High-throughput experiments in small-group deliberation

Keywords: Virtual-lab, High-throughput Experiments, Deliberation, Small-groups

Extended Abstract

The effectiveness of small group deliberation, like many problems in social science, depends on context. There is no one-size-fits-all solution for improving deliberation outcomes, so it is important to understand how interventions can be tailored to suit specific deliberative situations. However, traditional experimental methods, which are focused on finding support for particular theories, tend to limit examination to a few interventions in a small number of contexts. This makes it difficult to empirically measure the effect of context or make confident contextualized policies. This talk presents a method and experiment platform for conducting small-group discussion experiments at scale. This approach lets us explore a design space of all possible experiments, sampling adaptively to maximize our ability to predict out-of-sample data. We present preliminary data from a study to evaluate training interventions that support cross-partisan dialog, while varying group composition and topic partisanship.

Deliberation occurs whenever people come together to discuss issues, weigh alternatives, or make decisions. Scholars and practitioners have suggested numerous strategies to achieve positive outcomes from deliberation, such as training discussants in skills like active listening or enforcing rules and procedures for discussion (Gutmann and Thompson 2009; Fishkin et al. 2021). Although it seems obvious that no single intervention will be optimal for every deliberation, it is not obvious from the academic literature how interventions should be tailored to specific contexts. Surveying the literature, Mutz (2008) specifically criticizes the lack of systematization and clear specification of concepts that would enable comparability between different experiments. Without this systematic framework, she writes, "there is little hope that empirical research will usefully speak to deliberative theory, nor that theory will speak to practice."

This criticism is not unique to the deliberation literature; it could be leveled at essentially any nontrivial problem in social and behavioral science (Watts 2017; Yarkoni 2019; Bryan, Tipton, and Yeager 2021). All such problems exist in high-dimensional context spaces where every conceivable study design falls at a particular point in the space. In the standard experimental paradigm, each experiment samples at most a small number of interventions in a small number of contexts. These experiments are appealing because they are easy to motivate with theory and can be performed by a small team of researchers within typical academic budgets and timelines. However, to systematically explore the context-dependency of an intervention, we need a method for collecting and synthesizing orders of magnitude more experiments in a purposeful and commensurable way. In this talk, we propose such a method and demonstrate its use in testing interventions designed to improve cross-partisan dialog.

Mapping the design space. We start by identifying a set of factors that are likely to influence deliberation interventions. For example, the partisan makeup of the group, or the polarized nature of the discussion topic. These features define the "experimental design space": a high-dimensional landscape in which every point corresponds to a possible experiment, and which contains all possible experiments in our domain of interest.

Sampling experiments in an integrative manner. We use this design space to proactively sample experiments to cover the range of deliberative contexts. Initially, we choose samples to replicate previous experiments and to take a random sample over the space. We then use machine learning methods to actively "learn" over the landscape. These models use all the information we have collected, weighted by proximity in the design space, to make statistical predictions about the outcomes we should expect to see in parts of the space that we have not yet sampled. After making these out-of-sample predictions, we collect data at the predicted points and use this "testing" data to evaluate the accuracy of our statistical models. We also use these statistical models to identify "high-leverage" experimental conditions to prioritize in our next round of data collection.

Conducting individual experiments. To conduct individual experiments, we recruit a group of participants and ask them a series of questions to understand their characteristics, such as demographics and personality traits. We then assemble groups with various distributions of these characteristics. Participants enter a video call interface where they discuss a provided topic and are provided with audiovisual or interactive interventions. The deliberation is recorded, and we use video-processing algorithms to extract features that may explain the effect of the interventions. After the deliberation, participants are asked questions about their experience and changes in their opinions.

While this “high-throughput” approach requires more data collection than the average study it also generates more cumulative and mutually comparable knowledge per time and budget than traditional, isolated experiments (Almaatouq et al. 2022).

Putting the approach into practice. We demonstrate this approach with an experiment designed to test whether interventions shown to mitigate self-reported partisan animosity create similar changes in behavior during cross-partisan dialog (Voelkel et al. 2022). We initially begin with a small number of dimensions in the context space, defined by the party affiliation of each individual and the contentiousness and partisanship of the issues being discussed. We actively explore these dimensions, creating a map in which certain regions of the experimental design space are best served by one intervention and other regions by a different intervention.

References

Almaatouq, Abdullah, Thomas L. Griffiths, Jordan W. Suchow, Mark E. Whiting, James Evans, and Duncan J. Watts. 2022. “Beyond Playing 20 Questions with Nature: Integrative Experiment Design in the Social and Behavioral Sciences.” *The Behavioral and Brain Sciences*.

Bryan, Christopher J., Elizabeth Tipton, and David S. Yeager. 2021. “Behavioural Science Is Unlikely to Change the World without a Heterogeneity Revolution.” *Nature Human Behaviour* 5 (8): 980–89.

Fishkin, James, Alice Siu, Larry Diamond, and Norman Bradburn. 2021. “Is Deliberation an Antidote to Extreme Partisan Polarization? Reflections on ‘America in One Room.’” *The American Political Science Review* 115 (4): 1464–81.

Gutmann, Amy, and Dennis F. Thompson. 2009. *Why Deliberative Democracy?* Princeton University Press.

Mutz, Diana C. 2008. “Is Deliberative Democracy a Falsifiable Theory?” *Annual Review of Political Science* 11 (1): 521–38.

Voelkel, Jan, Michael N. Stagnaro, James Chu, Sophia Pink, Joseph S. Mernyk, Chrystal Redekopp, Matthew Cashman, James N. Druckman, David G. Rand, and Robb Willer. 2022. “Megastudy Identifying Successful Interventions to Strengthen Americans’ Democratic Attitudes.”

Watts, Duncan J. 2017. “Should Social Science Be More Solution-Oriented?” *Nature Human Behaviour* 1 (1): 1–5.

Yarkoni, Tal. 2019. “The Generalizability Crisis.” *PsyArXiv*. https://doi.org/10.31234/osf.io/jqw35.