

ARTICLE

When Deliberation Produces Persuasion rather than Polarization: Measuring and modeling Small Group Dynamics in a Field Experiment

Kevin M. Esterling^{1*} , Archon Fung² and Taeku Lee³

¹Department of Political Science, University of California – Riverside, ²JFK School of Government, Harvard University and

³Department of Political Science, University of California – Berkeley

*Corresponding author. E-mail: kevin.esterling@ucr.edu

(Received 21 June 2018; revised 12 April 2019; accepted 30 April 2019; first published online 19 December 2019)

This article proposes a new statistical method to measure persuasion within small groups, and applies this approach to a large-scale randomized deliberative experiment. The authors define the construct of ‘persuasion’ as a change in the systematic component of an individual’s preference, separate from measurement error, that results from exposure to interpersonal interaction. Their method separately measures persuasion in a latent (left–right) preference space and in a topic-specific preference space. The model’s functional form accommodates tests of substantive hypotheses found in the small-group literature. The article illustrates the measurement method by examining changes in study participants’ views on US fiscal policy resulting from the composition of the small discussion groups to which they were randomly assigned. The results are inconsistent with the ‘law of small-group polarization’, the typical result found in small-group research; instead, the authors observe patterns of latent and policy-specific persuasion consistent with the aspirations of deliberation.

Keywords: persuasion; measurement; randomized design; Bayesian statistics; deliberative democracy

Persuasion is central to any conception of democratic political communication (Broockman and Kalla 2016; Kalla and Broockman 2018; Minozzi et al. 2015; Mutz, Sniderman and Brody 1996). For example, one of the core tenets of deliberative democracy (Gutmann and Thompson 1996; Habermas 1984) holds that preferences among debate participants should be responsive to arguments, at least on occasion.¹ When debate participants recognize merits in each others’ claims, policy agreements possess legitimacy beyond that gained from majority-rule voting (Cohen 1989).

We propose a novel method for modeling persuasion within small groups, which is applicable when assignment to groups is randomized. Our method measures the extent to which individual preference change is caused by exposure to interpersonal interactions within a small group, after netting out measurement error. We partition measured persuasion into two components: (1) *latent persuasion*, which is the amount an individual changes on an underlying, left–right dimension that structures preferences across a set of policy items and (2) *topic-specific persuasion*, which is the amount an individual changes preferences on a given topic, such as a policy option, net of latent preferences (similar to Lauderdale, Hanretty and Vivyan 2018). Randomization is the key to identifying both of these components of persuasion; without randomization the model results are likely to be driven by confounding through self-selection processes.

¹The possibility of non-coercive persuasion is central to Gutmann and Thompson’s (1996, 52) conception of ‘reciprocity’ and Habermas’s (1984, 9) conception of ‘communicative action’.

© The Author(s) 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

We demonstrate this method by testing for the causal effects of exposure to small-group discussion at the ‘Our Budget, Our Economy’ nationwide town hall meetings organized by *AmericaSpeaks*, an event at which nearly 3,000 participants were randomly assigned to small-group discussion tables. The event was held on 26 June 2010 at town halls in nineteen separate cities, with 100–500 participants in each town hall. Within each town hall, participants’ seating assignments were randomized among small-group discussion tables, and we administered opinion surveys both before and after the event. We use this application to demonstrate our novel strategy to measure persuasion within small groups, and to assess the extent and nature of persuasion that occurred at this event.

Substantively, our application demonstrates that the amount and nature of persuasion we observe fulfills many of the normative aspirations of deliberative democracy. In particular, the patterns of persuasion we observe are inconsistent with the ‘law of small-group polarization’ that is the typical finding in small-group research (Sunstein 2002), as well as with motivated reasoning observed in partisan contexts (Bolsen, Druckman and Cook 2014). Instead, we observe measured persuasion at both the latent and topic levels, and find that participants who perceived the discussion to be well informed were the most likely to be persuaded to accept policies inconsistent with their initial predispositions. We argue that the careful design of deliberative small groups can be effective at triggering rational or System II (Broockman and Kalla 2016; Kahneman 2011) interactions.

MEASURING AND MODELING PERSUASION

The standard approach to measuring persuasion in the small-group literature evaluates changes in a discussion participant’s self-reported preferences from before to after a discussion event (for example, Grönlund, Herne and Setälä 2015; Schkade, Sunstein and Hastie 2010; Westwood 2015). In a basic small-group design, the researcher typically administers a survey to each participant before exposure to the group to measure their pre-treatment preferences on an item or topic, which we will label O_i^0 . Next, the researcher randomizes each participant’s assignment to a small group. This randomization varies the composition of the group to which each participant is exposed. For example, randomization will vary the distribution of ideological ideal points within a group, so in small groups they will randomly assign each participant to a group that is on average either liberal or conservative (or anything in between), and that is either diverse in ideology or homogeneous. The groups are invited to have a discussion, and afterwards the researcher will measure the respondents’ post-treatment preferences, O_i^1 .

In the basic design, the researcher will conduct a statistical test to see if there is a relationship between group exposure and the difference between the pre-treatment and post-treatment response, $(O_i^1 - O_i^0)$. Farrar et al. (2009, 619) is an exemplar of current practices, which models preference change in response to exposure to a small-group discussion as:

$$O_i^1 = \beta_0 + \beta_1 O_i^0 + \beta_2 H_i + \beta_3 \text{Site}_i + \epsilon_i \quad (1a)$$

$$H_i = \frac{1}{n_i - 1} \sum_j O_j^0, \quad j \in \{J_i : j \text{ is seated at } i\text{'s table, } j \neq i\} \quad (1b)$$

where the i^{th} respondent’s post-treatment preference on a given topic (O_i^1) is modeled as a function of her own pre-treatment preference (O_i^0), the average (H_i) of the pre-treatment preferences of her $(n_i - 1)$ discussion partners (indexed by $j \in J_i$), and separate intercepts for each location

(*Site_i*) at which the discussions were held. Farrar *et al.* (2009) model preference change as the difference in pre-post survey responses by subtracting $\beta_1 O_i^0$ from both sides of Equation 1.²

In the Farrar *et al.* (2009) study, as in our own application, the respondent is randomly assigned to discussion groups so the average of the pre-treatment preferences of her discussion partners (H_i) is also random, and under the normal assumptions for identifying a causal effect in a randomized control trial that we describe in more detail below, β_2 identifies the causal effect on the respondent’s change in response to the survey item from the pre-test to the post-test, ($O_i^1 - \beta_1 O_i^0$), that comes from exposure to a discussion group with a given composition of participants (see also Gastil *et al.* 2008; Klar 2014).³

The difference in pre- and post-survey responses, however, does not map onto persuasion as a construct because the pre-test and post-test responses each contain a stochastic component related to measurement error (Achen 1975; Ansolabehere, Rodden and Snyder 2008; Prior 2010), in addition to a systematic component that captures respondents’ preferences at a given time. Only a change in the systematic component that results from some intervention, such as interpersonal interactions within a discussion, should count as a valid measure of persuasion; random noise should not.⁴

To formalize the systematic component for preference change, for simplicity we assume a continuous, normally distributed opinion response at time t , O_i^t , and decompose the opinion response as:

$$O_i^t = \beta_0 + \theta_i^t + \zeta_i^t + \epsilon_i^t, \quad t \in \{0, 1\} \tag{2}$$

where θ_i^t is the respondent’s latent, left–right ‘ideal point’ that structures preferences across a range of issues (Hinich and Munger 1994),⁵ ζ_i^t is a topic-specific preference that remains after netting out latent preferences, and ϵ_i^t is the idiosyncratic component from measurement error that represents instability in the individual’s opinion response (Lauderdale, Hanretty and Vivyan 2018), all evaluated at time t ; $t = 0$ is the pre-test and $t = 1$ is the post-test value.⁶ If θ_i^t and ζ_i^t are invariant or fixed over time, then opinion change is driven only by the idiosyncratic component and is essentially noise.

The statistical task is to separate out systematic preference change in θ_i^t and ζ_i^t from random noise using a measurement strategy, and then to model the two systematic components directly. To derive a model of preference change over time from first principles, we can take the difference in Equation 2 between time $t = 1$ and $t = 0$:

$$O_i^1 = \beta_0^1 + \theta_i^1 + \zeta_i^1 + \epsilon_i^1 \tag{3a}$$

$$\beta_1(O_i^0 = \beta_0^0 + \theta_i^0 + \zeta_i^0 + \epsilon_i^0). \tag{3b}$$

²Including the pre-treatment response in the model as a right-hand-side variable identifies the β_1 coefficient, which allows the scale of the preference item to change over time. One can constrain $\beta = 1$ to set the scales equal.

³As we discuss more extensively below, the model tests for the causal effect of exposure to a given composition of participants in the discussion group, which is randomized in the study design, rather than exposure to the discussion itself, which is not randomized. Pre-treatment preference is an instrument for what participants say in discussion, and so the model identifies the complier average causal effect of exposure to a discussion (see Angrist, Imbens and Rubin 1996).

⁴Note that this definition of persuasion is not limited to rational persuasion (Habermas 1984); in the application below we demonstrate methods to assess the nature of persuasion including its rationality using the concept of construct validity.

⁵When the scale has a left–right orientation, the institutional literature labels this latent preference as the respondent’s ‘ideology’, which can typically be scaled using a single dimension (Clinton 2012; Poole and Rosenthal 1997). The assumption of unidimensionality is not necessary and the model below can accommodate an arbitrary number of dimensions through a more elaborate design.

⁶The complex structure of group data separately identifies the θ_i^t and ζ_i^t parameters. For example, in the application below, we identify θ_i^t by nesting questions within participants, and ζ_i^t by nesting participants within discussion groups.

Subtracting Equation 3a from Equation 3b and rearranging yields:

$$O_i^1 = \beta_0 + \beta_1 O_i^0 + \Delta\theta_i + \Delta\zeta_i + \epsilon_i \quad (4)$$

where $\beta_0 = \beta_0^1 - \beta_1 \beta_0^0$ and $\epsilon_i = \epsilon_i^1 - \beta_1 \epsilon_i^0$. With this derivation we have identified two new quantities, $\Delta\theta_i = \theta_i^1 - \beta_1 \theta_i^0$ which is the change in the respondent's pre- to post-discussion preferences in the *latent preference* space, and $\Delta\zeta_i = \zeta_i^1 - \beta_1 \zeta_i^0$ which is the change in the respondent's *topic-specific preference* for the outcome represented by O_i after accounting for changes in latent preferences. This derivation allows us to focus on these more substantively interesting preference changes, rather than only on the noisily measured changes in the survey response itself. We define *measured persuasion* as the change in each of these systematic components of the respondent's expressed preference.

Consider the two systematic components of preference change in turn. First, one can consider latent preferences to be a heuristic, such as left-right ideology, that enables individuals to make sense of and engage in policy debates involving complex matters even with limited information (Eatwell 1993; Hinich and Munger 1994). In this interpretation, $\Delta\theta_i$ captures changes in the latent structuring of their preferences that organizes their views across a range of policies. In the American context, ideology largely reduces to a single, latent dimension (Poole and Rosenthal 1997). For example, in the context of our application on US fiscal policy that we describe below, as an empirical matter all preferences load exclusively on a single latent dimension captured by θ .

Secondly, the structure of preferences within specific policy topics can be complex (Feldman and Johnson 2014; Treier and Hillygus 2009), and particularly at the elite level or within deliberative communication (Gutmann and Thompson 1996, 56; Habermas 1984, 99), reasoning about policy topics is not strictly constrained by ideology or to any other single latent dimension (see Tausanovitch and Warshaw 2017). Such an assumption would be overly restrictive and indeed a gross oversimplification of human cognition. For example, in the town hall event we study, participants were provided policy reading material and expert testimony to inform discussions, and so had the capacity to give reasons and exchange rationales that go beyond a heuristic defined by ideology. In this view, the $\Delta\zeta_i$ measure of topic-specific persuasion captures the amount of persuasion that occurs 'outside' of the latent scale.

Thus, within a small-group event, persuasive processes can operate at these two different levels. Note that this partitioning between $\Delta\theta_i$ and $\Delta\zeta_i$ does not create a hierarchy among latent and topic-specific reasoning. In the statistical model, the relative amount of each can vary freely across individuals.

As is common practice (for example, Farrar et al. 2009), we allow the scale of the response space to vary over time by multiplying both sides of Equation 3b by β_1 . For example, $\beta_1 < 0$ implies a plenary shift in preferences toward moderation and $\beta_1 > 0$ implies a plenary shift toward extremity. Note that in the case of both $\Delta\theta_i$ and $\Delta\zeta_i$, the change in systematic preferences is based on the underlying preference space rescaled by β_1 . One can fix the scales across the two time periods by setting $\beta_1 = 1$, which is equivalent to modeling the difference ($O_i^1 - O_i^0$) as an outcome (such as in Westwood 2015).

In general, including an outcome response variable measured pre-treatment, such as O_i^0 , on the right-hand side will lead to endogeneity bias since many of the individual-level determinants of an outcome in the pre-treatment period also determine the outcome in the post-treatment period. To see why in the case of modeling preference change, define $\omega_i^t = \theta_i^t + \zeta_i^t$, and note that $cov(\omega_i^0, \omega_i^1) \neq 0$, since $\theta_i^1 = \theta_i^0 + \Delta\theta_i$ and so θ_i^0 is contained in both ω_i^0 and ω_i^1 . In the statistical model below we correct for this potential bias by including θ_i^0 in the outcome equations. In essence, we guard against endogeneity bias under the assumption that the latent preference

scale is a strong predictor of both pre- and post-discussion preferences, and that the remaining variation in preferences O_i^0 and O_i^1 is random once a respondent's ideal point is accounted for.⁷ Thus we estimate the following equation:

$$O_i^1 = \beta_0 + \beta_1 O_i^0 + \beta_2 \theta_i^0 + \Delta\theta_i + \Delta\zeta_i + \epsilon_i \quad (5)$$

This model differs from the standard practice for modeling persuasion (for example, Farrar *et al.* 2009) in three important ways. First, our model includes pre-test preferences O_i^0 in a way that does not induce endogenous variable bias.

Secondly, our model focuses on the change in the respondent's latent and topic-specific preferences, represented by $\Delta\theta_i$ and $\Delta\zeta_i$, rather than the raw opinion change ($O_i^1 - O_i^0$) that is at best a noisy measure of persuasion. That is, in the model below, the components of measured persuasion $\Delta\theta_i$ and $\Delta\zeta_i$ serve as the outcomes of interest, which we model directly.⁸ These two components are similar to factor scores in that they both measure quantities that are unobserved directly in the data.

Thirdly, modeling the θ_i^0 , $\Delta\theta_i$ and $\Delta\zeta_i$ parameters jointly with the structural parameters β correctly propagates the uncertainty that comes from estimating these parameters using the statistical model. That is, the estimates of the β parameters are the marginal distributions having integrated over the sample space underlying θ_i^0 , $\Delta\theta_i$ and $\Delta\zeta_i$, a technique known as the 'method of composition' (Treier and Jackman 2013). As Treier and Jackman (2013) note, the method of composition accounts for the estimation uncertainty in each of the random-effect parameters. This is in contrast to using estimated factors scores as right-hand-side variables, as if the factor scores correctly recovered each random-effect parameter, which causes errors in variables bias in a regression.⁹

A research design that would enable this statistical strategy to measure the systematic component of preference change has two main requirements. First, the respondent must express preferences on three or more topics in both the pre- and post-discussion surveys in order to identify the underlying latent preference space in $\Delta\theta_i$. If multiple outcomes do not exist, then only $\Delta\zeta_i$ is identified. Secondly, the standard assumptions in evaluating randomized control trials must be met (Angrist, Imbens and Rubin 1996; Gerber and Green 2012) in order to identify the effect of group composition rather than confounds to each group's discussion. We discuss these assumptions below.

THE OBOE TOWN HALLS

We apply our measurement strategy in a test of small-group persuasion using a dataset from a randomized, large-scale deliberative field experiment. On 26 June 2010, nearly 3,000 individuals

⁷By adding θ_i^0 to the model, we further change the mapping of the scale of the underlying ideological spaces in $\Delta\theta_i$ from β_1 to $(\beta_1 + \beta_2)$. This is only a mathematical transformation and highlights that scales do not have a ratio level of measurement and so require a transformation to bridge one space into the other. If one had substantive reasons to assume the two scales are identical in a specific application, one can instead choose to estimate a restricted model with $\beta_1 = 1$, $\beta_2 = 0$, and then assume (and hope) endogeneity bias does not exist in the application.

⁸Modeling the change in noisy outcomes directly reduces the power of a statistical test. As we describe in Appendix Section A.5, we implement the Farrar *et al.* (2009) regression model using the data from our application. For the six outcomes, only two of the group composition effects show a statistically significant impact. Applying the Bonferroni correction for multiple comparisons shows that the effects for the set of outcomes are not significant at standard levels when using the Farrar model.

⁹We show in the appendix that using factor scores as right-hand-side variables attenuates the structural parameter estimates, a clear indicating of measurement error. We also show that one can approximate the full model solution by using the post-treatment factor score as a left-hand-side variable since the regression can account for measurement error on the dependent variable. This latter regression using factor scores is a simplified solution and thus serves as a good robustness check for the full model.

in nineteen cities convened in town hall meetings to discuss America's long-term fiscal future.¹⁰ The event, entitled 'Our Budget, Our Economy' (OBOE), brought together diverse citizen-deliberators, armed with background reading material, to discuss and prioritize policy options that would help put the nation's budget on a more sustainable long-term fiscal path. To recruit participants, the event organizer, *AmericaSpeaks*, worked with hundreds of local groups in each of the nineteen cities, from all walks of life, to create a group of participants that closely mirrors the demographic composition of each community (see Appendix Section A.1 for a description of the event, recruitment and the respondents' characteristics).¹¹ In addition, *AmericaSpeaks* worked with over thirty national organizations that research and advocate budget policies, both liberal and conservative, to develop technical background reading material that was factual, balanced and that represented the views of diverse perspectives.

On the day of the event, participants were randomly assigned to small-group discussion tables; the randomization occurred within each site.¹² They spent the entire day reading the materials, watching some instructional videos and discussing their policy views with others seated at their table. Given the diversity of the participants in the town halls, randomization served two purposes. First, randomizing participants to small-group discussions helped to assure that many participants were exposed to the views of co-discussants who were very different from themselves. In the absence of pre-determined seating assignments, participants are likely to seek out co-discussants who are like themselves (Fowler et al. 2011), or to sit with other participants with whom they arrived at the event, which in turn would minimize the diversity of viewpoints available at each table. Randomization washes out any existing social ties among participants and diversifies the views to which participants are exposed. Since the groups were small in number, typically ten participants, sampling variability under randomization assured that the composition of preferences would vary across tables, ranging from homogeneous to heterogeneous groups.

Secondly, random assignment allows us to identify the causal effects of exposure to different group compositions (Farrar et al. 2009) and therefore enables us to identify our measure of persuasion as a function of pre-discussion viewpoints. In the present case, the mix of pre-discussion viewpoints among participants at a given table is exogenous to the analysis. One might believe a better measure of persuasion would rely on the arguments actually made in the course of the discussion, say from a transcript of the session (for example, Karpowitz and Mendelberg 2007; Westwood 2015). This measurement strategy, however, cannot test for causal effects as the arguments offered during a discussion occur post-treatment; that is, since arguments are not randomly assigned, a statistical test based on arguments will lack internal validity.¹³ Instead, we rely on the composition of pre-test ideal points of the other participants at the respondent's table as an instrument of exposure to viewpoints during the discussion, since we can take the pre-test ideal points of the discussion partners as an exogenous and randomly assigned

¹⁰The event was held simultaneously in nineteen sites in nineteen different cities, and the sites were co-ordinated via videoconferencing technology. Six of the sites were designated 'large sites' with approximately 500 participants each: Albuquerque, Chicago, Columbia (SC), Dallas, Philadelphia, and Portland (OR). The remaining sites were smaller and had 100 or fewer participants: Los Angeles, Des Moines, Overland Park (Kansas City), Louisville, Augusta (ME), Detroit, Jackson (MS), Missoula, Portsmouth (NH), Grand Forks, Richmond, Caspar and Palo Alto. A table in Appendix Section A.3 gives the number of participants at each site.

¹¹The recruitment is similar to Barabas (2004). Since *AmericaSpeaks* could not compel a truly representative sample of citizens to participate in the experiment (see Fishkin and Luskin 2005; Luskin, Fishkin and Jowell 2002), we can only state the in-sample group dynamics. The in-sample results remain interesting since they test for dynamics among those who have a propensity to show up to a deliberation.

¹²Prior to the event, the organizers printed cards with table numbers, and then shuffled the cards before handing them to participants as they arrived. Randomization and balance tests show that the quality of the randomization was very good. See Appendix Section A.4 for a detailed analysis.

¹³Using post-treatment arguments as a causal variable would require the much stronger 'sequential ignorability' assumption from mediation analysis (Imai et al. 2011), which our design does not support. Hence, conditioning on the arguments made in the discussions would lack internal validity and expose the results to confounding.

encouragement to create the mix of arguments made in the discussion under an encouragement design (as in Farrar *et al.* 2009). To connect the group compositions to persuasion from discussion, our strategy must assume that there is a larger mix of conservative arguments made at a discussion where most of the participants have conservative pre-discussion ideal points compared to tables where most of the participants have liberal ideal points, and vice versa.

The group context in which deliberation occurs can affect the nature of discussion. In the OBOE deliberation, the discussion groups were typically not homogeneous; the discussion focused extensively on a single topic; the participants had access to balanced, factual reading material; and *AmericaSpeaks* assigned a moderator to each table. The moderator did not participate substantively in the discussion and was trained by the event organizers in techniques to ensure that everyone at the table had the chance to speak, to encourage everyone to participate, and to enforce a set of rules (written on cards located at the center of each table) that were designed to make each table a ‘neutral, safe space’ for expressing diverse views. We expect this careful design to induce deliberative exchanges within the small groups (Barabas 2004; Gastil *et al.* 2008; Gerber *et al.* 2016; Grönlund, Herne and Setälä 2015; Luskin, Fishkin and Hahn 2007), and so our findings might well depart from those of non-deliberative small-group studies (see Isenberg 1986).

DATA AND MODEL

The statistical model tests for the presence of persuasion within the small groups regarding various policy proposals considered at the event. At each of the nineteen town halls, we asked participants to complete a short survey as they arrived, before the event began, and to complete another survey at the conclusion of the event. We refer to the former as the pre-test survey, and the latter as the post-test survey. A total of 2,793 participants, seated at 339 tables across nineteen sites, filled out one or the other or (for the vast majority) both of these surveys.¹⁴

The pre- and post-test surveys each contained a block of items asking participants their policy preferences on a set of proposals. The block of six questions is preceded with ‘Here are several things the government could do to cut the budget deficit. Please tell us what you think about each approach to reducing the deficit.’ The response categories each have a five-point scale: ‘Strongly disagree’, ‘Disagree’, ‘Neither’, ‘Agree’ or ‘Strongly agree.’ The items are (labels for items shown below in bold font were *not* in the survey):

- Q1: Tax Rich** Raise income taxes on the very wealthy – individuals making \$250,000 or more and households making \$500,000 or more.
- Q2: Cut Programs** Cut discretionary federal programs and services by 5 per cent across the board.
- Q3: Cut Entitlements** Cut the growth of spending on entitlement programs such as social security and Medicare benefits.
- Q4: Cut Defense** Cut the spending on national defense and the military.
- Q5: Tax Both** Raise taxes on the middle class as well as the wealthy.
- Q6: Federal Sales Tax** Create a new federal consumption tax, which would be like a federal sales tax that would be on top of any state and local sales tax.

In the American context, the first four items have a clear left–right orientation: to solve the deficit, liberals prefer to tax the rich and cut defense; conservatives prefer to cut discretionary programs and entitlements. The remaining two items that advocate taxing the middle class and a federal sales tax cut across liberal–conservative ideology. The statistical model uses

¹⁴Because the analysis depends on table-level summary statistic functions, we drop all tables with fewer than five participants. This omits forty-six participants who were seated at twenty tables, which is less than 2 per cent of the sample.

the pre- and post-test values of these items; an indicator of whether the pre-test is missing (9 per cent of pre-tests are missing),¹⁵ a variable indicating a unique table identification number (among 339 tables total), and dummy variables indicating the site (out of the nineteen sites, omitting one site) for each participant. The Appendix Section A.2 provides summary statistics for all of the variables.

Statistical Model

Our statistical model estimates the effect of small-group composition on persuasion, exploiting the random assignment to groups and a measurement model. The full statistical model is given in Appendix Section A.7. In this section we ‘walk through’ the elements of the likelihood function in order to show how we measure persuasion, and how the parameters and functional form specifications allow us to test a variety of substantive hypotheses regarding persuasion that are found in the literature on small-group dynamics. The likelihood equation for a single categorical outcome is summarized in Equation 6a, which is a non-linear implementation of Equation 5.

$$O_{ik}^1 \sim \text{Ordered Logit}(\beta_{1k}O_{ik}^0 + \beta_{2k}\theta_i^0 + \beta_{3k}\text{Site}_i + \omega_{ik}), \quad (6a)$$

$$\omega_{ik} = \Delta\theta_i + \Delta\zeta_{ik}. \quad (6b)$$

We estimate this model simultaneously for each of six policy preference items. In this equation, i indexes N participants (each i is a potential ‘persuadee’) and k indexes $K = 6$ policies, which are labeled Q1 to Q6 above. The post-test policy preferences for each item and for each individual, O_{ik}^1 , are modeled as a function of her pre-test policy preference O_{ik}^0 , her pre-test left–right ideal point θ_i^0 , an indicator of the Site _{i} (city) of her event, and a random effect ω_{ik} that varies across individuals and policies. We describe each of these four elements in turn, noting for now that our main interest will focus on ω_{ik} .

The first component (O_{ik}^0) is the respondent’s pre-treatment response on the respective policy preference survey item. Including the pre-treatment opinion on the right-hand side ensures that the structural parameters in the model estimate the individual’s *change* in preference that occurs between the pre- and post-test surveys (Farrar et al. 2009). As we describe above, including the pre-treatment outcome on the right-hand side and estimating the β_{1k} parameter allows the scale of the post-treatment outcome to vary. Since O_{ik}^0 is categorical, we include a set of dummy variables indicating each of the first four response categories for the pre-test item (omitting the fifth category), and hence O_{ik}^0 is a matrix and β_{1k} is a vector. Using these dummy variables enables us to relax an assumption that each response category predicts the post-test response equally and in the same direction, and also allows the degree of scale compression and expansion to vary across the response options.

For the second component, we include θ_i^0 in the likelihood function to capture the endogenous dependence between the pre- and post-test responses on the outcome, which corrects for any endogenous variable bias that comes from including the pre-test item in the outcome equation (see Skrondal and Rabe-Hesketh, 2004, 107–8). We use pre-test responses to the tax rich, cut programs, cut entitlements and cut defense (Q1 to Q4) items to estimate each participant’s pre-treatment latent ideal point preference scale, since these items have a clear liberal–conservative

¹⁵See Appendix Section A.12 for sensitivity tests that assess the possible range of estimates that would result under different extreme distributions of missing pre-test data. Of those who filled out a pre-test questionnaire, 22 per cent failed to fill out a post-test survey. We impute missing post-test data as missing at random conditional on the respondent’s pre-test response on the policy item, her ideology and the ideological composition of her table.

orientation.¹⁶ We estimate each participant’s latent ideal point θ_i^0 dynamically within the model, as in a structural equation model, and hence the estimation uncertainty inherent in θ_i^0 is included in the likelihood.

For the third component we condition on the *Site* or city in which the participants’ event took place. Since randomization took place within sites these fixed effects allow us to control for any site-specific influences.

The fourth component of the likelihood function is a random effect, ω_{ik} , that varies across individuals and policies.¹⁷ ω_{ik} measures the amount of dependence among the participants’ preference changes in communication with each other (Anselin 1988), for both the latent preferences ($\Delta\theta_i$) and the topic-specific preferences ($\Delta\zeta_i$) and hence represents the amount of a respondent’s systematic preference change that is due to exposure to the discussion. In our application, since participants are randomly assigned to tables, we can state that any relationship between group composition and respondents’ preferences we observe is caused by interpersonal interactions, rather than due to confounding, omitted variables or homophily.¹⁸

Because we estimate this model for multiple items simultaneously, and since the policy items contain an underlying latent structure, we are able to decompose ω_{ik} into two components, shown in Equation 6b as a random effect that varies across individuals, $\Delta\theta_i$, and a second random effect that varies across both individuals and policies $\Delta\zeta_{ik}$. $\Delta\theta_i$ is a random-effect parameter nested jointly within the full set of policy items and hence captures a systematic shift in preferences along the underlying latent, liberal–conservative dimension that structures preferences across topics. $\Delta\zeta_{ik}$ is specific to each policy item and captures dependence in the preference changes among participants seated at a table for that item, net of the latent component.

Since we define *persuasion* as the component of pre-post preference change that is due to interpersonal interactions, our interests lie in modeling variation across individuals and policies in ω_{ik} and hence variation in $\Delta\theta_i$ and $\Delta\zeta_{ik}$. We model these two dimensions separately. We define $\Delta\theta_i$ in Equation 7a as a normally distributed random effect with conditional mean $\Delta\theta_i^*$ and variance equal to one.¹⁹

$$\Delta\theta_i \sim \phi(\Delta\theta_i^*, 1), \tag{7a}$$

$$\begin{aligned} \Delta\theta_i^* = & \alpha_1 H_i + (\delta_1 \cdot \text{Liberal}_i + \delta_2 + \delta_3 \cdot \text{Conservative}_i) \cdot H_i^2 \\ & + (\gamma_1 \cdot \text{Liberal}_i + \gamma_2 + \gamma_3 \cdot \text{Conservative}_i) \cdot S_i \\ & + \kappa_1 \cdot \text{Liberal}_i + \kappa_2 \cdot \text{Conservative}_i. \end{aligned} \tag{7b}$$

We model the conditional mean for $\Delta\theta_i$ in Equation 7b as a function of the latent ideal points of others seated at the respondent’s discussion table (H and S , defined next) as well as the respondent’s own ideology (liberal, moderate or conservative). Equation 7b contains four distinct variables. To create the Liberal_i and Conservative_i variables, we retrieve the pre-treatment ideal point for each participant and trichotomize this scale into three equally sized groups. H_i is defined in Equation 8a as the estimated mean of the pre-discussion ideal points of the discussants seated at

¹⁶We demonstrate in a separate analysis that there is a one-factor solution for this set of items, in which the first and last items had negative loadings and the other two positive; results not reported.

¹⁷We estimate the components of ω_{ik} using a non-linear spatial auto-regression model, as described in Congdon (2003, chapter 7).

¹⁸As we discuss below, the ω_i parameter captures any within-group dependence, and hence one must be careful in the study design not to introduce confounding group-specific interventions or influences that some groups are exposed to but not others.

¹⁹In an ordered-logit model, the scale of the linear index is not identified and hence we must set this variance parameter to a constant. In other applications this variance should be estimated.

i 's table, excluding i 's own ideal point. S_i is the variance of the ideal points of the other discussants at i 's table, again not including i 's own ideal point. These functions of ideal point estimates, H_i and S_i , are estimated dynamically within the structural equation model.

$$H_i = \text{mean}(\theta_{ij}^0), \quad (8a)$$

$$S_i = \text{mean}([\theta_{ij}^0]^2) - \text{mean}(\theta_{ij}^0)^2, \quad (8b)$$

$$\theta_{ij}^0 \in \{\theta_j^0 : j \text{ is seated at } i\text{'s table, } j \neq i\}. \quad (8c)$$

In Equation 8c, j indexes i 's discussion partners, and the two mean functions are:

$$\text{mean}(\theta_{ij}^0) = \sum_j (\theta_{ij}^0)/(N_i^-), \quad (9a)$$

$$\text{mean}([\theta_{ij}^0]^2) = \sum_j ([\theta_{ij}^0]^2)/(N_i^-). \quad (9b)$$

N_i^- is the number of participants sitting at i 's table, not including i

The parameterization and functional form of Equation 7b are designed to test substantive hypotheses from the literature on small-group dynamics. We have labeled each set of parameters with a different Greek letter (α , δ or γ) to indicate the hypothesis for each set of parameters tests. The parameter α_1 estimates the degree to which person i 's preferences depend on the ideal point composition of others seated at her table, which is the instrument for discussion using the pre-treatment ideological orientation of the participants seated at the respondent's table (Farrar et al. 2009; Gastil et al. 2008; Klar 2014). As Farrar et al. (2009) notes, since respondents' ideological ideal points are measured pre-treatment, we can take these as exogenous, and since group compositions are randomly assigned, we can take effects of this exposure to this measure of group composition to be causal.

The signs for the parameters δ_1 and δ_3 test whether polarization (Furnham, Simmons and McClelland 2000; Isenberg 1986; Schkade, Sunstein and Hastie 2010; Sunstein 2002; Sunstein 2008) is evident in the respondents' latent-space persuasion, separately for liberals and conservatives; we include δ_2 (corresponding to moderates) for completeness and we do not have expectations for its sign. To state expectations for the signs of δ_1 and δ_3 , note that we code the policy-preference items so that high values indicate a conservative response and low values indicate liberal, so higher scores on the latent ideal point scale indicate a conservative leaning. If liberals become more liberal, as the table becomes more liberal, then under a law of group polarization δ_1 should be negative as this would indicate that as a liberal respondent's table becomes more liberal, her latent preferences will become polarized and even more liberal (see the hypothetical curve in the left panel of Figure 1). The patterns should be symmetric for conservatives and so under polarization δ_3 should be positive. If polarization is not evident, then these parameters will not differ from zero. We note that the empirical deliberation literature proposes that structured deliberation inoculates groups from polarization (Barabas 2004; Gerber et al. 2016; Grönlund, Herne and Setälä 2015; Klar 2014; Luskin, Fishkin and Hahn 2007), and hence do not expect small-group polarization to emerge in this context.

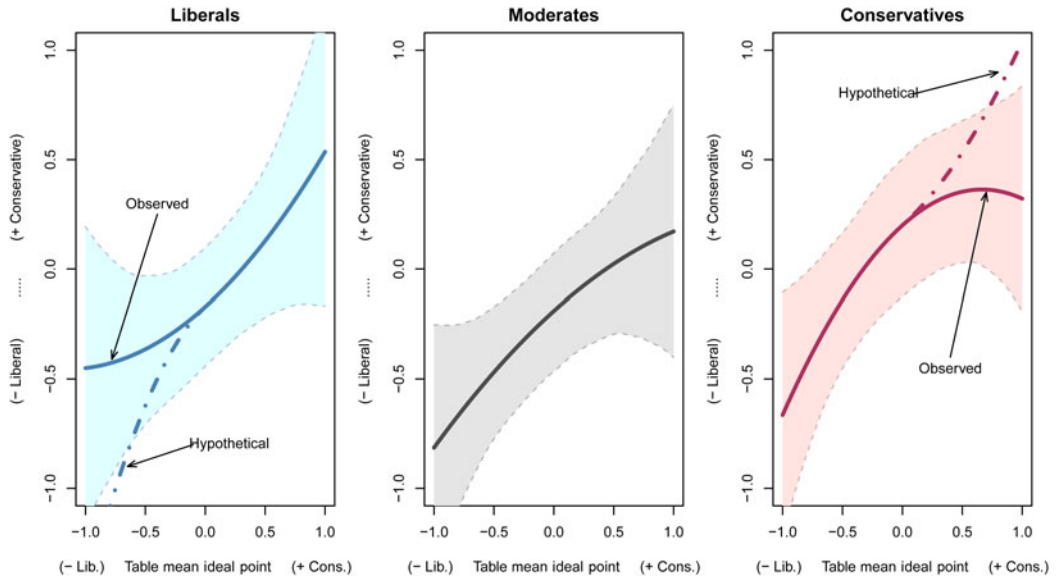


Figure 1. Latent persuasion: no evidence of polarization
 Note: If the ‘law’ of small-group polarization held true, then we would expect to see liberals becoming even more liberal as the table grew more liberal (a concave pattern) and vice versa for conservatives (a convex pattern). Instead we observe a linear relationship or diminishing returns, which is consistent with a mechanism of persuasive arguments within cross-cutting discourse. The confidence bands indicate 95 per cent highest posterior density intervals.

The parameters γ_1 , γ_2 and γ_3 test whether the dispersion of ideal points at a table – a pre-treatment measure of the extent of disagreement among the discussion participants – affects preference change separately for liberals, moderates and conservatives. While we do not have strong priors regarding the direction of this dynamic, it is possible that as the group becomes more divided (as the standard deviation of ideal points increases) participants will tend to selectively attend to the arguments that match their predispositions (see, for example, Bolsen, Druckman and Cook 2014; Edwards and Smith 1996; McGarty *et al.* 1994; Nyhan and Reier 2010; Tabor and Lodge 2006) and hence increase the within-group polarization. In this case, γ_1 should be negative, γ_3 , should be positive, and we have no prior expectations for γ_2 .

The second, policy-specific component of persuasion, $\Delta\zeta_{ik}$, is defined in Equation 10a as a normally distributed random effect with mean $\Delta\zeta_{ik}^*$ and variance one. $\Delta\zeta_{ik}^*$ is a function of the respondent’s own ideology and the policy-specific random effects $\Delta\zeta_{jk}$ of the other participants who are seated at i ’s table. Nesting this random effect within the participants of a given table enables us to assess the extent of dependence in the preference changes on the specific policy topic among table co-discussants, after netting out the covariates in the model as well as $\Delta\theta_i$. Methodologically, this random effect accommodates remaining spatial dependence within clusters (Congdon 2003, chapter 7).

$$\Delta\zeta_{ik} \sim \phi(\Delta\zeta_{ik}^*, 1), \tag{10a}$$

$$\Delta\zeta_{ik}^* = (\rho_{1k} \cdot \text{Liberal}_i + \rho_{2k} + \rho_{3k} \cdot \text{Conservative}_i) \cdot \text{mean}(\Delta\zeta_{ijk}). \tag{10b}$$

$$\Delta\zeta_{ijk} \in \{\Delta\zeta_{jk} : j \text{ is seated at } i\text{'s table, } j \neq i\}, \tag{10c}$$

where

$$\text{mean}(\Delta\zeta_{ijk}) = \sum_j (\Delta\zeta_{ijk}) / (N_i^-). \quad (11)$$

The parameters ρ_{1k} , ρ_{2k} and ρ_{3k} estimate the degree of dependence on each policy preference item among table participants for liberals, moderates and conservatives (respectively) after netting out each respondent's pre-treatment preference, her own ideal point, both before and after the discussion. If a ρ_{2k} is positive and significant, this indicates that if everyone else at the table has a shift in their expected post-test preference on policy k , then person i also can be expected to have a shift in the same direction on issue k ; conversely, if everyone else's preferences stay put, so does person i 's. (Negative rhos are very unusual in this type of model.) If this dependence is net of ideology, then ρ_1 and ρ_3 should test to zero.

We assert that the two components of ω_{ik} ($\Delta\theta_i$ and $\Delta\zeta_{ik}$) capture spatial dependence that comes from the respondents' exposure to her co-discussants. In particular, $\Delta\theta_i$ measures the extent to which the respondent's preferences change along a latent ideological dimension that results from exposure to discussion groups of varying preference compositions; $\Delta\zeta_{ik}$ measures the extent to which a respondent's post-discussion preferences are dependent on her co-discussant's post-discussion preferences on that specific topic for any other reasons. In both of these ways, the model measures dependence that comes from interpersonal interactions.

These twelve structural parameters, α_1 , δ , γ , κ , and ρ capture the effects of exposure to small-group discussion partners on persuasion within the small group, with each set of parameters evaluating the specific mechanisms for persuasion for both latent persuasion, measured by $\Delta\theta_i$, and for topic-specific persuasion for each policy, measured by $\Delta\zeta_{ik}$.

Interpretation and Assumptions

We can take exposure to the discussion group composition as a causal effect provided that the standard assumptions for identifying causal effects within randomized control trials are met (see Angrist, Imbens and Rubin 1996; Gerber and Green 2012). The first assumption is randomization, which is met by the study design in that the event organizers used a random assignment procedure to assign table numbers, and because the number of participants at each table was fixed; a participant could not reassign herself to a different table (Appendix Section A.4 describes an extensive randomization check and balance tests for the table assignments).

The second assumption is the stable unit treatment value assumption, which has two requirements: there is no communication across tables and no alternate versions of the treatment. The assumption of no communication across tables is somewhat strong for our application because tables were adjacent to each other, but one important design feature was that the tables were round, and as a physical configuration of the discussion space for each group the round shape strongly tended to focus discussion within a table and discouraged communication across tables. In addition, with hundreds of people in the event room, the discussion at other tables was mostly background noise. The assumption of no alternate versions of the treatment is met since no information relevant to the decision was introduced by a third party to some discussion groups during the discussion, but not to others. Otherwise, this information could create a group-specific dependence that would confound the effect of interpersonal interactions. The final assumption is the exclusion restriction, which requires that the random assignment process itself does not influence respondents' policy preferences other than through the group composition. This assumption is not testable, but it is difficult to think of ways that our random assignment procedures would have any direct effect on preferences.

Our proposed measurement of persuasion does not generalize to non-randomly assigned small groups or social networks, since the randomized control trial assumptions are unlikely to hold in

these situations. In naturally occurring discussion groups or networks, within-group dependence can occur due to confounding or homophily in addition to any influences from interpersonal interactions.

RESULTS

We estimate the model in OpenBUGS using Bayesian MCMC methods (Lunn *et al.* 2009) and provide details in Appendix Section A.8. We report the estimates for latent-space persuasion in Figure 1 (using $\hat{\alpha}_1$, $\hat{\delta}_1$, $\hat{\delta}_2$ and $\hat{\delta}_3$) to estimate the degree of persuasion conditionally on the mean of the group ideal points, separately for liberals, moderates and conservatives. Figure 2 shows the effect of (pre-test measured) ideological diversity on latent persuasion (estimated by $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{\gamma}_3$). The results for topic-specific persuasion ($\hat{\rho}_{1k}$, $\hat{\rho}_{2k}$ and $\hat{\rho}_{3k}$ for each of the six outcomes) are in Figure 3.

Latent-Space Persuasion

The curves in Figure 1, moving from left to right, show the effect of increasing the proportion of the participant's co-discussants that have ideal points at the conservative end of the latent preference scale, H_p , on the participant's change in the latent space ($\Delta\theta_i$). The middle panel of the figure (moderates) shows that $\hat{\alpha}_1$ is positive, substantively quite large and statistically significant, indicating that moderates' latent preferences respond to exposure to the mix of arguments they hear in the discussion. The left-hand (liberals) panel indicates that $\hat{\delta}_1$ is relatively small, positive in sign and not significantly different from zero, and the right-hand panel (conservatives) shows that $\hat{\delta}_2$ is small, negative in sign and also not significant.

These results for both liberals and conservatives are consistent with a linear pattern or even (by their point estimates) a diminishing return response to the table's ideal point composition when moving in the direction of the respondent's own ideological leaning. For example, as a table grows more conservative in composition, all participants tend to move in the conservative direction on the latent preference scale; but the right-hand panel shows that conservatives themselves do not become especially more conservative; this pattern is symmetric for liberals.

Given these patterns, we do not observe ideological polarization within these small groups, findings that are similar to Barabas (2004), Gerber *et al.* (2016) and Grönlund, Herne and Setälä (2015), who show that deliberative institutions can inoculate small groups against polarization. Under a law of polarization (Sunstein 2002) we would expect the curve in the right-hand panel to be convex or upward bending and the curve in the left-hand panel to be concave or downward bending, patterns indicated by the hypothetical (dashed) curves in Figure 1. The figure shows that the effect of latent-space persuasion is large, but similar for each group. That is, assuming that participants' pre-discussion ideal points are a good instrument for the quantity of ideologically informed arguments they make, these results show that the participants are persuaded by fellow co-participants' ideological appeals, but that ideologues are not especially persuaded by co-ideologues to become extreme.

Recall that participants were randomly assigned to tables, and as a result the effects of table composition can be taken as causal persuasion. Under a counter-argument, one might worry that the linear increasing effect we observe is simply driven by a conformity process, in that a liberal seated at a mostly conservative table might simply conform to conservative positions under social pressure and vice versa. We can argue that conformity is not at work, however, in that the respondents filled out their post-test surveys privately as their final activity of the day and they had no reason to reveal their post-test responses to their co-discussants. Thus, participants completed the post-test survey in an environment that lacked social monitoring (for elaboration, see Boster and Cruz 2003, 478).

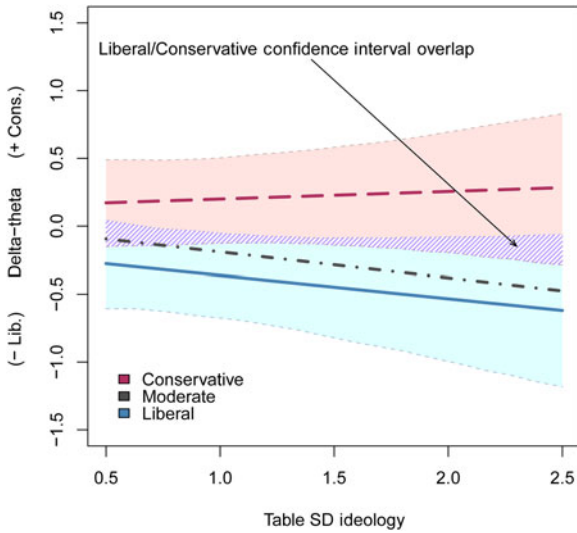


Figure 2. Effect of disagreement on preference direction

Note: As the table becomes more ideologically diverse, ideologues tend to reinforce their pre-existing views, although the effects are not statistically significant. The confidence bands indicate 95 per cent highest posterior density intervals (not shown for moderates).

One might also counter-argue that the diminishing effect we observe is due to a ceiling effect, in that liberals and conservatives might already be located near the endpoints of the latent preference scale with little additional room to move. This concern is mitigated in that, as we demonstrate in Appendix Section A.2, the distribution of ideal points follows a normal distribution so there are very few respondents who are located near the endpoint of the scale. Indeed, only 8.4 per cent of liberals chose the lowest category for each pre-test preference item, and no conservatives chose the highest category for each.

Within-Group Polarization

In addition to the mean ideal point of the group, the statistical model for latent persuasion also includes a second function that characterizes the dispersion of ideal points within each table: the standard deviation of pre-discussion ideal points among participants at each table. This function is an instrument for the diversity of viewpoints available at a given table. Participants might respond to diverse viewpoints by combining those views and so provide a response on the post-test survey that is closer to the center (Druckman and Nelson 2003). Alternatively, participants might use motivated reasoning to selectively attend to the arguments that tend to support their own preconceptions (see, for example, Bolsen, Druckman and Cook 2014; Edwards and Smith 1996; McGarty et al. 1994; Nyhan and Reier 2010; Tabor and Lodge 2006) and so increase in their polarization through a form of confirmation bias. We do not have strong prior expectations regarding either of these patterns.

In the model, the γ parameters test for any effect from a diversity of viewpoints at a table, evaluating the effect of increased ideological diversity on liberals, moderates and conservatives. Figure 2 shows the results. Considering first the point estimates, we find that with a greater diversity of views, liberals (and moderates) tend to become more liberal, while conservatives show no change.

These point estimates suggest that diversity among discussants – rather than ideological homogeneity – may increase polarization in a deliberative context. We note, however, that these point estimates are not statistically different from each other at standard levels. Thus the evidence for polarization from high levels of within-group disagreement at this event is relatively weak.

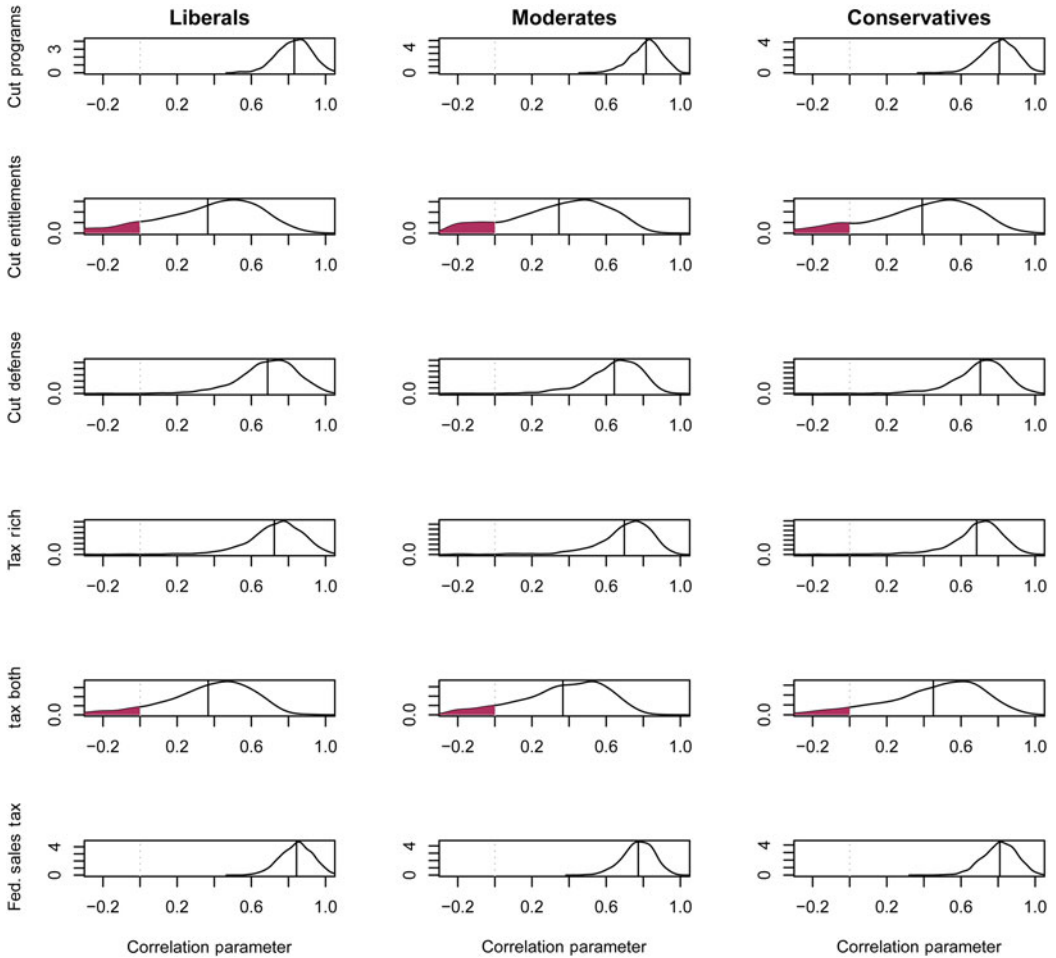


Figure 3. Topic-specific persuasion
 Note: This figure shows the posterior distributions for the ρ_k correlation parameters, which test for spatial dependence in respondents' changes in topic-specific preferences for each item. Note that the dependence is identical for liberals, moderates and conservatives across all items.

Topic-Specific Persuasion

Statements made in deliberation need not be constrained by any heuristic such as left-right ideology (Gutmann and Thompson 1996, 56; Habermas 1984, 99). We are able to assess the amount of persuasion that occurs outside the constraints of the latent preference scale in small-group discussions by examining the degree of dependence of respondents' post-treatment topic-specific preferences ($\Delta\zeta_{ik}$) within a group on each policy preference item.

Figure 3 shows the estimates of the $\rho_{k,k}$ correlation parameters assessing the degree of dependence in the topic-specific preference changes among table co-participants, separately by the ideology of the participant and the item. Overall, the figure indicates a very strong dependence of topic-specific preferences within tables since the $\rho_{k,k}$ parameters are large and significantly different from zero for the cut social programs, cut defense, tax rich and federal sales tax items, and the probability that ρ is different from zero is very large for the cut entitlements and both tax items. Remembering that assignment to tables is random, these results make a strong case for the existence of topic-specific persuasion.

Figure 3 shows that participants' topic-specific preferences are responsive to interactions that occur within the small-group discussions, and since the dependence is uniform between liberals, moderates and conservatives, we show that these preference changes are unrelated to the participant's left-right ideology. This finding is consistent with the aspirations of deliberative democracy in that participants appear to be responsive to reasons and rationales regarding policies that go beyond ideological appeals.

Evaluating the Nature of Persuasion

One could reasonably assert that not all opinion persuasion that is caused by interpersonal interactions should be labeled rational or deliberative (Habermas 1984). Instead, one might be persuaded by co-participants' arguments based on their personal characteristics rather than the substance of their arguments (Petty and Cacioppo 1986), and this non-deliberative persuasion can also induce dependence among responses that is due to interpersonal interaction at a table. While we do not have objective measures of the quality of discourse at each table (such as Steiner et al. 2004), we can gain a sense of the nature of topic-specific persuasion by examining the correlates of the estimated topic-specific random effects, $E(\hat{\zeta}_{ik})$, both in their direction and in their magnitude. We detail these supplemental tests in Appendix Section A.11.

In short, the only consistently significant correlation with each $\hat{\zeta}_{ij}$ we uncovered, both in direction and magnitude, is the perceived *informativeness* of the discussion. In addition, the signs of the correlations indicate that respondents' perceptions of the informativeness of the discussion covaries with movement toward favoring policies that solved the collective problem of the national debt – that is, toward increasing taxes and toward reducing spending. We find that among those who found the discussion to be informative, liberals tended to move toward conservative policies (cut programs and cut entitlements), conservatives moved toward favoring a liberal policy (tax rich), and both liberals and conservatives moved toward favoring the policies that do not load on the latent scale (tax middle class and the rich, and the federal sales tax).

These results are consistent with deliberative aspirations, in that participants who moderated their positions toward the collective goal of solving the debt crisis also perceived the discussion to be informative. While self-perceptions of informativeness do not measure the objective amount of rationality in discourse (Gerber et al. 2016), the correlation establishes the participants' subjective beliefs about the merits of the discussion, which in turn are likely to influence their views of the legitimacy of the event (Cohen 1989).

Missing Data Sensitivity Analysis and Full Replication Study

The online appendix provides analyses that examine the robustness and external validity of these results. First, Appendix Section A.12 analyzes the sensitivity of our findings to different assumptions regarding missing pre-test responses.

Secondly, Appendix Section A.13 reports the results of a replication study that uses data collected from a separate event to test the external validity of the causal findings we report in this article. The data come from the 2007 California *Speaks* health care policy event that was also conducted by America *Speaks*. These data are useful as an external validity test in that the 2007 and 2010 events were substantively very similar in design, but the 2007 study (1) took place three years earlier, (2) was conducted entirely within the state of California rather than nationally, (3) relied partially on randomized survey methods for recruitment and (4) was on health policy rather than fiscal policy. We show that all of the causal results hold up under the replication, including the inconsistency of the observed preference changes with any law of small-group polarization.

Discussion of the Application

The vast bulk of the social psychology literature on small-group discussion finds that small groups tend to polarize to ideological extremes (Isenberg 1986), and hence suggests that human interaction is largely incapable of measuring up to the rational ideals of deliberative democracy (Sunstein 2002). Our findings starkly contrast with this body of work in two ways. First, we do not observe a tendency toward polarization within our deliberative groups, a finding that is robust to replication. Secondly, participants in the aggregate report higher perceptions of the informedness of the discussion when their preferences tend to move toward policy views that are the opposite of their initial ideological predispositions.

As Simone Chambers (2018) notes, group interaction is a necessary but not sufficient condition for human reasoning, and discussion within small groups is likely to be more rational and constructive when the institutional setting is carefully designed to induce deliberative exchanges (Barabas 2004; Grönlund, Herne and Setälä 2015). In Kahneman's 2011 framework, the group context must be designed to trigger System II reasoning, which is deliberate, effortful and analytical, rather than System I reasoning which is based on immediate intuitions. In the *AmericaSpeaks* event, the organizers recruited diverse participants and used random assignment to ensure the groups reflected this diversity; they focused the events on a single topic and provided balanced factual reading materials; they established norms to govern the conversations; and they provided trained moderators for each discussion table who were instructed to facilitate the conversation but not interject their own opinions. In contrast to the minimal designs of most small-group research, these elements are likely the reason we observe persuasion that is consistent with effortful reasoning (Fishkin 2018; Neblo, Esterling and Lazer 2018).

While we did not vary the institutional setting in order to conduct a comparative institutional analysis of small-group design on persuasion, we argue the analogy holds in the comparison between two recent articles on persuasion in the context of voting. On the one hand, Kalla and Broockman (2018) find that in the context of an election campaign, competing frames and partisanship diminish candidates' ability to persuade voters; they find persuasion effects to be near zero for campaign contacts and advertisements in this setting. On the other hand, Broockman and Kalla (2016) find that, outside of a major election, a brief conversation in which a canvasser encourages a voter to actively take the perspective of others can durably reduce transphobia. The authors argue that, in contrast to the typical election campaign contacts, conversation that is centered on perspective taking induces System II reasoning, which in turn can help to bridge a persistent societal divide.

CONCLUSION

This article develops a novel measurement strategy to evaluate persuasion within small groups, which we apply to a large-scale deliberative town hall. Thus we hope to make both methodological and substantive contributions to the literature on persuasion in small-group processes.

Methodologically, we wish to underscore the importance of measurement when testing hypotheses about persuasion. Our methods help to focus the statistical test on the systematic components of preference change that are due to interpersonal interactions, rather than the total variance in preference change that includes some unknown random or noise component. This explicit focus on measurement also allows us to identify substantively important dimensions of persuasion. In our case the distinction between latent-space and topic-specific persuasion is important to understand the full dynamics of deliberative interaction. The methods we propose are very general and can be applied to any small-group interaction where participants have been randomly assigned to small groups (as in Farrar *et al.* 2009).

Substantively, in our application, we find that the persuasion we observe fulfilled many of the normative aspirations for deliberative democracy. Participants are responsive to their co-discussants' ideological appeals, but within the deliberative setting we do not observe a

tendency toward ideological polarization (in contrast to Sunstein 2002). In addition, we find that liberals and conservatives tend to be responsive to non-ideological appeals, which we label ‘topic-specific persuasion’, and that the extent of topic-specific persuasion covaries with the participant’s perception that the discussion was informative. We also find that the correlation between informativeness and persuasion was most evident for liberals on conservative policies and conservatives on liberal policies. Given the polarized nature of contemporary political discourse, particularly on national fiscal matters, we believe that reinforcing deliberative institutions might prove an effective way to address many of our pressing common problems.

Supplementary material. Data replication sets are available in Harvard Dataverse at: <https://doi.org/10.7910/DVN/XDDN2Y> and online appendices are available at <https://doi.org/10.1017/S0007123419000243>.

Acknowledgements. This project was funded by a very generous grant from the John D. and Catherine T. MacArthur Foundation. We thank Jason Barabas, Peter Esaiasson, Thomas Leeper, Robert Lupton, John Ryan, Georgia Warnke, and seminar participants at the Center for the Study of Stratification and Inequality at Tohoku University, and at UC Center Sacramento, for very valuable comments. We gratefully acknowledge the collaboration with and support from Carolyn Lukensmeyer, Joe Goldman, and all of the staff who worked at *AmericaSpeaks*.

References

- Achen CH (1975) Mass political attitudes and the survey response. *American Political Science Review* 69(4), 1218–1231.
- Angrist JD, Imbens GW and Rubin DB (1996) Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434), 444–455.
- Anselin L (1988) *Spatial Econometrics: Methods and Models*. Boston, MA: Kluwer Academic.
- Ansolabehere S, Rodden J and Snyder JM (2008) The strength of issues: using multiple measures to gauge preference stability, ideological constraint, and issue voting. *American Political Science Review* 102(2), 215–232.
- Barabas J (2004) How deliberation affects public opinion. *American Political Science Review* 98(4), 687–701.
- Bolsen T, Druckman JN and Cook FL (2014) The influence of partisan motivated reasoning on public opinion. *Political Behavior* 36, 235–262.
- Boster FJ and Cruz MG (2003) Persuading. In Dill J (ed.), *The Persuasion Handbook: Development in Theory and Practice*. London: Sage Publications, pp. 477–494.
- Broockman D and Kalla J (2016) Durably reducing transphobia: a field experiment on door-to-door canvassing. *Science* 352(6282), 220–224.
- Chambers S (2018) Human life is group life: deliberative democracy for realists. *Critical Review* 30(1–2), 36–48.
- Clinton JD (2012) Using roll call estimates to test models of politics. *Annual Review of Political Science* 15, 79–99.
- Cohen J (1989) Deliberation and democratic legitimacy. In Hamlin A and Pettit P (eds), *The Good Polity: Normative Analysis of the State*. New York: Basil Blackwell, pp. 17–34.
- Congdon P (2003) *Applied Bayesian Modelling*. Hoboken, NJ: John Wiley & Sons, Ltd.
- Druckman JN and Nelson KR (2003) Framing and deliberation: how citizens’ conversations limit elite influence. *American Journal of Political Science* 47(4), 729–745.
- Eatwell R (1993) Introduction: what are political ideologies? In Eatwell R and Wright A (eds), *Contemporary Political Ideologies*, London: Pinter, pp. 1–20.
- Edwards K and Smith EE (1996) A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology* 71(1), 5–24.
- Esterling K, Fung A and Lee T (2019), “Replication Data for: When Deliberation Produces Persuasion Rather Than Polarization: Measuring and Modeling Small Group Dynamics in a Field Experiment”, <https://doi.org/10.7910/DVN/XDDN2Y>, Harvard Dataverse, V1, UNF:6:4JvR2VKXLFLnJIM7lHdpzQ== [fileUNF]
- Farrar C et al. (2009) Does discussion group composition affect policy preferences? Results from three randomized experiments. *Political Psychology* 30(4), 615–647.
- Feldman S and Johnson C (2014) Understanding the determinants of political ideology: implications of structural complexity. *Political Psychology* 35(3), 337–358.
- Fishkin JS (2018) *Democracy When the People are Thinking: Revitalizing Our Politics Through Public Deliberation*. New York: Oxford University Press.
- Fishkin JS and Luskin RC (2005) Experimenting with a democratic ideal: deliberative polling and public opinion. *Acta Politica* 40(3), 284–298.
- Fowler JH et al. (2011) Causality in political networks. *American Politics Research* 39(2), 437–480.
- Furnham A, Simmons K and McClelland A (2000) Decisions concerning the allocation of scarce medical resources. *Journal of Social Behavior and Personality* 15(2), 185–200.

- Gastil J, Black L and Moscovitz K** (2008) Ideology, attitude change, and deliberation in small face-to-face groups. *Political Communication* 25, 23–46.
- Gerber AS and Green DP** (2012) *Field Experiments: Design, Analysis and Interpretation*. New York: W.W. Norton.
- Gerber M et al.** (2016) Deliberative abilities and influence in a transnational deliberative poll (EuroPolis). *British Journal of Political Science* 48, 1093–1118.
- Grönlund K, Herne K and Setälä M** (2015) Does enclave deliberation polarize opinions? *Political Behavior* 37, 995–1020.
- Gutmann A and D Thompson** (1996) *Democracy and Disagreement: Why Moral Conflict Cannot be Avoided in Politics, and What Should be Done about It*. Princeton, NJ: Princeton University Press.
- Habermas J** (1984) *The Theory of Communicative Action, Volume One: Reason and the Rationalization of Society* Boston, MA: Beacon Press.
- Hinich M and Munger MC** (1994) *Ideology and the Theory of Political Choice*. Ann Arbor: University of Michigan Press.
- Imai K et al.** (2011) Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies. *American Political Science Review* 105(4), 765–789.
- Isenberg DJ** (1986) Group polarization: a critical review and meta-analysis. *Journal of Personality and Social Psychology* 50(6), 1141–1151.
- Kahneman D** (2011) *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kalla JL and Broockman DE** (2018) The minimal persuasive effects of campaign contact in general elections: evidence from 49 field experiments. *American Political Science Review* 112(1), 148–166.
- Karpowitz CF and Mendelberg T** (2007) Groups and deliberation. *Swiss Political Science Review* 13(4), 645–662.
- Klar S** (2014) Partisanship in a social setting. *American Journal of Political Science* 58(3), 687–704.
- Lauderdale BE, Hanretty C and Vivyan N** (2018) Decomposing public opinion variation into ideology, idiosyncrasy, and instability. *Journal of Politics* 80(2), 707–712.
- Lunn D et al.** (2009) The BUGS project: evolution, critique, and future directions. *Statistics in Medicine* 28, 3049–3067.
- Luskin RC, Fishkin JS and Hahn KS** (2007) Consensus and Polarization in Small Group Deliberation. Technical report, Presentation at the Midwest Political Science Association, Chicago, Ill. Available at <https://www.researchgate.net/publication/253165745>.
- Luskin RC, Fishkin JS and Jowell R** (2002) Considered opinions: deliberative polling in Britain. *British Journal of Political Science* 32, 455–487.
- McGarty C et al.** (1994) The effects of salient group memberships on persuasion. *Small Group Research* 25(2), 267–293.
- Minozzi W et al.** (2015) Field experiment evidence of substantive, attributional, and behavioral persuasion by members of Congress in online town halls. *Proceedings of the National Academy of Sciences* 112(13), 3937–3942.
- Mutz DC, Sniderman PM and Brody RA** (1996) Political persuasion: the birth of a field of study. In Mutz DC, Sniderman PM, and Brody RA (eds), *Political Persuasion and Attitude Change*. Ann Arbor: University of Michigan Press, pp. 1–14.
- Neblo MA, Esterling KM and Lazer DM** (2018) *Politics with the People: Building a Directly Representative Democracy*. New York: Cambridge University Press.
- Nyhan B and Reifler J** (2010) When corrections fail: the persistence of political misperceptions. *Political Behavior* 32(2), 303–330.
- Petty RE and Cacioppo JT** (1986) The elaboration likelihood model of persuasion. In Berkowitz L (ed.), *Advances in Experimental Social Psychology*, Vol. 19. Cambridge, MA: Academic Press, pp. 123–205.
- Poole KT and Rosenthal H** (1997) *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.
- Prior M** (2010) You've either got it or you don't? The stability of political interest over the life cycle. *Journal of Politics* 72(3), 747–766.
- Schkade D, Sunstein CR and Hastie R** (2010) When deliberation produces extremism. *Critical Review* 22(2–3), 227–252.
- Skrondal A and Rabe-Hesketh S** (2004) *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman and Hall.
- Steiner J et al.** (2004) *Deliberative Politics in Action: Analysing Parliamentary Discourse*. New York: Cambridge University Press.
- Sunstein CR** (2002) The law of group polarization. *Journal of Political Philosophy* 10(2), 175–195.
- Sunstein CR** (2008) *Republic.com 2.0*. Princeton, NJ: Princeton University Press.
- Tabor CS and Lodge M** (2006) Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science* 50(3), 755–769.
- Tausanovitch C and Warshaw C** (2017) Estimating candidates' political orientation in a polarized Congress. *Political Analysis* 25(2), 167–187.
- Treier S and Hillygus DS** (2009) The nature of political ideology in the contemporary electorate. *Public Opinion Quarterly* 73(4), 679–703.
- Treier S and Jackman S** (2013) Democracy as a latent variable. *American Journal of Political Science* 52(1), 201–217.
- Westwood SJ** (2015) The role of persuasion in deliberative opinion change. *Political Communication* 32(4), 509–528.

Cite this article: Esterling KM, Fung A, Lee T (2021). When Deliberation Produces Persuasion rather than Polarization: Measuring and modeling Small Group Dynamics in a Field Experiment. *British Journal of Political Science* 51, 666–684. <https://doi.org/10.1017/S0007123419000243>