Reply to Engzell: Maybe in plain sight but out of focus

Nate Breznau[a,1] (ID) and Eike Mark Rinke[b] (ID)

We thank Engzell (1) for making points about our study (2) that are both important independent of our findings and valuable as additional perspectives on the evidence reported. Overall, Engzell reminds us about careful hypothesis specification, measurement, and causal inference. These are things coming to the fore in recent methodological discussions (3–5), and with which we wholeheartedly agree and see as key takeaways from our main findings.

As a reminder, the main point of the research reported in our article was neither to identify the best possible test for a given hypothesis nor to conduct a replication of previous findings. It also was not to identify if any particular teams successfully used the data to test causality—although being nonexperimental in nature certainly does not exclude data from the potential to test causal claims (6, 7). Instead, our study held potential sources of variance in research design constant at levels that emulate a situation commonly encountered in the social sciences: a given hypothesis and data suspected to be appropriate for testing it. Then, when different teams were free to take their own approaches to this hypothesis, we observed the variation in findings and analytical decisions that followed.

This resulting variation was broad and remained after adjusting for all observed model components including teams' potential usage of up to eight different measures drawn from the data as input variables predicting variance in their numerical outcomes and subjective conclusions. In this process, we identified that between-team variance was 11.2% of total variance, which is not only a nontrivial proportion (8) but also the basis for broad disagreement in the subjective conclusions of the teams. As we report in the article, "13.5% (12 of 89) of the team conclusions were that the hypothesis was not testable given these data, 60.7% (54 of 89) were that the hypothesis should be rejected, and 28.5% (23 of 89) were that the hypothesis was supported." Such between-team variation is important because a single subjective conclusion is usually a headline takeaway of a given scientific publication. Therefore, our study demonstrates that any single study's outcomes might differ and should be interpreted as a draw from an unobserved distribution whose shape and weights are mostly unknown or unappreciated, often only inferred from simulations.

Moving forward, we believe the scientific community has much to learn in the interpretation of interresearcher and intermodel variance, although we recognize that some important first steps have been taken (9, 10). Engzell's letter coupled with our findings provide a reminder that it will be an important task for future many-analyst studies to examine more systematically the degree to which greater specificity in theory and hypotheses, identification of estimands, attention to measurement, and use of causal inference might reduce analytical flexibility and variation in scientific results. If the reduction is massive, as Engzell presumes, our concerns about the "hidden universe of uncertainty" may have solutions hiding in plain sight. At this time, we lack the evidence required to know whether this is clearly the case.

1. P. Engzell, A universe of uncertainty hiding in plain sight. *Proc. Natl. Acad. Sci. U.S.A.* (2022).
2. N. Breznau *et al.*, Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2203150119 (2022).
3. K. Auspurg, J. Brüderl, Has the credibility of the social sciences been credibly destroyed? Reanalyzing the "many analysts, one data set" project. *Socius* **7**, 1–14 (2021).
4. A. M. Scheel, Why most psychological research findings are not even wrong. *Infant Child Dev.* **31**, e2295 (2022).
5. I. Lundberg, R. Johnson, B. M. Stewart, What is your estimand? Defining the target quantity connects statistical evidence to theory. *Am. Sociol. Rev.* **86**, 532–565 (2021).
6. H. M. Blalock, *Causal Inferences in Nonexperimental Research* (The University of North Carolina Press, 1964).
7. J. Pearl, The foundations of causal inference. *Soc. Methodol.* **40**, 75–149 (2010).
8. J. J. Hox, *Multilevel Analysis: Techniques and Applications* (Routledge, ed. 2, 2010).
9. C. Young, K. Holsteen, Model uncertainty and robustness: A computational framework for multimodel analysis. *Sociol. Methods Res.* **46**, 3–40 (2017).
10. C. Young, Model uncertainty and the crisis in science. *Socius* **4**, 1–7 (2018).