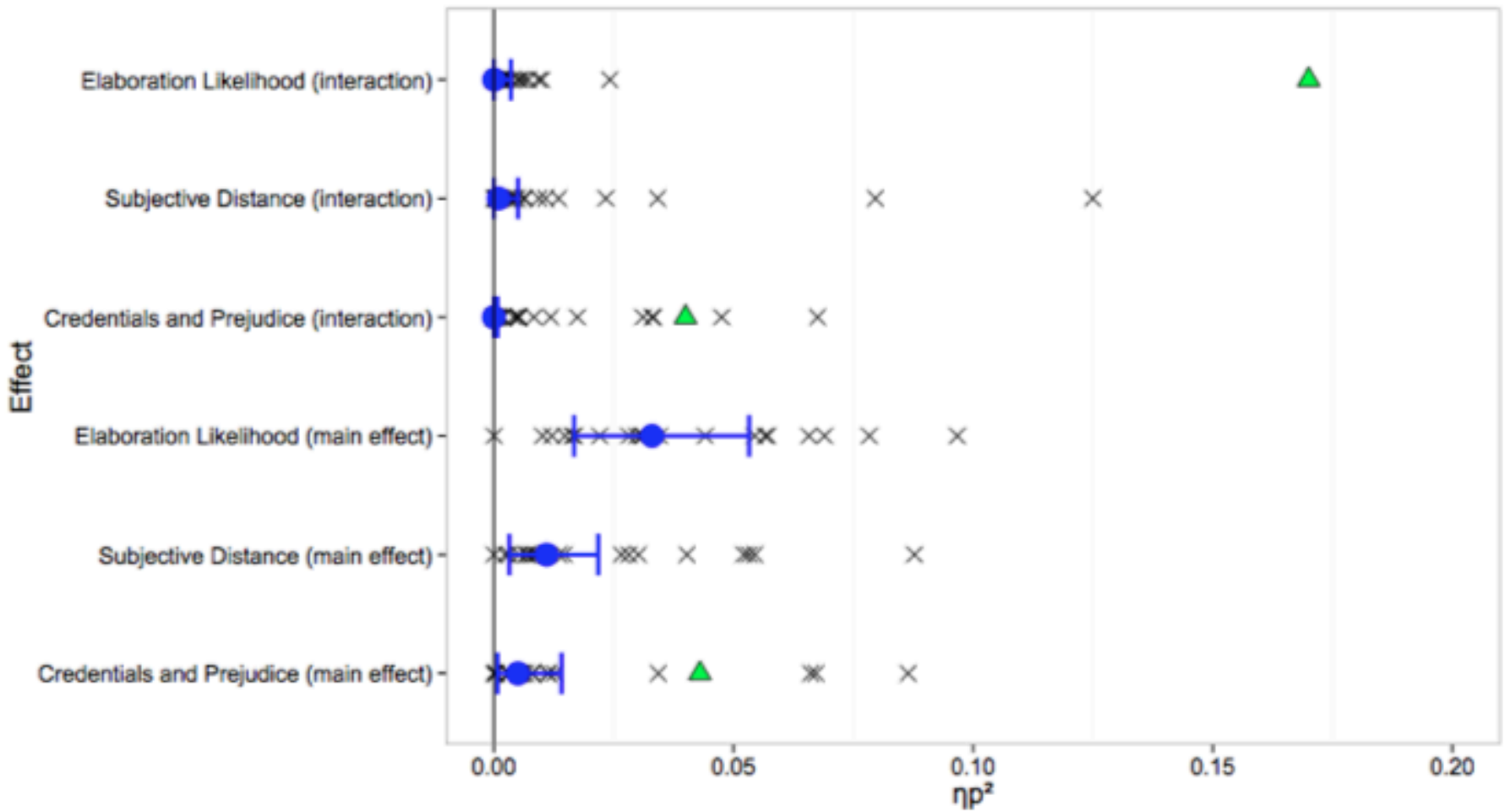
Effects in Many Labs 3

#	*Study Citation*	cites	study
1	Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. Psychological Science, 17(12), 1068-1074.	410	2a
2	Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. Journal of personality and social psychology, 81(1), 33.	408	1
3	Ross, M., & Wilson, A. E. (2002). It feels like yesterday: self-esteem, valence of personal past experiences, and judgments of subjective distance. Journal of personality and social psychology, 82(5), 792.	182	2
4	Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive psychology, 5(2), 207-232.	5659	3
5	Stroop Task (originally from Stroop, 1935, adapted for this project from Inzlicht, M., & Gutsell, J. N. (2007). Running on empty neural signals for self-control failure. Psychological Science, 18(11), 933-937.)	147	1
6	Szymkow, A., Chandler, J., IJzerman, H., Parzuchowski, M., & Wojciszke, B. (2013). Warmer hearts, warmer rooms. Social Psychology, 44(2), 167-176.	9	1
7	Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. Journal of personality and social psychology, 45(4), 805.	583	1
8	Conceptual replication of De Fruyt, F., Van De Wiele, L., & Van Heeringen, C. (2000). Cloninger's psychobiological model of temperament and character and the five-factor model of personality. Personality and individual differences, 29(3), 441-452. Uses unsolvable anagrams task from Aspinwall, L. G., & Richter, L. (1999). Optimism and self-mastery predict more rapid disengagement from unsolvable tasks in the presence of alternatives. Motivation and Emotion, 23(3), 221-245.	145	1
9	Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. Cognition, 75(1), 1-28.	801	1
10	Jostmann, N. B., Lakens, D., & Schubert, T. W. (2009). Weight as an embodiment of importance. Psychological science, 20(9), 1169-1174.	144	2

Ebersole et al., 2015



Ebersole et al., 2015



Ebersole et al., 2015

Many Labs 2

- 28 effects (split into two slates)
- 104 samples; ~15,000 participants
- 36 nations/regions (Serbia, Poland, New Zealand, Netherlands, Canada, Jamaica, UK, USA, Australia, South Africa, Colombia, Turkey, Costa Rica, Spain, Chile, Brazil, India, United Arab Emirates, Tanzania, Malaysia, Italy, Nigeria, Germany, Belgium, France, Czech Republic, Hong Kong, China, Uruguay, Sweden, Mexico, Portugal, Japan, Hungary, Switzerland, Taiwan)

Klein & 100+ collaborators, *Perspectives on Psychological Science*

ML2 Samples

- ~104 independent samples
- 

ML2 Effects (Slate 1)

1. Huang (2014). Living in the north is not necessarily favorable:… (52%, D = .34)
2. Kay (2013). A functional basis for structure-seeking: Exposure… (36%, D = .19)
3. Alter (2007). Overcoming intuition: metacognitive difficulty…(43%, D = .29)
4. Graham (2009). Liberals and conservatives rely on different…(76%, D = .45)
5. Rottenstreich (2001). Money, kisses, and electric shocks: On…(58%, D = .44)
6. Bauer (2012). Cuing Consumerism Situational Materialism…(54%, D = .45)
7. Miyamoto (2002). Cultural variation in correspondence bias:…(73%, D = .97)
8. Inbar (2009). Disgust sensitivity predicts disapproval of gays. (56%, D = .34)
9. Critcher (2008). Incidental environmental anchors. (44%, D = .18)
10. Van Lange (1997). Development of prosocial, individualistic,…(N/A)
11. Hauser (2007). A Dissociation Between Moral Judgments…(85%, D = .1.53)
12. Anderson (2012). The local-ladder effect social status…(58%, D = .34)
13. Ross (1977). The “false consensus effect”: An egocentric bias…(78%, D = .6)

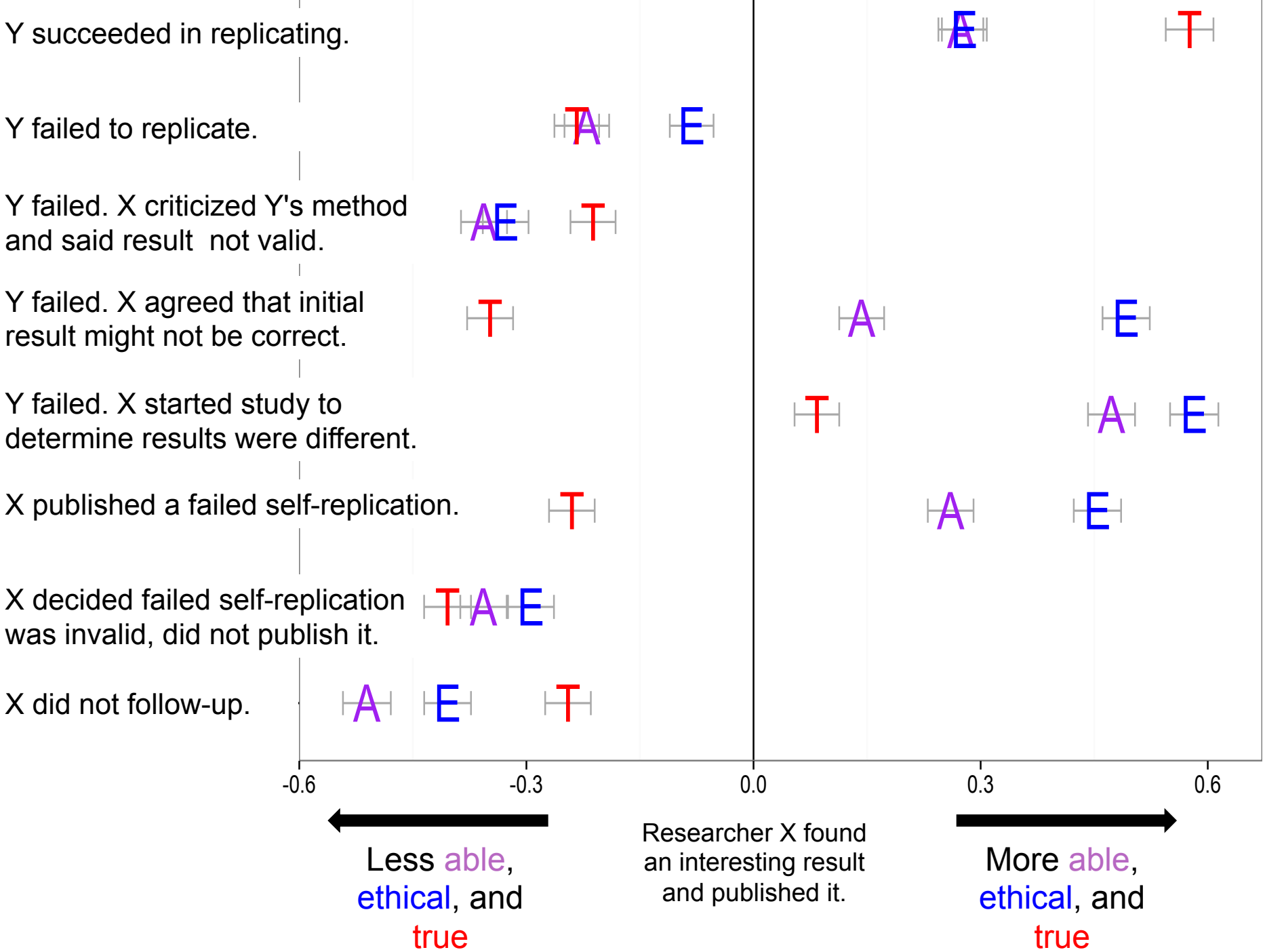
ML2 Effects (Slate 2)

15. Giessner (2007). High in the hierarchy: How vertical location…(48%, D = .26)
16. Tversky (1981). The framing of decisions and the psychology…(86%, D = .7)
18. Risen (2008). Why people are reluctant to tempt fate. (52%, D = .22)
19. Savani (2010). What counts as a choice? US Americans are…(41%, D = .19)
20. Norenzayan (2002). Cultural preferences for formal versus… (63%, D = .83)
21. Hsee (1998). Less is better: When low-value options are…(73%, D = .52)
22. Gray (2009). Moral typecasting: divergent perceptions…(87%, D = .68)
23. Zhong (2006). Washing away your sins: Threatened morality…(38%, D = .39)
24. Schwarz (1991). Assimilation and contrast effects in part…(55%, D = .18)
25. Shafir (1993). Choosing versus rejecting: Why some options…(71%, D = .34)
26. Zaval (2014). How warm days increase belief in global warming.(50%, D = N/A)
27. Knobe (2003). Intentional Action and Side Effects in Ordinary… (76%, D = .91)
28. Tversky (1978). Studies of similarity. (65%, D = .35)

Reputation and Replication

Surveyed 4,939 U.S. adults

Ebersole, Axt, & Nosek, 2015


Y succeeded in replicating.
Y failed to replicate.
Y failed. X criticized Y's method and said result not valid.
Y failed. X agreed that initial result might not be correct.
Y failed. X started study to determine results were different.
X published a failed self-replication.
X decided failed self-replication was invalid, did not publish it.
X did not follow-up.
-0.6
-0.3
0.0
0.3
0.6
Less able, ethical, and true
Researcher X found an interesting result and published it.
More able, ethical, and true