Reply to Bas et al.: The difference between a genuine tendency and a context-specific response

Nina Mazar[a,1] [ID], Christian T. Elbaek[b] [ID], and Panagiotis Mitkidis[b] [ID]

The scope of our work (1) was to examine the evidence for the claim (2–4) that there exists a common pattern (i.e., genuine tendency or predictable bias) wherein individuals evaluate experiments as relatively less appropriate than the universal implementation of its unobjectionable policies, termed relative experiment aversion (EA). For this claim to hold, two requirements must be met: An experiment must be rated relatively less favorably than its treatment arms [i.e. Mean (A/B test) < Mean (A+B pooled) or Mean (A/B test) < Min (A, B), refs. 1–5], and this observation should be generalizable and not an artifact of the specific context such as the choice of experimental design, method, or analytical tool (5–7).

We ran one direct and six conceptual replication studies, utilizing previous works' largest effect-size scenarios (2, 4) resulting in 18 A/B test observations (see Tables 1 and 2 in ref. 1). Focusing on previous work's (2–4) dependent variable—appropriateness of an agent's decision—we observed only one instance where respondents rated the experiment less favorably than its treatment arms: when we ran a direct replication (see ref. 1, Study 5).

Bas et al. (8) argue that this one successful direct replication constitutes a "robust empirical finding" that EA exists. However, we argue it constitutes a robust empirical finding that a very specific set of stimuli extensively used in previous work (2, 4) elicits relatively less favorable attitudes toward experiments. Our and others' work (1, 9–11) provide evidence that reasonable (6, 7) minor changes to these stimuli—as part of a triangulation-effort to assess the generalizability and robustness of "a genuine aversion to randomized evaluation"-claim (ref. 4, p. 18948)—not only weaken but make completely disappear (or even inverse) these differences (see Table 1 here with two new conceptual replication studies; OSF: https://osf.io/whz3b/).

Additionally, Bas et al. (8) claim that our deviations from the original studies were not systematically and orthogonally manipulated. This is factually incorrect, as is evident from Tables 1 and 2 in ref. 1. For example, we orthogonally manipulated DV order (Study 4) and tense (Studies 1a, 1b). Moreover, Bas et al. (8) claim that in our scenarios unlike in the original ones respondents had to evaluate several hospitals. This, again, is factually incorrect. Same as in the original studies (2, 4), respondents in our between-subject evaluations evaluated only one hospital, and respondents in the within-subject evaluations evaluated three hospitals. Finally, Bas et al. (8) erroneously speculate that "lengthy descriptions and additional DVs may have reduced participants' attentiveness in our studies, resulting in uninformative responses." Yet, our scenarios were similar in length to the original studies (see SI Appendix, Tables S1–S3, ref. 1 and Table 1 here) and we found no evidence for EA even when presenting the main DV first (i.e., before additional DVs; see the four relevant A/B test observations in Table 1).

In sum, we do not find evidence for both necessary requirements to claim that there exists a genuine aversion to experiments. However, further inquiries would be useful to policymakers to better understand the specific contexts that reliably foster favorable versus unfavorable relative attitudes towards experiments.

[1]To whom correspondence may be addressed. Email: nina@ninamazar.com.

Published December 4, 2023.

Table 1. Comparison of scenario-wordings in original (2) and our (new and ref. 1) between-subject "Best Drug: Walk-In Clinic" studies

	Meyer et al. (2), Study 5a and direct replication in Mazar et al. (1), Study 5	Our conceptual replications of Meyer et al. (2), Study 5a		
		NEW 2023, Study 4-2	NEW 2023, Study 4-1	Mazar et al. (1), Study 4[*]
Intro	Several drugs have been approved by the US Food and Drug Administration as safe and effective for treating high blood pressure. Doctor Jones works in a multi-doctor walk-in clinic where patients see whichever doctor is available. Some doctors in the clinic prescribe drug A for high blood pressure, while others prescribe drug B. Both drugs are affordable and patients can tolerate their side effects.	Imagine the following. Several drugs have been approved by the US Food and Drug Administration as safe and effective for treating high blood pressure. Clinic A [B][AB] is a multidoctor walk-in clinic where patients see whichever doctor is available. So far, some doctors in the clinic have prescribed drug A for high blood pressure, while others have prescribed drug B. Both drugs are affordable and patients can tolerate their side effects.	Imagine the following. Several drugs have been approved by the US Food and Drug Administration as safe and effective for treating high blood pressure. These drugs can save lives but not everyone responds to the treatment with them. Clinic A [B][AB] is a multidoctor walk-in clinic where patients see whichever doctor is available. So far, some doctors in the clinic have prescribed drug A for high blood pressure, while others have prescribed drug B. Both drugs are affordable and patients can tolerate their side effects.	Imagine you need medical treatment for high blood pressure. Two drugs have been approved by the US Food and Drug Administration as safe and effective for treating high blood pressure. These drugs can save lives but not everyone responds to the treatment with them. Clinic A [B][AB] is a multidoctor walk-in clinic where patients see whichever doctor is available. So far, some doctors in the clinic have prescribed drug A for high blood pressure, while others have prescribed drug B. Both drugs are affordable and patients can tolerate their side effects.
Cond A[B]	Doctor Jones wants to provide good treatment to his patients, so he decides that his patients who need high blood pressure medication will be prescribed drug A [B].	The director of the clinic wants to provide good treatment to their patients, so s/he decides that from now on all new patients who need high blood pressure medication will only be prescribed drug A [B].	The director of the clinic wants to provide good treatment to their patients, so s/he decides that from now on all new patients who need high blood pressure medication will only be prescribed drug A [B].	The director of the clinic wants to provide good treatment to their patients and streamline the care, so s/he randomly decides that from now on all new patients who need high blood pressure medication will only be prescribed drug A [B].
Cond A/B test	Doctor Jones thinks of two different ways to provide good treatment to his patients, so he decides to run an experiment by randomly assigning his patients who need high blood pressure medication to one of two test conditions. Half of patients will be prescribed drug A, and the other half will be prescribed drug B. After a year, he will only prescribe to new patients whichever drug has had the best outcomes for his patients.	The director of the clinic wants to provide good treatment to their patients, so s/he decides to run an experiment by randomly assigning their patients who need high blood pressure medication to one of two test conditions. Half of the patients will be prescribed drug A, and the other half will be prescribed drug B. After the experimental phase, the director will assess which drug, A or B, has had the best outcomes for their patients, and, from then on, all new patients who need high blood pressure medication will only be prescribed that drug.	The director of the clinic wants to provide good treatment to their patients, so s/he decides to run an experiment by randomly assigning their patients who need high blood pressure medication to one of two test conditions. Half of the patients will be prescribed drug A, and the other half will be prescribed drug B. After the experimental phase, the director will assess which drug, A or B, has had the best outcomes for their patients, and, from then on, all new patients who need high blood pressure medication will only be prescribed that drug.	The director of the clinic wants to provide good treatment to their patients and streamline the care, so s/he decides to run an experiment by randomly assigning their patients who need high blood pressure medication to one of two test conditions. Half of the patients will be prescribed drug A, and the other half will be prescribed drug B. After the experimental phase, the director will assess which drug, A or B, has had the best outcomes for their patients, and, from then on, all new patients who need high blood pressure medication will only be prescribed that drug.
First Q	How appropriate is Doctor Jones' decision? [very inappropriate–very appropriate]	What do you think of the director's decision? [very inappropriate–very appropriate]	What do you think of the director's decision? [very inappropriate–very appropriate]	What do you think of the director's decision? [very inappropriate–very appropriate]
Sample	MTurk US (N = 303 || replication N = 792)	Prolific US, Representative (N = 451)	Prolific US, Representative (N = 450)	Prolific US (N = 449)
A/B test effect	Mean (A/B test) < Mean (A+B) $P < 0.001$ || replication $P < 0.001$ $d = -0.64$ || replication $d = -0.84$	Mean (A/B test) ~ Mean (A+B) $P = 0.436$ $d = 0.08$	Mean (A/B test) ~ Mean (A+B) $P = 0.243$ $d = 0.12$	Mean (A/B test) > Mean (A+B) $P < 0.001$ $d = 0.40$

*Condition "A/B present tense (not SR, not different DV order)."
Note: All scenarios are worded in present tense and have the main DV question first (i.e., the question about the appropriateness of the agent's decision).

1. N. M. Mazar, C. T. Elbaek, P. Mitkidis, Experiment aversion does not appear to generalize. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2217551120 (2023).
2. M. N. Meyer *et al.*, Objecting to experiments that compared two unobjectionable policies or treatmes. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10723–10728 (2019).
3. M. N. Meyer *et al.*, Reply to Mislavsky et al.: Sometimes people really are averse to experiments. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23885–23886 (2019).
4. P. R. Heck, C. F. Chabris, D. J. Watts, M. N. Meyer, Objecting to experiments even while approving of the policies or treatments they compare. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18948–18950 (2020).
5. 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).
6. M. R. Munafò, G. D. Smith, Repeating experiments is not enough. *Nature* **553**, 399–401 (2018).
7. T. Yarkoni, The generalizability crisis. *Behav. Brain Sci.* **45**, E1 (2022).
8. B. Bas, J. Vosgerau, R. Ciulli, No evidence that experiment aversion is not a robust empirical phenomenon. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2317514120 (2023).
9. R. Mislavsky, B. Dietvorst, U. Simonsohn, Critical condition: People don't dislike a corporate experiment more than they dislike its worst condition. *Mark. Sci.* **39**, 1092–1104 (2019).
10. R. Dur, A. Non, P. Prottung, B. Ricci, "Who's Afraid of Policy Experiments?" (Discussion Paper No. 23–027/V, Tinbergen Institute, 2023), https://ideas.repec.org/p/tin/wpaper/20230027.html.
11. M. Fischer, E. Grewenig, P. Lergetporer, K. Werner, H. Zeidler, "The E-word – on the public acceptance of experiments" (Discussion Paper No. 16511, IZA, 2023), https://docs.iza.org/dp16511.pdf.