

Commentary to Almaatouq, Griffiths, Suchow, Whiting, Evans, and Watts

Experiment commensurability does not necessitate research consolidation

Milena Tsvetkova

London School of Economics and Political Science
Houghton Street
London WC2A 1AE
United Kingdom

Abstract

Integrative experimental design promises to foster cumulative knowledge by changing how we design experiments, build theories, and conduct research. I support the push to increase commensurability across experimental research but raise several reservations regarding results-driven and large-team-based research. I argue that it is vital to preserve academic diversity and adversarial debate via independent efforts.

The proposed integrative experimental design approach consists of three steps: 1) explicitly define the design space of the experiments in terms of features of the decision situation and the population sample, 2) systematically sample from that design space, and 3) build theories by quantifying the outcome heterogeneity over that space. This approach will guarantee commensurability between different experiments and findings and foster cumulative knowledge. The authors' concept of "research cartography" is brilliant – the idea is to itemize, standardize, categorize, and quantify the information that we typically and only partially reveal in the Methodology and Discussion sections of our research papers. The image of a Wikidata-style database containing all experimental (and in fact, any other) social and behavioral knowledge is incredibly appealing! The Cooperation Databank, for instance, offers a glimpse of how such a database could look like (Spadaro et al., 2020). Developing research cartography will help identify research gaps, established findings, and controversial problems. The approach will also aid the reuse and re-analysis of existing data to answer new research questions (Almaatouq et al., 2022; Rand et al., 2012; Tsvetkova et al., 2018). In short, whether retrospective or prospective, a comprehensive and systematic research cartography will help consolidate knowledge and stimulate new research.

The integrative experimental design approach, however, presses further – steps 2) and 3) propose to consolidate research and theory-testing efforts by sampling and generalizing over many points in the experimental design space simultaneously, rather than "one at a time." Yet, these steps are not necessary for commensurability and more importantly, carry negative implications for diversity, innovation, and productive debate in academic research. There are several issues I would like to raise here.

First, the proposed paradigm threatens to entrench and exacerbate existing inequalities within and between scholarly communities. Participating in global research consortia may be open to many but who leads these consortia will likely befall on those with status,

prestige, and funding. It is hard to overlook the fact that the authors speak from a position of privilege – they work at prestigious US universities, with access to hefty research funds and numerous PhD students and postdoctoral researchers. The large-scale research they propose is simply not accessible to many experimental researchers.

Second, the proposed paradigm aims to optimize efficiency in research but this is a misguided ideal. Academic research is not just about results but also about exploration and discovery, critique and debate, learning and training. Consolidating research activities in hierarchically structured labs or consortia with established protocols and routines may reduce labor costs but stifle entrepreneurship, critical thinking, and iconoclastic innovation. Based on some of the authors' empirical examples, the complexities of group synergies imply that different problems would be best addressed by teams of different size and composition (Almaatouq et al., 2021; Mao et al., 2016; Straub et al., 2023). This calls for independence, plurality, redundancy, and diversification of research effort, not consolidation.

Third, in the social and behavioral sciences, raising the question is often more important than finding out the answer. Much like the observer effect related to measuring physical systems, studying a social system changes it. Posing a specific social research problem can shape political debate, policy decisions, organization strategies, and collective behavior. Consolidating funding and research efforts forebodes a monopoly over setting research agendas and directions, the muffling of marginalized voices, the side-lining of localized problems, and the suppression of new perspectives and paradigms. Large-scale integrative experiments may be good for providing definitive evidence to integrate and reconcile existing theories but restricted when it comes to launching new research agendas.

Related to the latter issue, the proposed result-driven active-learning sampling strategy for experiments threatens to shift the focus to effects that are sizeable but not necessarily meaningful or important. Specifically, certain combinations of context and population features may be impossible or unlikely and hence, practically irrelevant. In short, the integrative experimental design approach does not alleviate and may even exacerbate the thorniest problem of experimental research – external validity. Explaining all variation is not always the best strategy for good or efficient science: the power of good general theories is not that they are universally true but that they apply to statistically likely/common situations and hence, they are useful.

I acknowledge that the authors present integrative experiment design as an additional, and not the only true, approach to experimentation in the social and behavioral sciences. I assumed an exaggeratedly antagonistic stance here to caution against consolidation. There are alternatives, such as adversarial collaboration (Killingsworth et al., 2023; Mellers et al., 2001), that can help reconcile contradictory findings without compromising debate, plurality, and diversity.

Competing interests: None

Funding: This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

References

Almaatouq, A., Alsobay, M., Yin, M., & Watts, D. J. (2021). Task complexity moderates group synergy. *Proceedings of the National Academy of Sciences*, 118(36), e2101062118. <https://doi.org/10.1073/pnas.2101062118>

Almaatouq, A., Rahimian, M. A., Burton, J. W., & Alhajri, A. (2022). The distribution of initial estimates moderates the effect of social influence on the wisdom of the crowd. *Scientific Reports*, 12(1), Article 1. <https://doi.org/10.1038/s41598-022-20551-7>

Killingsworth, M. A., Kahneman, D., & Mellers, B. (2023). Income and emotional well-being: A conflict resolved. *Proceedings of the National Academy of Sciences*, 120(10), e2208661120. <https://doi.org/10.1073/pnas.2208661120>

Mao, A., Mason, W., Suri, S., & Watts, D. J. (2016). An experimental study of team size and performance on a complex task. *PLOS ONE*, 11(4), e0153048. <https://doi.org/10.1371/journal.pone.0153048>

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, 12(4), 269–275. <https://doi.org/10.1111/1467-9280.00350>

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, 489(7416), 427–430. <https://doi.org/10.1038/nature11467>

Spadaro, G., Tiddi, I., Columbus, S., Jin, S., Teije, A. ten, & Balliet, D. (2020). *The Cooperation Databank*. PsyArXiv. <https://doi.org/10.31234/osf.io/rveh3>

Straub, V. J., Tsvetkova, M., & Yasseri, T. (2023). *The cost of coordination can exceed the benefit of collaboration in performing complex tasks* (arXiv:2009.11038). arXiv. <https://doi.org/10.48550/arXiv.2009.11038>

Tsvetkova, M., Wagner, C., & Mao, A. (2018). The emergence of inequality in social groups: Network structure and institutions affect the distribution of earnings in cooperation games. *PLOS ONE*, 13(7), e0200965. <https://doi.org/10.1371/journal.pone.0200965>