



REPLY TO MISLAVSKY ET AL.:

# Sometimes people really are averse to experiments

Michelle N. Meyer<sup>a,1</sup>, Patrick R. Heck<sup>a,b</sup>, Geoffrey S. Holtzman<sup>c</sup>, Stephen M. Anderson<sup>d</sup>, William Cai<sup>e</sup>,  
Duncan J. Watts<sup>f,g,h</sup>, and Christopher F. Chabris<sup>b</sup>

In response to our article (1), Mislavsky et al. (2) claim that “experiment aversion” does not exist because they found no evidence of it in their own research on low-stakes corporate experiments (3) and because our studies used between- rather than within-subjects designs.

First, as we noted, we do not expect (and did not ourselves find) an A/B effect in every scenario, and we called for research on how the effect might vary across contexts.

Second, we deliberately used a between-subjects design to maximize external validity: Universal implementation of policies usually occurs without mention of foregone alternatives, whereas A/B tests inherently acknowledge those alternatives. The belief that experiments deprive people of potentially beneficial interventions, but universally implemented policies do not, is not a “confound” to be avoided (2) but, rather, a key mechanism underlying the A/B effect. It is a consequential failure to recognize that “the world outside the experiment is often just the A condition of an A/B test that was never conducted” (1).

Third, nevertheless, to determine how much of the A/B effect is caused by this particular mechanism, we conducted studies 4 through 6 (not mentioned at all by Mislavsky et al.). In studies 5 and 6, all respondents were told that some doctors in a walk-in clinic prescribe an unnamed “drug A” to all their hypertensive patients, while others prescribe “drug B.” This is exactly what Mislavsky et al. call for—a vignette where “in the policy scenarios the [agent] could choose 1 of 2 treatments for everyone”—yet still we found a large

A/B effect among both laypersons and healthcare professionals.

Fourth, we also anticipated and addressed Mislavsky et al.’s concern about internal validity of a between-subjects design—that the A/B effect could simply be the aggregate of people objecting for different reasons to the A and B policies. Our studies 1 and 2 already showed that objection rates to experiments can be higher than the sum of objections to either arm. Moreover, in studies 4 through 6, there is no rational reason for respondents to prefer one anonymous drug over another, so, unlike peanut allergies and lactose intolerance, the total objections by people who read about an A/B test cannot logically be the sum of people who object to receiving “drug A” and people who object to “drug B.” Yet still we see a large A/B effect.

Finally, although the best test of the A/B effect is our between-subjects design, we note that in their simulated within-subjects design Mislavsky et al. chose the 7 scenarios we studied with the smallest A/B effects (including one nonsignificant effect). When we replace just 3 of these 7 that had the smallest effects with 3 of our larger-effect studies and run their own code the aggregate evidence supports—by their own criterion—“experiment aversion” (Fig. 1).

In sum, Mislavsky et al.’s claim that people never exhibit experiment aversion is unwarranted. The A/B effect—people’s persistent but nonuniversal tendency to object to experiments comparing policies they do not object to—cannot be explained away as an “artifact” or “confound.”

<sup>a</sup>Center for Translational Bioethics and Health Care Policy, Geisinger Health System, Danville, PA 17821; <sup>b</sup>Autism and Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA 17837; <sup>c</sup>Department of Psychology, Franklin & Marshall College, Lancaster, PA 17603; <sup>d</sup>Department of Psychology, The Pennsylvania State University, University Park, PA 16802; <sup>e</sup>Department of Management Science and Engineering, Stanford University, Stanford, CA 94305; <sup>f</sup>Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104; <sup>g</sup>Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA 19104; and <sup>h</sup>Department of Operations, Information and Decisions, Wharton School, University of Pennsylvania, Philadelphia, PA 19104

Author contributions: M.N.M., P.R.H., and C.F.C. designed research; P.R.H. performed research and analyzed data; M.N.M., P.R.H., D.J.W., and C.F.C. wrote the paper; and G.S.H., S.M.A., and W.C. edited versions of the paper.

The authors declare no competing interest.

Published under the [PNAS license](#).

<sup>1</sup>To whom correspondence may be addressed. Email: michellenmeyer@gmail.com.

First published November 12, 2019.

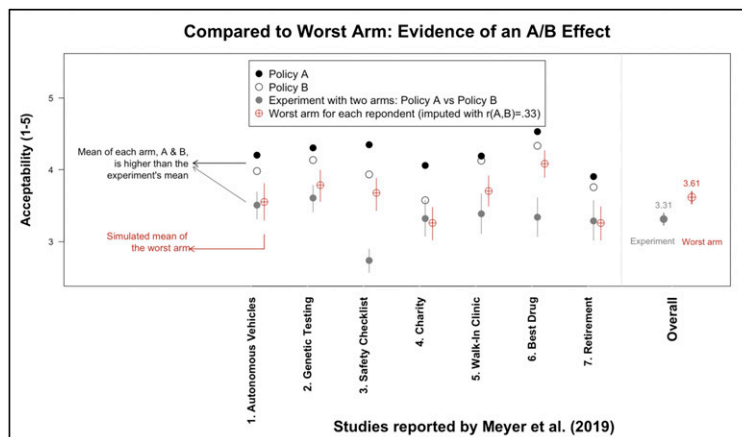


Fig. 1. Simulation run using Mislavsky et al.'s R code. We modified their simulation—which they used to claim that there was no evidence for experiment aversion—by replacing the 3 weakest studies (“Basic Income,” “Health Worker Recruitment,” and “Teacher Well-being”) with 3 studies that found a large A/B effect (“Safety Checklist,” “Walk-In Clinic,” and “Best Drug”). We used the same input correlation ( $r = 0.33$ ) that Mislavsky et al. obtained from 99 Amazon Mechanical Turk participants. The overall comparison (Far Right) now produces evidence of an A/B effect, where the mean of the worst arm ( $M = 3.61$ ) is greater than (and outside the 95% CI limits of) the mean of the experiment ( $M = 3.31$ ).

- 1 M. N. Meyer et al., Objecting to experiments that compare two unobjectionable policies or treatments. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10723–10728 (2019).
- 2 R. Mislavsky, B. J. Dietvorst, U. Simonsohn, The minimum mean paradox: A mechanical explanation for apparent experiment aversion. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23883–23884 (2019).
- 3 R. Mislavsky, B. Dietvorst, U. Simonsohn, Critical condition: People don't dislike a corporate experiment more than they dislike its worst condition. *Mark. Sci.*, 10.1287/mksc.2019.1166.