Testing the Benefits of Public Deliberation

William Minozzi, The Ohio State University
Ryan Kennedy, University of Houston
Kevin M. Esterling, University of California, Riverside
Michael A. Neblo, The Ohio State University
Ryan Jewell, University of Houston

Word Count: 9685

Abstract. Public deliberation grows increasingly prevalent, yet remains costly in terms of money and time. Accordingly, some suggest supplanting talk-based practices with individual, "deliberation within." Yet we have little evidence either way on the additional value of public deliberation over its individual variant. We evaluate the benefits of public deliberation with a field experiment. With the cooperation of two sitting U.S. Senators, we recruited several hundred of their constituents to deliberate on immigration reform. Participants were randomly assigned to either deliberate publicly in an online discussion, to deliberate individually, or to an information-only control. Across several measures, public deliberation added more value than individual deliberation. We find, moreover, little evidence to ground worries that differences in education, race, conflict avoidance, gender, or gender composition of deliberating groups will render public talk less valuable than individual deliberation.

Verification Materials. The data and materials required to verify the computational reproducibility of the results, procedures and analyses in this article are available on the American Journal of Political Science Dataverse within the Harvard Dataverse Network, at https://doi.org/10.7910/DVN/EGUVG0.

Acknowledgements. Presented at the Annual Meeting of the Midwest Political Science Association, April 4, 2019. We gratefully acknowledge the Congressional Management Foundation, Sen. Thom Tillis, and Sen. Mike Crapo. This project was funded through the generous support of the Templeton Foundation's Intellectual Humility in Public Discourse program, administered by Michael Lynch at the University of Connecticut.

Keywords. deliberative democracy; field experiment; education; race; conflict avoidance; gender

Introduction

Deliberative democratic practices are increasingly prevalent. In addition to being the subject of a vibrant research program (Gastil 2018; Dryzek et al. 2019), deliberation is beginning to have real impact. Governments rely on deliberative bodies to help craft and vet policy (Carson, Gastil, Hartz-Karp, and Lubensky 2013; Mccombs and Reynolds 1999; Warren and Pearse 2008), and deliberative practices deepen the link between lawmakers and constituents (Neblo, Esterling, and Lazer 2018). Participation in deliberative forums yields many salutary effects for participants, including knowledge gains (Esterling, Neblo, and Lazer 2011), increased participation (Gastil, Deess, and Weiser 2002; Minozzi, Neblo, Esterling, and Lazer 2015), invigorated engagement (Delli Carpini, Cook, and Jacobs 2004), reduction in adverse emotional responses (Baek, Wojcieszak, and Delli Carpini 2012), and improvements in trust in government and political efficacy (Gastil, Pierre Deess, Weiser, and Simmons 2010). Notwithstanding important problems and inequities that plague deliberation (Mutz, 2006; Karpowitz and Mendelberg 2014; Sanders 1997; Ugarriza and Nussio 2016), the upside potential for deliberative democracy remains bright (Neblo 2015).

But deliberation is expensive. In-person deliberative events can require participants to travel, arrange child care, and miss work (Fishkin 2011). Such costs have fallen thanks to Internet-based technologies (Papacharissi 2002). Online deliberation not only shrinks costs, it also expands horizons, both geographically (Stromer-Galley 2003) and socially (Wojcieszak and Mutz 2009). But even with online deliberative forums, sessions can be lengthy, participants must still agree on a time to meet and actually follow through, and someone must coordinate these schedules. Consequently, participation can be costly, both to plan and incentivize, requiring cash payments, collaboration with political officials, or both.

Accordingly, scholars have suggested ways to subsidize or supplant deliberative forums. Although deliberation seems quintessentially talk-based, interpersonal, and public, it also requires reflective, intrapersonal, and private action, most importantly weighing reasons. Goodin (2000), who notes this logical necessity, goes on to argue that individual deliberation should supplement public deliberation. As with public deliberation, the costs of informing one's individual deliberations have declined in recent years. The Internet provides a wealth of policy evidence and analysis for the great majority of people in established democracies; almost 90% of the U.S. has broadband Internet access (Anderson, et al. 2018).

Public deliberation also involves downsides that individual deliberation can avoid. Equal participation and mutual respect are the pivotal requirements necessary to achieve the benefits of talk-based practices, yet it is not clear that real-world instances of public deliberation come close to meeting these standards (Sanders 1997). Indeed, there are well known impediments. Psychological differences, such as an aversion to conflict (Ulbig and Funk, 1999; Mutz, 2006), mean that some individuals may be less inclined to participate, and thus not to share equally in the benefits. Differences in education can limit discursive capacity (Ugarizza and Nussio 2016), deepening disparities. And those who wish to contribute—especially women and nonwhite participants—may be made reticent or even prevented from doing so in some contexts (Sanders 1997; Karpowitz and Mendelberg 2014; but see also Hickerson and Gastil 2008).

It is therefore reasonable to ask what the marginal value of specifically public deliberation might be (Goold et al. 2012). Yet we have little empirical evidence on the added value of public deliberation over its individual variant. Goodin (2000, 2003) and Mercier and Landemore (2012) offer strong, though essentially theoretical arguments for the necessity of public deliberation. Previous research has established that public deliberation stimulates internal deliberative change

(Esterling, Neblo, and Lazer 2011; Goodin and Niemeyer 2003). But existing studies measure the effects of public deliberation with comparison either to pretest surveys, uninformed control subjects, or, at best, participants who were provided information, but not actively pushed toward authentic, individual deliberation.

We assay the benefits of public deliberation with a field experiment conducted from August to October 2017. We worked with the Congressional Management Foundation to recruit two sitting U.S. Senators to participate in a study focused on immigration reform.[1] The Senators did not interact directly with their constituents in the deliberative events. Instead, we provided the Senators a report analyzing the results of the sessions, and we made the participants aware in advance we would communicate their responses in this way.

Participants were residents of one of the states represented by the Senators and assigned to one of three conditions. The first two stimulated public and individual deliberation, respectively. The first condition involved participation in an online group discussion with other residents of their state, structured around an issue guide and short videos that discuss policy options and elucidate tradeoffs. The second walked participants through a questionnaire that simulated the online discussion, encouraging individual deliberation by requiring subjects to watch the same videos from the discussions and confront the same tradeoffs. To independently measure the value of both public and individual deliberation, some participants were assigned to a control group. Members of the control group were provided access to the same information and the same videos but could skip this part of the experiment if they wished. All participants responded to surveys after completing these tasks to measure the diffuse and subjective benefits that can attend

[1] The participating lawmakers were Sen. Mike Crapo (R-ID) and Sen. Thom Tillis (R-NC).

participating in effective deliberation, such as their satisfaction with the experience, knowledge gains, affective reactions, and civic attitudes. We then used multilevel models to provide overall estimates of the benefits of public deliberation.

Our experiment and analysis contribute methodologically in the use of state-of-the-art causal inference techniques and multilevel models to average over dozens of survey questions properly accounting for multiple comparison problems. Substantively, moreover, the paper contributes to our understanding of the political psychology of attitude change, in addition to the theory and practice of deliberation. Ultimately, we found that encouraging individual deliberation did offer clear benefits over merely providing access to information. Yet public deliberation produced an overall effect almost double that of individual deliberation, even though our sessions were online rather than face-to-face. Further, we found no evidence that important limits sometimes observed in face-to-face public deliberation—such as disparities in the experience that arise from conflict avoidance or less education, or social dynamics that can arise based on race and gender—reduced these benefits. If these limits were intrinsic to deliberation as a method of communication, the limits should be apparent irrespective of whether the interaction is online or face-to-face. We conclude that the added value of public deliberation persists across many lines of difference.

The Goals and Effects of Individual Deliberation

Among the many conceptions of democratic legitimacy, deliberative democracy (DD) stands out as one of the most ambitious. Advocates argue that, in contrast to majoritarian theories of democracy, DD allows for good and bad policy solutions rather than merely more and less powerful or numerous actors (Elster 1986). And unlike some theories of democracy, DD sets out

standards that its proponents claim reliably produce—and, according to some, are even constitutive of (Habermas 1996; Neblo 2015)—authentically better solutions, rather than vesting authority in a privileged subset of political actors (Brennan 2017) or depending on the statistical properties of crowds (Page 2007; Surowiecki 2004).

Yet these qualities come at a cost. Facially deliberative practices require citizens to gather together to talk, and in so doing search, challenge, tolerate, respect, reflect, revise, and so on. Such practices may seem prohibitively expensive, not only because of limited human cognitive faculties (Taber and Lodge 2006), but in literal terms of time and effort, especially given the size constraints of deliberation relative to the scale of mass democracy (Goodin 2000, 2003).

Despite (and perhaps because of) these costs, deliberation is thought to provide a wide variety of benefits for individual participants, including awareness and tolerance of the "other side" (Mutz 2006), gains in issue-specific knowledge (Esterling, Neblo, and Lazer 2011), awareness of competing rationales (Price, Cappella, and Nir 2002), and revision of policy attitudes (Minozzi, Neblo, Esterling, and Lazer 2015). More generally, deliberating deepens civic engagement (Delli Carpini, Cook, and Jacobs 2004), giving rise to positive affective experiences (Baek, Wojcieszak, and Delli Carpini 2012), and increased trust in government and efficacy (Gastil, Pierre Deess, Weiser, and Simmons 2010).

Even though these deliberative effects have been well established, it is far from clear which part of the process drives the change or is even the ultimate source of its normative appeal. From a theoretical perspective, Goodin (2000) argues that the conversational aspect of deliberation is catalytic rather than constitutive. That is, for Goodin, the act of talking with others stimulates cognitive processes that are themselves the proper goal of deliberation, and, as such, talk is merely a useful way to achieve the ensuing benefits. On this account, we should shift our attention from

"making people 'conversationally present' [to] making them 'imaginatively present'" (83). If we accept this view that the good in deliberation is inherently individual rather than public, the goods from public deliberation are instrumental.

Alternative accounts highlight the explicit role that talk plays, while acknowledging the value of individual deliberation. Mercier and Landemore (2012) and Landemore (2012) argue that, under conditions of diversity, deliberation fosters better collective decision making, not in spite of cognitive limits such as motivated reasoning, but rather because individuals are evolutionarily disposed to make competing arguments (Chambers 2018). Rather than falling prey to "the law of group polarization" (Sunstein 2002), this argumentative theory of reasoning holds that reasons can be generated through biased processes, yet still conduce toward better outcomes, provided that some aggregative mechanism like majority rule is employed. In stronger terms, Neblo (2015) draws on Brandom (1998) and Habermas (1996) to advance an inferentialist theory, according to which the conversational process does not just give rise to better choices, but is actually constitutive of them. But both positions also leave room for individual deliberation, recognizing that the goods provided by deliberation are not limited to behaviors exhibited in conversation.

The Differential Effects of Public Deliberation

Regardless of whether one accepts Goodin's argument that the goods derived from public deliberation are only instrumental, the relationship between public and individual deliberation is intricate. In a careful case study of a citizens' jury, Goodin and Niemeyer (2003) identify the pre-deliberation period as the moment of greatest attitude change. In their study of deliberative town halls, Esterling, Neblo, and Lazer (2011) also identified the preparatory period as the most important component of knowledge gains. Both these studies, however, rely on comparisons of

pre-test and post-test outcomes rather than direct comparison of public and individual deliberation. Further, Goodin and Niemeyer (2003) lack a control group, and Esterling, Neblo, and Lazer (2011) compare public deliberation to mere information provision, rather than encouragement of individual deliberation.

Yet public deliberation may be better at achieving these benefits than individual deliberation. First, as Mercier and Landemore (2012) explain, the process of arguing stimulates cognitive processes that might otherwise lie dormant. In response to challenges, a participant must generate a response, either responding to the challenge or revising her previous position. Through spreading activation (Collins and Loftus 1975), these conversations may yield a broader set of concepts, reasons, and considerations than one can generate individually. The social aspect of public deliberation differentiates it from individual reflection as well, and social settings play an important role in preference formation (Klar 2014). At a more basic level, it can be challenging to motivate interest and engagement in the absence of interpersonal communication (Nystrand and Gamoran 1991). Thus, our overarching hypothesis is that, while both public and individual deliberation plausibly give rise to a host of benefits, public deliberation should be more effective than individual deliberation.

There are also reasons to believe that inducements to deliberate may have differential effects based on characteristics of the participants. Individuals vary in both dispositions and identities that can affect their experiences of deliberation. We are particularly interested in the possibility that public deliberation may be less valuable than individual deliberation for certain individuals or in certain situations. While some of these concerns regarding disparities have been primarily observed in face-to-face deliberation; it is nonetheless important for us to investigate whether these disparities reproduce in an online setting.

First, more conflict avoidant individuals may experience public deliberation in a less positive way. The conflict avoidant are less likely to engage in political discussion generally (Ulbig and Funk 1999), and the consequences of their exposure to disagreement are more likely to include disengagement (Mutz 2006). Consequently, as conflict avoidance increases, people may not be able to share the benefits of public deliberation. The value of individual deliberation, in contrast, should not vary for the conflict avoidant.

Beyond psychological traits, differences in identity may play a role in conditioning the experience of public deliberation. Sanders (1997) suggests that (public) deliberation may not be particularly valuable for members of relatively disempowered groups—including women, non-white, and non-college educated participants. These disparities may arise via multiple mechanisms, both directly through bare expression of power and indirectly via introjection and self-censoring.

The evidence for this difference critique of public deliberation is nuanced. On the one hand, Karpowitz and Mendelberg (2014) find that women do experience deliberation differently from men, but that this experience is dependent on the gender mix of the deliberating groups. Similarly, Ugarriza and Nussio (2016) report that the quality of discourse depends on the education levels of the deliberators. And, all else equal, membership in the majority enhances the influence a deliberator can have (Myers 2017), which might lead members of any minority group to prefer the individual variety. On the other hand, Hickerson and Gastil (2008) find only trace evidence of disparities in the benefits of jury participation. None of these studies, however, focus specifically on the additional value of public deliberation. Therefore, we take it as an open empirical question whether inequities in participation offset some of the benefits of public deliberation.

Hypotheses about the Benefits of Deliberation

Given the known costs associated with public deliberation, we investigate the benefits that result from the interpersonal experience, benefits that go beyond mere exposure to information or even individual reflection on new information. While the benefits we examine are not generally financial, we posit that they *do* enter into respondents' utility function when considering whether participation is worthwhile, and *should* count toward our notion of social utility.

Deliberation requires the participants' attention and engagement. As a result of this engagement, we hypothesize that participants will find the experience to be more helpful and informative, and regard the overall practice to be both useful to lawmakers and important for democracy. To the extent that public deliberation engages participants more successfully than individual deliberation, satisfaction should also be higher.

> ***Attitudes toward Experience Hypothesis.*** Participants will be more satisfied with the experience of public deliberation than individual deliberation, and more satisfied with both these experiences than mere exposure to information.

Deliberative processes also should lead at least some participants to change their views, which in turn might lead them to assess the time spent as meaningful and worthwhile (thus there is some potential overlap between the two concepts). As they weigh reasons and encounter different opinions, individuals who deliberate should sometimes be persuaded (Minozzi, Neblo, Esterling, and Lazer 2015). At its best, the "unforced force of the better argument" (Habermas 1984) should prevail, meaning that these changes will be for the better. These changes might include perceived knowledge gains, declines in confusion, profession of respect, and testament to actual change in attitudes. These subjective changes are important on their own. To the extent that extreme expressive partisanship (Huddy, Mason, and Aarøe 2015) is symptomatic of problems for a democracy, the open expression of change is consistent with democratic health. Again, these

changes are more likely in public deliberation than its individual variant, both because of deepened engagement and according to the argumentative theory of reasoning. And both varieties should dominate changes resulting from mere exposure to information.

> ***Perceptions of Changed Minds Hypothesis.*** Participants will report changing their minds after deliberating, and the changes should be greater for public deliberation than individual deliberation.

For many, deliberation typically entails learning facts that may be integrated into a coherent basis for their policy attitudes. As such, participants in deliberation should know more about the issue at hand than their non-deliberating counterparts, a knowledge gain that in turn has presumptive value to respondents and public discourse. If active learning is more easily achieved in public deliberation than in individual deliberation, gains should be correspondingly greater.

> ***Knowledge Gain Hypothesis.*** Participants will know more about the issue under deliberation than others who do not deliberate, and the gain should be larger for public deliberation than individual deliberation.

Beyond learning facts, deliberation may elicit positive or negative affective responses from participants as a result of thinking about challenging political problems. Anger is a powerful emotion in political action and often moves in distinct ways relative to the surveillance systems developed in theories of affective intelligence (Albertson and Gadarian 2015). Deliberation may reveal that seemingly intractable problems are actually solvable, and that more common ground exists among heterogeneous citizens than might otherwise have been apparent (Wojcieszak and Mutz 2009), thus reducing negative emotions including anger (Feldman et al. 2004). This effect should be particularly acute in public deliberation, which, at its best, yields experiential evidence of both difference and tolerance (Mutz 2006), and may thereby activate individuals' "disposition" systems (e.g., feelings of hope, pride, and enthusiasm, see Marcus, Neuman, and MacKuen 2000). Similarly, discussion of an issue may trigger attention to attendant uncertainty, and in so doing

activate the "surveillance" system associated with anxiety. Neblo (2020) explicitly links these three clustered phenomena to high quality deliberation, and so, to the extent that public deliberation is higher quality than individual deliberation, the emotional effects of public deliberation should be larger.

> ***Affective Hypotheses.*** Deliberation should increase enthusiasm, reduce anger, and increase anxiety; and these effects should be larger for public deliberation than individual deliberation.

Beyond the issue at hand, deliberation may kindle renewed feeling of civic engagement, trust, and efficacy, although there is mixed evidence on this question in the literature (Myers and Mendelberg 2013). For efficacy, the empirical evidence suggests larger effects for external than for internal efficacy (Esterling, Neblo, and Lazer 2011; Hertzum, Andersen, Andersen, and Hansen 2002; Morrell 2005; Nabatchi 2010; Pierce, Neeley, and Budziak 2008), although Gastil and Dillard (1999) found evidence of increase internal efficacy. Elsewhere, Gastil, Pierre Deess, Weiser, and Simmons (2010) found that participation in jury service, an institutionally important species of deliberation, improved citizens' trust in government and political leaders. There is little evidence, however, about the *relative* roles of public and individual deliberation on these questions. But to the extent that improvements in engagement due to public deliberation are substantial, it is reasonable to expect corresponding differences in effects.

> ***Trust and Efficacy Hypotheses.*** Deliberative participants will have increased trust in government, and internal and external efficacy, and these effects will be larger for public than for individual deliberation.

Experimental Design and Methods

We recruited participants through Qualtrics and randomly assigned them to one of three conditions. Our aim was to hold participants' access to information constant while varying the locus of their subsequent mode of engagement with that information. In total, we analyze evidence

from 402 experimental participants, which is similar in size to that from other experimental studies of deliberation. For example, in their landmark study of gender and deliberation, Karpowitz and Mendelberg (2014) report on 470 experimental participants (e.g., p. 109).

Members of the **Public Deliberation** group participated in an online, small-group discussion about immigration reform, conducted through the "Common Ground for Action" (CGA) platform. During the course of their deliberation, members of this treatment group were exposed to information about current immigration policy. The loosely scripted sessions were moderated and walked participants through three approaches to immigration reform: a path to citizenship, focus on border security, and employment regulation. Participants first ranked six policies, two associated with each approach. For each approach, they were prompted to consider pros and cons, and then engage with other discussion group members via a text-based chat. Participants then registered via a graphical interface whether they now favored the policy, and whether they could accept the associated drawbacks. Sessions lasted forty-five minutes to an hour, after which participants were directed to a post-survey. In all, we fielded 39 discussion groups, ranging from two to nine in size, for a total of 198 public deliberators who completed surveys.[2]

[2] Due to vagaries of scheduling, eight additional participants attended solo sessions that only included the moderator, for a group size of 1. We omit these observations from our analysis. Including them does not change the results because there are so few of these observations. Further, we identified three individuals who participated in two sessions; we dropped the second observation for these individuals. Finally, 32 individuals who participated in discussion groups did not click through to complete the survey, and therefore do not appear in our analysis.

The first of the other two treatment groups provides the main rival to the **Public Deliberation** group. Members of the **Individual Deliberation** group participated in a replica of each stage of the CGA platform. Members of this group were given access to the same background information and, in our key manipulation, were walked through a survey-based simulacrum of the CGA platform in a manner that frustrated satisficing (e.g., by requiring a certain amount of time before advancing screens, etc.), but without access to small group discussion. This treatment condition was conducted solely through an Internet-based survey.

Finally, members of the **Control** group were also given access to all the information, and prompted to answer the same questions, as the other groups. However, these participants could also easily skip the readings and videos and move onto the subsequent questionnaire. This opens the opportunity for satisficing behavior on the part of the respondent. An alternative way to view the manipulation is that the **Control** group was merely given the opportunity to take the treatment, while the **Individual Deliberation** group was "required" to.[3] By constructing the experiment in this way, we effectively simulate ambient conditions, in which a wealth of policy relevant information is easily available over the Internet. Incentives for accessing, processing and reflecting on that information are the key causal forces of interest. We recruited 102 participants each for the **Individual Deliberation** and **Control** groups.[4]

[3] Of course, we did not literally *force* or require participants to do anything, and they could drop out of the study, as per our IRB protocol.

[4] As always, our sample size is only sufficient to yield reliable estimates of our outcomes provided that they are of a minimum detectable effect (MDE) size. In the Appendix (p. A17), we calculate MDEs that would yield 80% power for each of our outcomes and treatments. In general, these

Our hypotheses focus on nine outcomes. We administered a survey immediately after the session, including 41 questions that are nested in the nine categories that we take as outcomes.[5] As we explain below, we use these repeated measures of the nine substantive outcomes in a multilevel model. The model nests questions within the question-categories that are of substantive interest to us, and estimates the treatment effects as (semi-pooled) random coefficients across the categories to improve efficiency of the estimates while guarding against multiple comparisons problems (Gelman, Hill and Yajima 2012). We scaled all survey item responses to lie between 0 and 1.

First, to measure participants' *Attitudes toward the Session*, we asked six questions about their experience. These measures capture general satisfaction with the online session: whether participants think it was worthwhile enough to participate in the activities again, whether their member of Congress should pay attention to the resulting outcomes from the session, future interest in participation, and more general beliefs about deliberation. Since we focus these questions on the session the questions are interpretable for respondents in each of the three experimental conditions. Second, we measure *Perceptions of Changed Minds*, broadly understood, with eight items. Here, we seek to measure whether participants perceived themselves to learn, gain clarity and reduce confusion, develop respect for the opposition, and change policy positions and rationales. The subjective perception of changing one's mind suggests that the respondent found the session constructive and worthwhile, and that they gained novel information. Third, we

MDEs are on the order of 3% to 10%, depending on the outcome and treatment, and they are larger for the Individual condition than the Public condition, given the difference in sample size.

[5] See Appendix (pp. A2-A7) for question wording for all outcomes and descriptive statistics for conditioning covariates.

measured objective *Knowledge* about the issue under discussion. To do so, we focused on facts that were included in both the PDF and videos that were available to all participants. From these materials, we identified six quiz questions, which we coded as correct (1) or incorrect (0); "Don't knows" were coded as incorrect.[6] Fourth, respondents were asked about emotions they felt when thinking about the issue of immigration, including *Enthusiasm* (enthusiastic, hopeful, proud), *Surveillance* (anxious, worried, afraid) and *Anger* (anger, bitterness, contempt, hatred), the last of which we reverse-coded to measure *Reduced Anger*. Finally, we measured *Trust* in government, leaders, and officeholders with three items; and *Internal Efficacy* and *External Efficacy* with four items each.[7]

In addition to these outcomes, we measured several covariates: *Party* (seven point); *Ideology* (seven point); *Political Interest*; *Previous Political Activity* (sum of talking politics with peers, talking politics online, contributing to campaigns, and volunteering for campaigns); *Need for Cognition* (sum of two five-point items); *Conflict Avoidance* (first latent dimension from five items); *Age* (in years); and indicators for *Female*; *Asian*, *Black*, *Latino*, and *White* (excluded categories are *Male* and *Other*); *Urban*, *Suburban*, and *Rural* (excluded category is *Small Town*); *College Graduate* and *Some College* (excluded category is *High School Only*); *Full-time*

[6] Results are similar if we instead drop "don't know" responses (see Appendix p. A15-6).

[7] To gauge whether our items cohered into their presumed scales, we estimated α for all nine outcomes. The α values range from 0.7 to 0.9, with the exception of the *Knowledge* scale which has $\alpha = 0.4$. While that level is very low, the scale includes a set of relevant factual questions that we intentionally varied in difficulty, and hence we continue to group those items together. See Appendix (p. A18) for details.

Employment; and *Family Immigration History* (first generation, second, and third; excluded category is "other").

Attrition, Covariate Balance, and Weighting

Because the three conditions in our experiment involved differences in scheduling and the costs of actual attendance, there are potentially important differences between groups of respondents. Given the large number of covariates and comparisons between conditions, we expected some imbalance to occur by chance. More importantly, participants in the *Control* and *Individual* conditions both completed a survey with duration 10-20 minutes, and either survey could be taken at any time within a pre-specified window. In contrast, the *Public* condition required participants to both have time available during which to attend a scheduled online event, and to show up at the appointed time. As a result, even if randomization yielded balanced groups at the assignment stage, selective attrition may have yielded different comparison groups.

We therefore examined balance between conditions by comparing the distributions of covariates for each of three pairs of treatment conditions, and we do find imbalances on some covariates for some comparisons. To quantify balance, we compare standardized differences in covariates between pairs of conditions. We focus on standardized differences at the 95% and 80% levels; i.e., with magnitude greater than 1.96 and 1.28, respectively. In each panel of Figure 1, the left end of each line indicates the magnitude of the standardized difference for a covariate. Between the *Control* and *Individual* conditions, where we expected imbalance only by chance, we see little imbalance: none at the 95% level, and only three cases at the 80% level. As expected, imbalance is worse when comparisons include the *Public* condition. In a few cases, standardized differences exceed the 95% level.

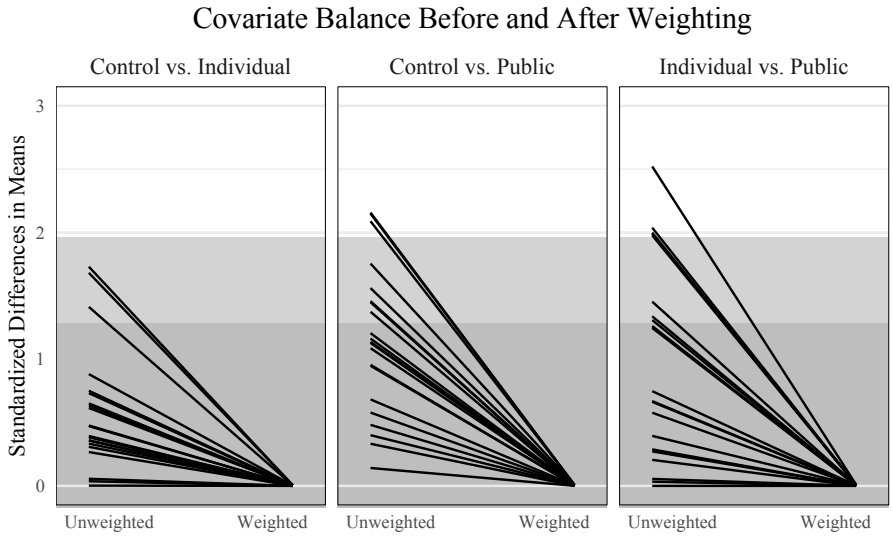


Figure 1. The figure displays covariate balance tests for each covariate and pair of treatment conditions. The lighter gray band depicts the 95% quantile, and the darker one shows 80%. Each line represents the effect of weighting on the standardized difference in covariates for each pair of conditions. In general, weighting reduces imbalance to acceptable levels.

To cope with imbalance, we used covariate balancing generalized propensity scores (Fong, Hazlett, and Imai 2018; for applications, see Davis and Morse 2018 and Charnysh 2019). This method identifies weights that minimize imbalance between groups; the weights are then used in regressions. We used the nonparametric, generalized version of the method because we have three treatment groups, and require balancing across all pairs of conditions. To evaluate, we reprise our balance tests, recalculating standardized differences after weighting. In all cases, the weights reduce imbalance substantially (see Figure 1). In fact, after weighting, the largest standardized difference has magnitude of 0.01.[8]

[8] After weighting, we can reject all null hypotheses of differences in covariate between treatments and the control condition (Hartman and Hidalgo 2018).

Estimation and Statistical Inference

Given these weights, we estimated weighted linear mixed effects models. The structure of our data lends itself to this multilevel modeling approach. The rectangular dataset has 402 respondents (198 in the *Public* deliberation condition, 102 in *Individual* deliberation, and 102 in control) and 41 survey items that we use to measure our nine outcomes of interest. To estimate average treatment effects, we fit the multilevel model

$$
\begin{aligned}
y_i &\sim N(\alpha_i + \beta_i\, Individual_i + \gamma_i\, Public_i,\ \sigma^2) \\
\alpha_i &= \alpha_0 + \alpha_{outcome[i]} + \alpha_{question[i]} + \alpha_{respondent[i]} \\
\beta_i &= \beta_0 + \beta_{outcome[i]} + \beta_{question[i]} \\
\gamma_i &= \gamma_0 + \gamma_{outcome[i]} + \gamma_{question[i]},
\end{aligned}
$$

where i is an observation, y_i is a response, and $Individual_i$ and $Public_i$ are dichotomous indicators of treatment. The intercepts, α_i, combine random intercepts at the levels of the category (*Attitudes toward the Session*, *Perceptions of Changed Mind*, *Knowledge*, etc.), individual question, and respondent.[9] We model the treatment effects, β_i and γ_i, using random coefficients. The multilevel model simultaneously estimates the full set of treatment effects by modeling the treatment effects as random coefficients, which shrinks the point estimates toward the grand mean, and estimates the standard errors of the treatment effect coefficients given the information in the full model and data. As Gelman, Hill, and Yajima (2012) note, the multilevel model addresses the problem of multiple comparisons by making the point estimates more conservative, and so obviates the need to make post-analysis corrections such as Bonferroni.

[9] We also estimated a model including deliberation group-level random intercepts (with each of the respondents in the public condition nested in their CGA groups, and the two control conditions treated as separate groups). Results are similar; see the Appendix (p. A13-4).

In particular, we are interested in the effects of *Individual* and *Public* deliberation at the outcome level. Therefore, we estimate these treatment effects as the sums of the overall coefficients and the outcome-level coefficients. That is, we focus estimates of $\beta_0 + \beta_{outcome}$ and $\gamma_0 + \gamma_{outcome}$ for each of our nine outcomes.

We estimated these models using the lmer function in the lme4 package in R (Bates, Mächler, Bolker, and Walker 2015) For statistical inference, we used the nonparametric block bootstrap. As with most surveys, there is sporadic missing data. For covariates, we imputed using multiple imputation via Amelia II (Honaker, King, and Blackwell 2011); no outcome variables were imputed. Appendix A2 reports all descriptive statistics.[10] Throughout our presentation, we focus on graphical depictions of results in the text and relegate tables to the Appendix.

Results

The main effects of both types of deliberation assess the impact of each deliberation condition relative to control across the full set of items that we use to measure the benefits of deliberation. Both are positive and significantly different from zero. In the case of *Individual* deliberation, the average treatment effect (β_0) was about 3.3 percentage points (bootstrapped 95% interval = [1.5%, 5.1%]). The overall effect of *Public* deliberation (γ_0) was almost double, at 5.8% [4.2%, 7.4%].

[10] Specifically, we (1) block resampled over respondents within treatment groups and deliberative groups, (2) imputed missing values for covariates, creating 10 complete datasets for each resample, (3) estimated CBPS weights for each complete dataset, (4) fit the model for each dataset, and (5) averaged coefficients over imputations. We did so 1000 times to yield distributions of quantities of interest. All 95% intervals are two-tailed and refer to the bootstrap distributions.

The difference-in-differences between the two was statistically significant (bootstrapped p = 0.004).[11] While these magnitudes may seem modest in an absolute sense, we note that we are comparing deliberation to a baseline of strong information provision, similar to an intensive online tutorial, rather than to the naturalistic baseline of no or low information.

We found varying levels of support for our hypotheses. Estimates of average treatment effects are shown in Figure 2, produced by adding the main coefficient (β_0) to that for each category ($\beta_{outcome}$).[12]

First, we find evidence supporting the ***Attitudes toward Experience Hypothesis***. *Public* deliberation caused a 10 percentage point rise in *Attitudes toward the Session* relative to *Control*, as seen in the top row of Figure 2. In contrast, the effect of *Individual* deliberation on *Attitudes toward the Session* is smaller, at 4%. Both effects are significantly different from zero, but participants were more satisfied with *Public* deliberation; the difference-in-differences is significant ($p < 0.001$). In terms of satisfaction, the benefits from public deliberation are substantial.

[11] For all differences-in-differences tests, the bootstrapped p-values we report are the fraction of the 1000 replicates in which the estimated effect for the *Individual* treatment was larger than that for the *Public* treatment.

[12] We also estimated separate scales for each of our nine outcomes using the first principal component of each, and then fit nine separate linear models, one per outcome. Results are similar; see Appendix for details (pp. A18-9).

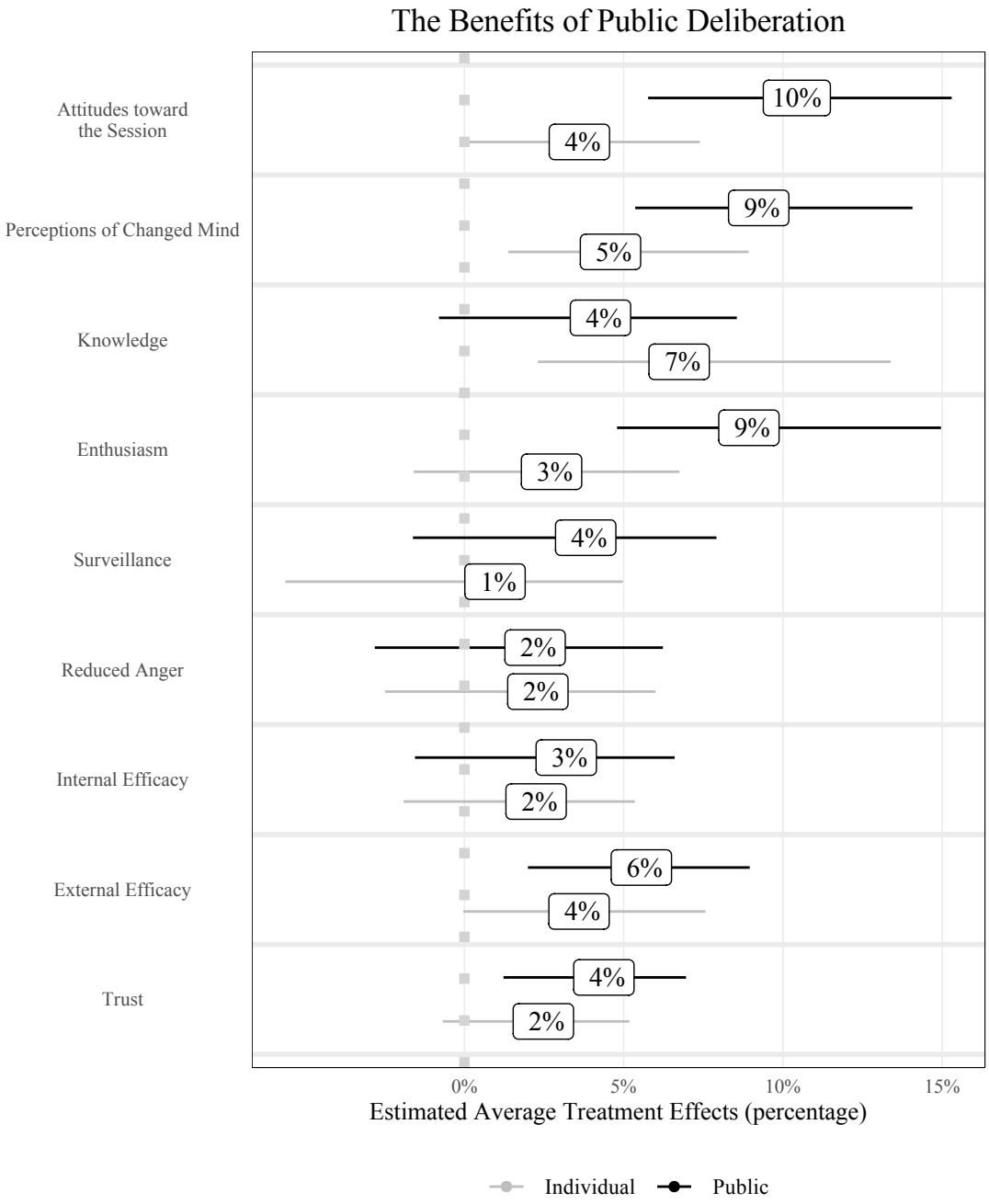


Figure 2. The figure reports average treatment effects in terms of percentages of the outcome's range, for each of the nine outcomes. All effects are estimates of the difference between the relevant treatment group and the information-only control group. Estimates are based on multilevel models with covariate balancing weights, and depicted with 95% intervals. Tabular versions of regression results appear in the Appendix (p. A8).

The evidence for the ***Perceptions of Changed Mind Hypothesis*** suggests that subjects perceived larger changes in themselves after *Public* deliberation. The 9% increase in *Perceptions of Changed Mind* caused by *Public* deliberation is larger than the 5% rise caused by *Individual* Deliberation. The difference-in-differences is significant ($p = 0.018$). Thus, participants both more satisfied with public deliberation, and more likely to believe they had changed their minds.

Intriguingly, the difference between the deliberative experiences was inverted for the ***Knowledge Hypothesis***. In terms of correct answers on our factual quiz, *Individual* deliberation was responsible for a 7% increase, while *Public* deliberation caused only a 4% increase. The difference between effects was insignificant ($p = 0.4$). Thus, while individual deliberation does seem to rival (and perhaps exceed) public deliberation in terms of raw knowledge, that effect is either too small or imprecisely estimated to infer a robust difference. In any case, the subjective experience of having learned was corroborated by the absolute increase in both conditions.

In testing our ***Affective Hypotheses***, we see that *Public* deliberation increased *Enthusiasm* on the immigration issue, causing a 9% increase. The effect of *Individual* deliberation was smaller, at 3%. The difference between the two is significant ($p = 0.004$). We observe smaller differences between effects in *Surveillance* emotions ($p = 0.3$). Neither type of deliberation successfully *Reduced Anger* on the issue. These two null findings deserve further inquiry to assess whether they are the result of different sub-groups—perhaps based on *ex ante* activation—moving in opposite directions and washing each other out. The *Surveillance* emotions are also ambiguous. One might be anxious over uncertainty or because of risk—i.e., one may not know how to evaluate an object, or may worry that a negatively evaluated possibility will obtain. Increased engagement can relieve anxiety from uncertainty yet increase anxiety from risk (Bryner, Devine, and Neblo 2010).

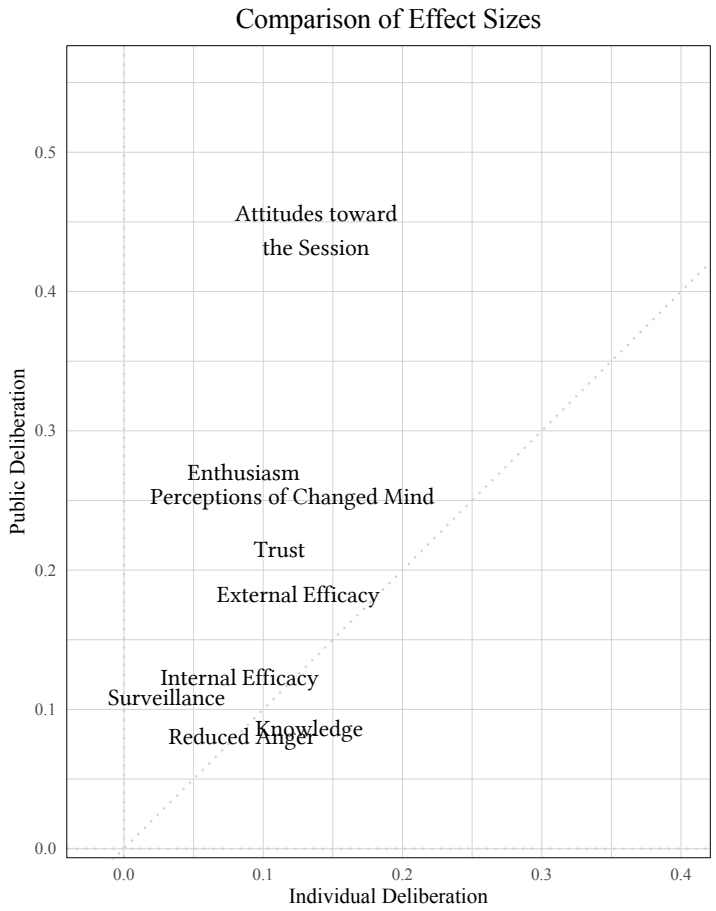


Figure 3. The figure plots Cohen's *d* for each category and treatment. Effect sizes have been separated slightly for clarity.

Finally, we test the ***Trust and Efficacy Hypotheses***. In keeping with previous studies, we find that *Public* deliberation increased *External Efficacy*, though the differences between conditions are both substantively weak and insignificant for both *Internal* and *External* varieties (p = 0.6 and 0.4, respectively). Similarly, although it appears that *Public* deliberation moved the needle on *Trust*, the difference between conditions was again insignificant (p = 0.3).

Figure 3 summarizes our findings graphically with a depiction of the effect sizes of *Individual* and *Public* deliberation on each of our nine categories. All categories that appear above the 45° line indicate that the value of *Public* deliberation exceeds that of its *Individual* counterpart.

Overall, the benefits of *Public* deliberation are clearest for subjective perceptions of the experience, and for gains in *Enthusiasm*. In general, the effect sizes are relatively small.

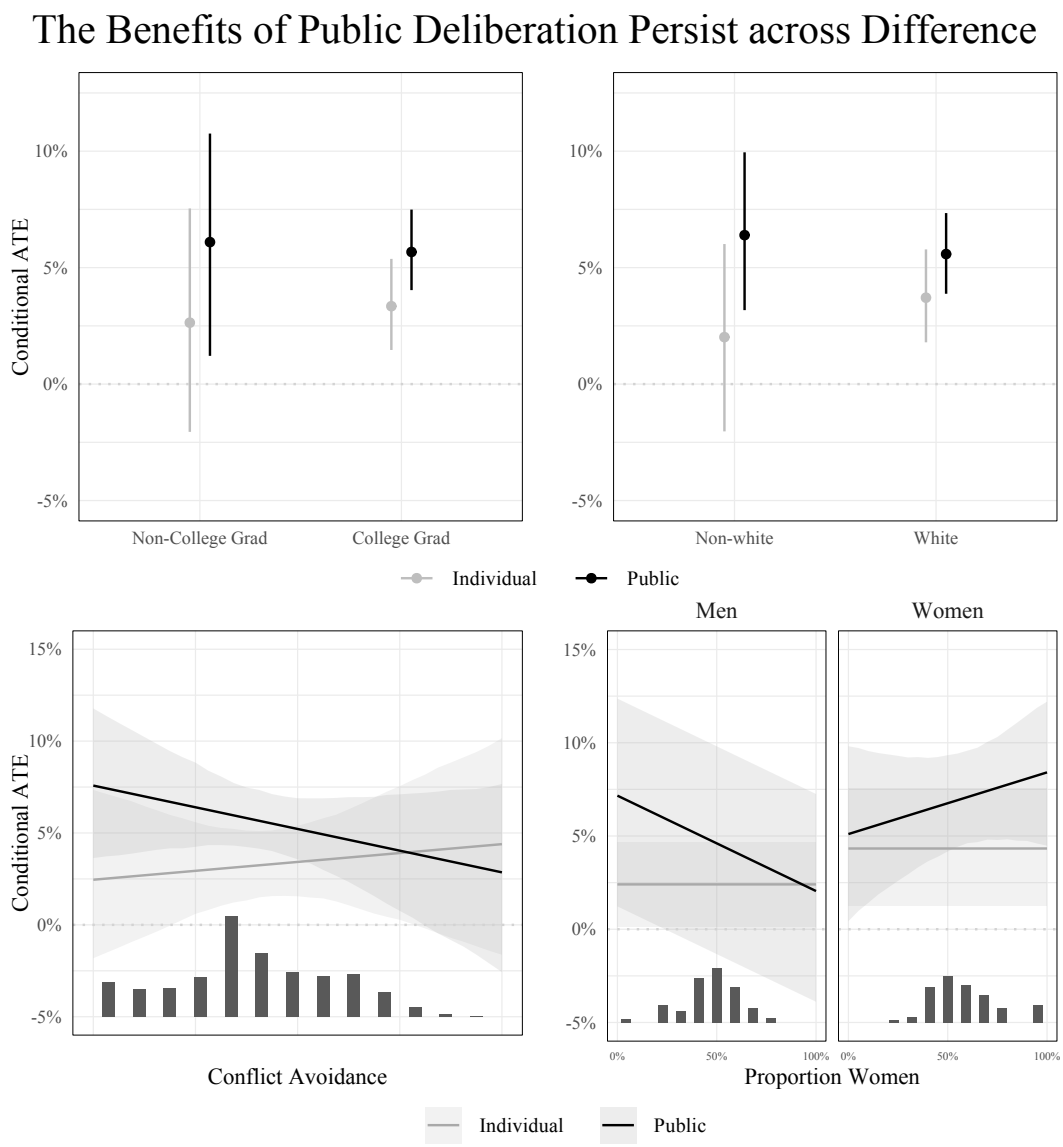


Figure 4. The figure reports conditional average treatment effects as percentages of the outcome's range. Estimates are based on linear mixed effects regression with covariate balance weighting, and depicted with 95% confidence intervals. The bottom two panels include histograms of the distributions of the conditioning variables. Tabular versions of regression results appear in the Appendix (pp. A9-A12).

Potential Disparities in the Benefits of Public Deliberation

We have established that the overall benefits of public deliberation outstrip the benefits from individual reflection at least with respect to some aspects of the experience. But public deliberation faces important limitations that might lead these benefits to be concentrated among only a few participants. For example, the conflict inherent in political discussion might cause the conflict avoidant to benefit less from public deliberation than the conflict acceptant. Similarly, people of color, those who lack a college degree, and women might not benefit from public deliberation as much as white, college-educated men. In each case, the value of individual deliberation may eclipse that of public discussion.

To probe for the conditional effects that would accompany such disparities, we fit separate multilevel models including multiplicative interactions and constitutive terms.[13] We note that our sample size is likely to be too small to precisely estimate interaction effects—unless they are very large. And so, our point in this analysis is to gauge the extent to which the benefits of public deliberation are clearly distributed inequitably, rather than to offer well powered tests of conditional hypotheses.

First, we examine whether conflict avoidant individuals benefit less from public than individual deliberation (Ulbig and Funk 1999; Mutz 2006). We see little evidence of conditionality based on conflict avoidance (see lower left panel of Figure 4). Overall, the estimated effect of *Public* deliberation eclipsed that of *Individual* deliberation for all but the most conflict avoidant,

[13] We do not include random slopes for interactions at the outcome level because these more complicated models failed to converge. We therefore focus our discussion on overall conditional effects. We continue to include random slopes for treatment conditions at those levels.

of whom there were very few in our sample. The coefficients on the interaction of *Conflict Avoidance* with *Public* is -3% [-9%, 2%], while that with *Individual* is 1% [-5%, 7%]. More substantively, in the figure, the 95% intervals of the two effects clearly overlap throughout the range of the conditioning variable. More substantively, the overall difference between the effects of *Public* and *Individual* deliberation was 3% [1%, 5%] for participants with *Conflict Avoidance* at the lowest observed tercile, and 2% [0%, 3%] for those at the highest one—positive in both cases. We conclude that, at least in this case, both those who seek and avoid conflict shared in the value of public deliberation. Put another way, we do not see clear, unambiguous evidence that the conflict acceptant reaped the vast share of the benefits from public deliberation.

Beyond differences in psychological traits, the less privileged may also not share equally in the benefits of public deliberation. In particular, people of color and those without college educations may not benefit, either because the more privileged actively dominate discussions, or because the less privileged introject such power disparities and consequently participate less fully. About 81% of the sample held a college degree, and 73% were white, allowing us to test for these possibilities.[14]

The top two panels of Figure 4 reveal little evidence of difference in benefits by either college education or race/ethnicity. In fact, the point estimates for the effects of *Public* deliberation are actually very slightly larger for non-college graduates and for people of color. Specifically, the

[14] The proportions of college graduates are virtually identical across conditions, while the distribution of white participants is more varied (68% in *Public*, 75% in *Individual*, 79% in *Control*). See Table A3 in the Appendix (p. A6) for details on deliberative-group level mean covariates.

coefficient on the interaction of *Non-College Graduate* and *Public* is 0% [-5%, 5%], and that with *Individual* is -1% [-6%, 5%]. Similarly, the coefficient on the interaction of *Non-white* and *Public* is 1% [-3%, 5%], and that with *Individual* is -2% [-6%, 3%]. In more substantive terms, the overall difference between *Public* and *Individual* treatments was 3% [-1%, 7%] for non-college graduates, and 2% [1%, 4%] for college graduates. Similarly, that difference was 4% [1%,8%] for non-white participants, and 2% [0%, 4%] for white participants. We conclude that the benefits of deliberation were shared across differences in education and race/ethnicity—or at least that any disparities are not large enough for us to detect.

Finally, we consider conditional effects by participants' gender and the gender mix of the deliberating groups. Karpowitz and Mendelberg (2014, KM) suggest that many potentially beneficial aspects of the deliberative setting depend on the style and performance of individual participants, which in turn depend at least in part on whether women are in the minority. Further, they argue that the costs are relatively larger for women participants than for men. It is therefore plausible that *Public* deliberation could even be counterproductive relative to *Individual* deliberation, especially for women and when they are in the minority. Therefore, we focus our analysis on comparisons of the relative benefits of *Public* deliberation in groups dominated by men, and those dominated by women, for participants of both genders.

Overall, 48% of our participants identified as women, although women were slightly overrepresented in the *Public* deliberation condition (52%) as compared to the *Individual* deliberation (44%) and *Control* (43%) conditions. The *Proportion of Women* in our 39 discussion groups was consequently slightly elevated, with a mean of 52% and a standard deviation of 19%, ranging from groups of all men to all women. The Pearson correlation between group size and *Proportion of Women* was very small, at 0.05. To analyze these effects, we fit a multilevel model

including three-way interactions of participant gender, treatment indicators, and *Proportion of Women*, replicating KM's analysis.

Importantly, our deliberations differ from KM in two ways that might have reduced differences by gender. First, our discussion groups occurred online rather than in-person. To the extent that nonverbal cues are necessary to stimulate differences in gender (versus introjection), online forums might attenuate gender biases. Second, KM's forums involved incentivized collective choice and a varying decision rule, finding that women are disadvantaged in majority rule settings, but not with a unanimity rule. KM suggest that the pivotal mechanism for the difference in biases by decision rule is that majority rule encourages participants to engage in contestation, while unanimity rule fosters consensus-building. Common Ground for Action forums, in contrast, use a supermajoritarian preference aggregation for informational purposes, without incentivized collective choice.[15]

While both of these differences might limit gender differences, there is also good reason to expect that such biases might persist. First, almost all forums included participants with a mix of gender identities. Participants used their first names, which often reveal an individual's gender, and, to the extent that gendered behavior is mediated through language and especially contestation, the fact that forums were online may by irrelevant. Indeed, if limitations from such things as conflict aversion and disparities were intrinsic to deliberation, as critics fear, then these limitations should persist across different modes for deliberation. Second, although Common Ground for Action forums did not include incentivized collective choice or use majority rule, the forums are

[15] The forums use an 80% rule to identify which policy solutions are "in the common ground," and thus should be strongly considered for adoption.

designed to foster contestation. Participants are encouraged to talk about points on which they disagree, and to offer arguments that might persuade others. So it is plausible that we will observe conditional effects of *Public* deliberation depending on both gender and gender mix.

The bottom right panel of Figure 4 displays overall treatment effects for both men and women, and for varying proportions of women in the deliberating groups. The effect of *Individual* deliberation is a flat line for each gender, as it does not depend on the composition of the group. Across the board, we see that the effect of *Public* deliberation dominates that of *Individual* deliberation. For example, when discussion groups are composed of equal proportions of men and women, the difference-in-differences between treatment effects is nearly identical for men, 2% [0%, 4%], and women, 2% [0%, 5%]. We conclude that, at least in balanced groups, there are no gender differences in the benefits of public deliberation.

We also see suggestive evidence that is consistent with KM: both women and men seem to derive somewhat larger benefits from discussion in groups in which their gender is the more numerous. Consider two hypothetical groups of five, one with one woman, and the other with one man. According to our model, the difference between the effects of *Public* and *Individual* deliberation in the group with one woman is 1% [-2%, 5%] for the lone woman, yet it is 4% [0%, 7%] for the men. Comparatively, in the group with one man, that difference is 3% [0%, 6%] for the woman, and 1% [-4%, 5%] for the lone man. Thus, our findings are consistent with the direction of KM's findings.

Conclusion

We reported on head-to-head comparisons of public and individual deliberation, testing a wide range of hypotheses and probing for conditional effects based on dispositions and identities. While

on some accounts the normative appeal of deliberation is inherently individual, our findings suggest that the benefits from public deliberation can be substantial. Both in overall terms, as well as attitudes toward the session, subjective perceptions of change and reflection, and enthusiasm about the issue, the effects of public deliberation supersede those of individual reflection. The only case in which individual deliberation rivals its public counterpart was in knowledge gains, and that difference is small and imprecisely measured. Unlike previous studies that merely provide subjects with information, subjects assigned to our individual deliberation group were effectively encouraged to engage in such reflection. So, while depth of processing across the public and individual groups may have contributed to our findings, we are confident that the differences we identified are due not only to inducement, but also to genuine differences in experience as a result of conversation.

The conditional effects we report also destabilize the notion that the disparities associated with face-to-face public deliberation are reproduced in this online setting. We found no systematic evidence that the conflict avoidant gained more from individual than public deliberation. While we found suggestive evidence of differences in deliberation effects depending on gender mix in discussion groups, we found no circumstances in which the value of individual deliberation eclipsed that of public deliberation for differences in education or race/ethnicity; and further, the disparities were symmetric across genders. At the least, the upshot of our research is to place a relatively low ceiling on the magnitude of inequality in the benefits of public deliberation.

Based on our experiment, we conclude that the value added of public deliberation is clearest and largest for participants' subjective experiences: their attitudes toward their experiences and their perceptions of their minds having changed. But effects are relatively small overall, as our depiction reveals. From this, we conclude that, on the one hand, public deliberation may be most

valuable insofar as it incentivizes individuals to actually think about difficult policy questions. But, once the deliberation has actually occurred, we see only sporadic gains—in enthusiasm, which may be an outlier, given our sample size. That effect sizes are small does not necessarily diminish the value of deliberation full stop, but it does mean that once the expenses of public deliberation are taken into account, it remains plausible that its costs outweigh its benefits.

Importantly, future study should focus on better identifying conditional effects with larger samples and more precise instrumentation. While our sample size was not unreasonably small—we report on 402 participants—we nevertheless found only imprecise estimates of several treatment effects. That could be because deliberation, either public or individual, simply does not offer such benefits. Or, it could be because the sample sizes we use did not offer sufficient statistical power to reliably estimate such effect sizes. Simultaneously, the significant estimates we report may actually be overestimates, or errors of magnitude (Gelman and Carlin 2014). In particular, future study should focus on the effects of public deliberation on the emotional subsystems, as we found effects only for the enthusiasm and not for anger or anxiety. All told, our study offers a guide for future empirical studies of deliberation, which should seek samples substantially larger than what we report on here.

Although public deliberation remains expensive—for both participants and planners—those costs appear to produce substantial value. If we had found little evidence of differences in the effects of public and individual deliberation, that would have called into question the deep investments that institutions and governments have made in the practice. Instead, our findings suggest that, public deliberation may well be worth that cost.

References

Albertson, Bethany and Gadarian, Shana Kushner, 2015. *Anxious politics: Democratic citizenship in a threatening world*. Cambridge University Press.

Anderson, Monica, Posts, Bio, Andrew Perrin, et al. 2018. *11% of Americans don't use the internet. Who are they?* http://www.pewresearch.org/fact-tank/2018/03/05/some-americans-dont-use-the-internet-who-are-they/.

Baek, Young Min, Magdalena Wojcieszak, and Michael X Delli Carpini. 2012. "Online versus face-to-face deliberation: Who? Why? What? With what effects?" *New Media & Society* 14(3): 363–383.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67(1): 1–48.

Brandom, Robert. 1998. *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

Brennan, Jason. 2017. *Against Democracy*. Princeton University Press.

Bryner, S., Devine, C. and Neblo, M., 2010. Is Anxiety About Immigration a Response to Perceptions of Risk or Uncertainty? Paper presented at the Annual Meeting of the American Political Science Association, Washington, DC, 1–4 September.

Carson, Lyn, John Gastil, Janette Hartz-Karp, and Ron Lubensky. 2013. *The Australian Citizens' Parliament and the Future of Deliberative Democracy*. Penn State Press.

Chambers, Simone. 2018. "Human Life Is Group Life: Deliberative Democracy for Realists." *Critical Review* 30(1–2): 36–48.

Charnysh, V., 2019. Diversity, Institutions, and Economic Outcomes: Post-WWII Displacement in Poland. *American Political Science Review*, *113*(2), pp.423-441.

Collins, Allan M, and Elizabeth Fee Loftus. 1975. "A spreading-activation theory of semantic processing." *Psychological Review* 86(2): 407.

Davis, C.L. and Morse, J.C., 2018. Protecting Trade by Legalizing Political Disputes: Why Countries Bring Cases to the International Court of Justice. *International Studies Quarterly*, *62*(4), pp.709-722.

Delli Carpini, Michael X, Fay Lomax Cook, and Lawrence R Jacobs. 2004. "PUBLIC DELIBERATION, DISCURSIVE PARTICIPATION, AND CITIZEN ENGAGEMENT: A Review of the Empirical Literature." *Annu. Rev. Polit. Sci.* 7(1): 315–344.

Dryzek, John S. et al. 2019. "The crisis of democracy and the science of deliberation." *Science* 363(6432): 1144–1146.

Elster, Jon. 1986. *Sour Grapes*. Cambridge University Press.

Esterling, Kevin M, Michael A Neblo, and David M J Lazer. 2011a. "Means, Motive, and Opportunity in Becoming Informed about Politics: A Deliberative Field Experiment with Members of Congress and Their Constituents." *Public Opin. Q.* 75(3): 483–503.

Feldman, S., Huddy, L. and Cassese, E., 2004, October. Exploring the Structure of Political Affect: The Distinct Effects of Anxiety and Anger. In *Dynamics of Affective Intelligence Workshop at the University of Michigan*.

Fishkin, James S. 2011. *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford University Press.

Gastil, John, E Pierre Deess, Philip J Weiser, and Cindy Simmons. 2010. *The Jury and Democracy: How Jury Deliberation Promotes Civic Engagement and Political Participation*. Oxford University Press.

Gastil, John. 2018. "The Lessons and Limitations of Experiments in Democratic Deliberation." *Annu. Rev. Law. Soc. Sci.* 14(1): 271–291.

Gastil, John, E Pierre Deess, and Phil Weiser. 2002. "Civic Awakening in the Jury Room: A Test of the Connection between Jury Deliberation and Political Participation." *J. Polit.* 64(2): 585–595.

Gastil, John, and James P Dillard. 1999. "The aims, methods, and effects of deliberative civic education through the national issues forums." *Commun. Educ.* 48(3): 179–192.

Gelman, Andrew and John Carlin. 2014. Beyond Power Calculations: Type S (Sign) and Type M (Magnitude) Errors. 2014. *Perspectives on Psychological Science*, 9(6) 641–651.

Gelman, Andrew, Jennifer Hill, and Masanao Yajima. 2012. "Why We (Usually) Don't Have to Worry About Multiple Comparisons." *Journal of Research on Educational Effectiveness* 5(2): 189–211.

Goodin, Robert E. 2000. "Democratic Deliberation Within." *Philos. Public Aff.* 29(1): 81–109.

Goodin, Robert E. 2003. *Reflective Democracy*. New York: Oxford University Press.

Goodin, Robert E, and Simon J Niemeyer. 2003. "When Does Deliberation Begin? Internal Reflection versus Public Discussion in Deliberative Democracy." *Polit. Stud.* 51(4): 627–649.

Goold, Susan Dorr, Michael A. Neblo, Scott Y H Kim, Raymond De Vries, et al. 2012. "What is Good Public Deliberation?" *The Hastings Center Report* 42(2): 24–26.

Habermas, Jurgen. 1996. *Between Facts and Norms*. Cambridge, MA: MIT Press.

Habermas, Jurgen. 1984. 1 *Theory of Communicative Action*. Boston: Beacon Press.

Hartman, Erin, and F. Daniel Hidalgo. "An equivalence approach to balance and placebo tests. American Journal of Political Science. 62(4):1000-13.

Hertzum, Morten, Hans H. K. Andersen, Verner Andersen, and Camilla B. Hansen. 2002. "Trust in information sources: seeking information from people, documents, and virtual agents." *Interacting with Computers* 14(5): 575–599.

Honaker, James, Gary King, and Matthew Blackwell. 2011. "Amelia II: A program for missing data." *Journal of Statistical Software* 45(7): 1–47.

Huddy, Leonie, Lilliana Mason, and Lene Aarøe. 2015. "Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity." *Am. Polit. Sci. Rev.* 109(1): 1–17.

Karpowitz, Christopher F, and Tali Mendelberg. 2014. *The Silent Sex: Gender, Deliberation, and Institutions*. Princeton University Press.

Klar, Samara. 2014. “Partisanship in a Social Setting.” *American Journal of Political Science* 58(3): 687–704.

Landemore, Hélène. 2012. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton University Press.

Marcus, George E., W. Russell Neuman, and Michael MacKuen. 2000. *Affective Intelligence and Political Judgment*. University of Chicago Press.

Mccombs, Maxwell, and Amy Reynolds. 1999. *The Poll With A Human Face: The National Issues Convention Experiment in Political Communication*. Routledge.

Mercier, Hugo, and Hélène Landemore. 2012. “Reasoning Is for Arguing: Understanding the Successes and Failures of Deliberation.” *Polit. Psychol.* 33(2): 243–258.

Minozzi, William, Michael A Neblo, Kevin M Esterling, and David M J Lazer. 2015. “Field experiment evidence of substantive, attributional, and behavioral persuasion by members of Congress in online town halls.” *Proc. Natl. Acad. Sci. U. S. A.* 112(13): 3937–3942.

Morrell, Michael E. 2005. “Deliberation, Democratic Decision-Making and Internal Political Efficacy.” *Political Behavior* 27(1): 49–69.

Mutz, Diana C. 2006. *Hearing the Other Side: Deliberative Versus Participatory Democracy*. Cambridge University Press.

Myers, C. Daniel. 2017. “Interests, Information, and Minority Influence in Deliberation.” *The Journal of Politics* 79(3): 804–822.

Myers, C Daniel, and Tali Mendelberg. 2013. “Political Deliberation.” In *The Oxford Handbook of Political Psychology*, eds. Leonie Huddy, David O Sears, and Robert Jervis. New York: Oxford University Press, p. 699–734.

Nabatchi, Tina. 2010. “Deliberative democracy and citizenship: In search of the efficacy effect.” *Journal of Public Deliberation* 6(2): 8.

Neblo, Michael A. 2006. “Change for the better? Linking the mechanisms of deliberative opinion change to normative theory.” In Chicago, IL.

Neblo, Michael A. 2015. *Deliberative Democracy between Theory and Practice*. Cambridge University Press.

Neblo, Michael A. 2020. "Impassioned Democracy: The Roles for the Emotions in Deliberative Theory." *American Political Science Review*.

Neblo, Michael A, Kevin M Esterling, and David M J Lazer. 2018. *Politics with the People: Building a Directly Representative Democracy*. Cambridge University Press.

Nystrand, Martin, and Adam Gamoran. 1991. “Instructional Discourse, Student Engagement, and Literature Achievement.” *Research in the Teaching of English* 25(3): 261–290.

Page, Scott E. 2007. “Making the Difference: Applying a Logic of Diversity.” *AMP* 21(4): 6–20.

Papacharissi, Zizi. 2002. “The virtual sphere: The internet as a public sphere.” *New Media & Society* 4(1): 9–27.

Pierce, Jason L, Grant Neeley, and Jeffrey Budziak. 2008. “Can deliberative democracy work in hierarchical organizations?” *Journal of Public Deliberation* 4(1): 14.

Price, Vincent, Joseph N Cappella, and Lilach Nir. 2002. “Does Disagreement Contribute to More Deliberative Opinion?” *Political Communication* 19(1): 95–112.

Sanders, Lynn M. 1997. “Against Deliberation.” *Polit. Theory* 25(3): 347–376.

Stromer-Galley, Jennifer. 2003. “Diversity of Political Conversation on the Internet: Users’ Perspectives.” *J. Comput. Mediat. Commun.* 8(3).

Sunstein, Cass R. 2002. “On a Danger of Deliberative Democracy.” *Daedalus* 131(4): 120–124.

Surowiecki, James. 2004. *The Wisdom of Crowds: Why the Many are Smarter Than the Few and how Collective Wisdom Shapes Business, Economies, Societies, and Nations*. Doubleday.

Taber, Charles S, and Milton Lodge. 2006. “Motivated Skepticism in the Evaluation of Political Beliefs.” *Am. J. Pol. Sci.* 50(3): 755–769.

Ugarriza, Juan E, and Enzo Nussio. 2016. “There is No Pill for Deliberation: Explaining Discourse Quality in Post-conflict Communities.” *Swiss Polit. Sci. Rev.* 22(1): 145–166.

Ulbig, Stacy G, and Carolyn L Funk. 1999. “Conflict Avoidance and Political Participation.” *Political Behavior* 21(3): 265–282.

Warren, Mark E, and Hilary Pearse. 2008. *Designing Deliberative Democracy: The British Columbia Citizens’ Assembly*. Cambridge University Press.

Wojcieszak, Magdalena E, and Diana C Mutz. 2009. “Online Groups and Political Discourse: Do Online Discussion Spaces Facilitate Exposure to Political Disagreement?” *J. Commun.* 59(1): 40–56.

Online Appendix for

Testing the Benefits of Public Deliberation

William Minozzi, The Ohio State University
Ryan Kennedy, University of Houston
Kevin M. Esterling, University of California, Riverside
Michael A. Neblo, The Ohio State University
Ryan Jewell, University of Houston

August 2020

Table of Contents

Section A1. Question Wording for Outcome Variables.

Attitudes toward the Session. Items 2-5 reverse keyed, scaled to 0-1.

ATS1. How satisfied are you with the policy session as a whole? (1 = Not at all satisfied, 2, 3, 4 = Moderately Satisfied, 5, 6, 7 = Extremely satisfied, DK)

To what degree do you agree or disagree with the following statements? (1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, 5 = Strongly disagree)

ATS2. I found this session to be helpful and informative.

ATS3. In the future, I would be interested in participating in sessions like this one on other political issues.

ATS4. I feel like I learned a lot from participating in this session.

ATS5. It would be useful for my Member of Congress to see the results of this session.

ATS6. Discussions with my fellow citizens on topics like this are important in our democracy.

Perceptions of Changed Mind. Reverse-keyed item CM2, scaled to 0-1.

PCM1. Compared to what you knew about this issue, how much more do you know about this issue after this policy session? (1 = I understand the issue less well, 2, 3 = I understand the issue about the same, 4, 5 = I understand the issue a little better, 6, 7 = I understand the issue much better, DK)

PCM2. Compared to how you felt about this issue, how confused about this issue do you feel after this policy session? (1 = much more confused, 2 = a little more confused, 3 = about the same as before, 4 = a little less confused, 5 = much less confused, DK)

PCM3. Compared to how you saw things before this policy session began, how much would you say you respect the views of people who disagree with you on this issue? (1 = I respect their views a lot less, 2 = I respect their views a little less, 3 = about the same as before, 4 = I respect their views a little more, 5 = I respect their views a lot more, DK)

Some people change their minds after a policy session, but others don't. Do any of these describe you? (0 = No, 0.5 = Maybe, 1 = Yes, DK)

PCM4. I support more strongly the actions I favored before the session.

PCM5. I support new actions now that I didn't support before.

PCM6. I now oppose some actions that I favored before the session.

PCM7. I am more sympathetic to some actions that I still oppose.

PCM8. I now recognize clear tradeoffs for some actions that I still support.

Knowledge. Six items, correct responses (in bold) coded as 1, all others 0.

K1. About what proportion of Americans support some form of immigration reform? (1/4, 1/3, 1/2, ***2/3***, 3/4, DK)

K2. Under current law, are undocumented immigrants who came here as minors and graduate high school automatically eligible to become citizens? (Yes, ***No*** DK)

K3. About what percentage of farmworkers in the U.S. are undocumented workers? (10%, 20%, ***50%,*** 75%, DK)

K4. Under current law, are undocumented immigrants who have lived in the U.S. for five years or more, and have no criminal record, eligible to apply for citizenship? (Yes, ***No***, DK)

K5. Under current law, do most undocumented workers pay into social security, even if they are not eligible for benefits? (***Yes***, No, DK)

K6. About how much economic activity do economists estimate undocumented immigrants are responsible for? ($7 billion, $75 billion, ***$150 billion***, $300 billion, DK)

Enthusiasm, Surveillance, & ***Reduced Anger.*** Three, three, & four items respectively, reverse keyed (except for *Reduced Anger*), scaled 0-1.
When thinking about the issue of immigration, to what extent would you say that you feel the emotions listed below? (1 = Strongly, 2 = Somewhat, 3 = A little, 4 = Not at all)

E1. Enthusiastic
E2. Hopeful
E3. Proud
S1. Anxious
S2. Worried
S3. Afraid
A1. Hatred
A2. Contempt
A3. Bitterness
A4. Anger

Internal efficacy. Four items, items 1 & 2 reverse-keyed, scaled 0-1.
To what degree do you agree or disagree with the following statements? (1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, 5 = Strongly disagree)

IE1. I consider myself well-qualified to participate in politics.
IE2. I feel that I could do as good a job in public office as most other people.
IE3. Sometimes politics and government seem so complicated that a person like me can't really understand what's going on.
IE4. I often don't feel sure of myself when talking with other people about politics and government.

External efficacy. Four items, items 3 & 4 reverse-keyed, scaled 0-1.
To what degree do you agree or disagree with the following statements? (1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, 5 = Strongly disagree)

EE1. People like me don't have any say about what the government does.
EE2. If public officials are not interested in hearing what the people think, there is really no way to make them listen.
EE3. Under our form of government, the people have the final say about how the country is run, no matter who is in office.
EE4. There are many legal ways for citizens to successfully influence what the government does.

Trust. Six items, correct responses (in bold) coded as 1, all others 0.
How much of the time do you think the following statements are true?
(1 = Just about always, 2 = Most of the time, 3 = Only some of the time, 4 = Almost never)

T1. You can trust the people who run our government to do what is right.
T2. When government leaders make statements to the American people on television or in the newspapers, they are telling the truth.
T3. The people we elect to public office try to keep the promises they have made during the election.

Section A2. Descriptive Statistics

Table A1. Descriptive Statistics for Covariates and Group Size

Variable	Mean	SD	Obs.Min	Obs.Max	# Missing
Conflict Avoidance	-0.36	0.33	-1	0.6	13
Female	0.48	0.5	0	1	14
Proportion Women	0.54	0.19	0	1	0
Party	-0.06	0.67	-1	1	16
Ideology	-0.02	0.58	-1	1	23
Political Interest	0.44	0.47	-1	1	7
Prev. Political Actions	0.07	0.58	-1	1	0
Need for Cognition	0.66	0.37	-0.75	1	11
Age	47.2	12.6	22	76	13
Asian	0.04	0.21	0	1	28
Black	0.07	0.25	0	1	28
Latino	0.06	0.23	0	1	28
White	0.74	0.44	0	1	28
Some College	0.2	0.4	0	1	0
College	0.8	0.4	0	1	0
Fulltime Employment	0.9	0.3	0	1	10
First Gen. Immigrant	0.1	0.3	0	1	0
Second Gen. Immigrant	0.1	0.3	0	1	0
Third+ Gen. Immigrant	0.2	0.4	0	1	0
Urban	0.2	0.4	0	1	0
Suburban	0.3	0.5	0	1	0
Rural	0.1	0.3	0	1	0
Number of Participants	6.1	1.7	2	9	0

All continuous covariates are rescaled to have theoretically possible ranges of –1 to 1. Summaries of *Proportion Female* and *Number of Participants* are calculated only for participants in the **Public Deliberation** condition. All statistics are calculated based on unimputed data.

Table A2. Descriptive Statistics for Outcomes.

Variable	Mean	SD	Obs. Min	Obs. Max	# Missing
ATS1	0.7	0.3	0	1	14
ATS2	0.7	0.2	0	1	4
ATS3	0.8	0.2	0	1	4
ATS4	0.6	0.3	0	1	4
ATS5	0.8	0.2	0	1	4
ATS6	0.9	0.2	0	1	4
PCM1	0.4	0.3	0	1	5
PCM2	0.6	0.2	0.2	1	15
PCM3	0.6	0.2	0	1	14
PCM4	0.6	0.4	0	1	22
PCM5	0.3	0.4	0	1	16
PCM6	0.1	0.3	0	1	17
PCM7	0.5	0.4	0	1	16
PCM8	0.7	0.4	0	1	14
K1	0.4	0.5	0	1	0
K2	0.7	0.5	0	1	0
K3	0.4	0.5	0	1	0
K4	0.5	0.5	0	1	0
K5	0.5	0.5	0	1	0
K6	0.2	0.4	0	1	0
E1	0.4	0.3	0	1	8
E2	0.5	0.3	0	1	8
E3	0.3	0.3	0	1	8
S1	0.5	0.3	0	1	8
S2	0.6	0.3	0	1	8
S3	0.3	0.3	0	1	9
A1	0.1	0.2	0	1	9
A2	0.2	0.3	0	1	9
A3	0.2	0.3	0	1	9
A4	0.2	0.3	0	1	9
IE1	0.7	0.2	0	1	6
IE2	0.6	0.3	0	1	6
IE3	0.7	0.3	0	1	6
IE4	0.6	0.3	0	1	6
EE1	0.6	0.3	0	1	5
EE2	0.5	0.3	0	1	5
EE3	0.4	0.3	0	1	5
EE4	0.7	0.3	0	1	5
T1	0.3	0.2	0	1	6
T2	0.3	0.2	0	1	6
T3	0.4	0.2	0	1	6

Note. Variable names refer to Appendix A1. Items A1-A4 (Anger) were reversed for analysis.

Table A3. Descriptive Statistics for Mean Covariates for Deliberative Groups

Variable	Weighted	Mean	SD	Obs.Min	Obs.Max	Cor. w/ Group Size
Conflict Avoidance	No	-0.38	0.14	-0.6	-0.1	0.11
	Yes	-0.37	0.14	-0.6	-0.1	0.16
Female	No	0.51	0.21	0	1	0.09
	Yes	0.48	0.22	0	1	0.04
Proportion Women	No	0.52	0.19	0	1	0.09
	Yes	0.52	0.19	0	1	0.09
Party	No	-0.08	0.40	-0.8	0.6	0.08
	Yes	-0.06	0.40	-0.8	0.7	0.05
Ideology	No	-0.08	0.35	-0.8	0.7	0.08
	Yes	-0.05	0.36	-0.8	0.7	0.08
Political Interest	No	0.50	0.23	0	0.9	-0.22
	Yes	0.47	0.23	-0.1	0.9	-0.27
Prev. Political Action	No	0.17	0.32	-0.6	0.8	-0.14
	Yes	0.11	0.31	-0.6	0.7	-0.18
Need for Cognition	No	0.66	0.22	0	1	0.11
	Yes	0.65	0.22	0	1	0.06
Age	No	47.43	6.45	32.2	61.5	-0.05
	Yes	47.42	6.62	31.4	61.7	-0.08
Asian	No	0.06	0.10	0	0.3	0.03
	Yes	0.05	0.09	0	0.3	0.03
Black	No	0.08	0.11	0	0.4	0.10
	Yes	0.07	0.11	0	0.4	0.02
Latino	No	0.08	0.12	0	0.5	-0.15
	Yes	0.07	0.10	0	0.4	-0.14
White	No	0.68	0.20	0.2	1	0.17
	Yes	0.71	0.2	0.2	1	0.22
Some College	No	0.15	0.21	0	1	-0.15
	Yes	0.16	0.22	0	1	-0.15
College	No	0.81	0.23	0	1	0.11
	Yes	0.80	0.23	0	1	0.12
Fulltime Employment	No	0.91	0.14	0.5	1	0.10
	Yes	0.92	0.13	0.5	1	0.08
First Gen. Immigrant	No	0.08	0.12	0	0.4	0.37
	Yes	0.08	0.11	0	0.4	0.35
Second Gen. Immigrant	No	0.11	0.14	0	0.5	-0.17
	Yes	0.09	0.12	0	0.4	-0.15
Third+ Gen. Immigrant	No	0.19	0.20	0	0.7	-0.18
	Yes	0.22	0.23	0	0.7	-0.17
Urban	No	0.22	0.20	0	0.8	0.20
	Yes	0.22	0.21	0	0.8	0.22
Suburban	No	0.25	0.20	0	0.8	0.04
	Yes	0.28	0.22	0	0.8	0.09
Rural	No	0.15	0.19	0	0.7	-0.15
	Yes	0.13	0.17	0	0.7	-0.19

Note to Table A3 appears on next page.

Note to Table A3 (p. A6)
All continuous covariates are rescaled to have theoretically possible ranges of –1 to 1. All cells report descriptive statistics on means (weighted or not, as indicated) at the deliberative group-imputation level, then averaged over 10 imputations. The rightmost column reports the correlation between *Number of Participants* and deliberative group mean values for each covariate.

Section A3. Regression Models from Main Text

Table A4. Details on Main Multilevel Model.

Main Effects	**Estimate [95% Interval]**
Individual Deliberation (β_0)	0.033 [0.015, 0.051]
Public Deliberation (γ_0)	0.058 [0.042, 0.074]
Intercept (α_0)	0.508 [0.494, 0.522]
Group-level Effects	**SD**
Respondent ($n_{respondents}$ = 402)	
Intercept ($\alpha_{respondent}$)	0.079
Outcome ($n_{outcomes}$ = 9)	
Individual Deliberation ($\beta_{outcome}$)	0.051
Public Deliberation ($\gamma_{outcome}$)	0.044
Intercept ($\alpha_{outcome}$)	0.131
Question ($n_{questions}$ = 41)	
Individual Deliberation ($\beta_{question}$)	0.031
Public Deliberation ($\gamma_{question}$)	0.041
Intercept ($\alpha_{question}$)	0.163
Residual (σ^2)	0.015

n_{total} = 16510. The table presents details on the main multilevel model from the text (see model in "Estimation and Statistical Inference" section. The model was estimated using the lmer function from the lme4 package in R, using covariate balancing propensity score weights. To ease convergence, we fixed the group-level correlations between coefficients to be 0. Confidence intervals refer to bootstrap distribution of coefficients, with block resampling within each of the treatment conditions. Missing outcome observations, of which there were 300, were removed.

Table A5. Details on Non-College Graduate Model.

Main Effects	**Estimate [95% Interval]**
Individual Deliberation (β_0)	0.033 [0.015, 0.054]
Public Deliberation (γ_0)	0.057 [0.040, 0.075]
Non-College Grad.	–0.036 [–0.078, 0.007]
Non-College Grad. × *Individual*	–0.007 [–0.061, 0.046]
Non-College Grad. × *Public*	0.004 [–0.050, 0.054]
Intercept (α_0)	0.515 [0.500, 0.529]

Group-level Effects	**SD**
Respondent ($n_{respondents} = 402$)	
Intercept ($\alpha_{respondent}$)	0.078
Outcome ($n_{outcomes} = 9$)	
Individual Deliberation ($\beta_{outcome}$)	0.051
Public Deliberation ($\gamma_{outcome}$)	0.044
Intercept ($\alpha_{outcome}$)	0.131
Question ($n_{questions} = 41$)	
Individual Deliberation ($\beta_{question}$)	0.031
Public Deliberation ($\gamma_{question}$)	0.041
Intercept ($\alpha_{question}$)	0.163
Residual (σ^2)	0.015

$n_{total} = 16510$. See note on Table A3 for details on estimation and inference.

Table A6. Details on Non-White Model.

Main Effects	**Estimate [95% Interval]**
Individual Deliberation (β_0)	0.037 [0.018, 0.058]
Public Deliberation (γ_0)	0.056 [0.039, 0.073]
Non-White	–0.005 [–0.039, 0.028]
Non-White × *Individual*	–0.017 [–0.061, 0.030]
Non-White × *Public*	0.008 [–0.030, 0.047]
Intercept (α_0)	0.509 [0.494, 0.524]

Group-level Effects	**SD**
Respondent ($n_{respondents}$ = 402)	
Intercept ($\alpha_{respondent}$)	0.079
Outcome ($n_{outcomes}$ = 9)	
Individual Deliberation ($\beta_{outcome}$)	0.051
Public Deliberation ($\gamma_{outcome}$)	0.044
Intercept ($\alpha_{outcome}$)	0.131
Question ($n_{questions}$ = 41)	
Individual Deliberation ($\beta_{question}$)	0.031
Public Deliberation ($\gamma_{question}$)	0.041
Intercept ($\alpha_{question}$)	0.163
Residual (σ^2)	0.015

n_{total} = 16510. See note on Table A3 for details on estimation and inference.

Table A7. Details on Conflict Avoidance Model.

Main Effects	**Estimate [95% Interval]**
Individual Deliberation (β_0)	0.037 [0.011, 0.063]
Public Deliberation (γ_0)	0.046 [0.023, 0.070]
Conflict Avoidance	–0.021 [–0.066, 0.025]
Conflict Avoidance × *Individual*	0.012 [–0.054, 0.071]
Conflict Avoidance × *Public*	–0.030 [–0.087, 0.021]
Intercept (α_0)	0.501 [0.484, 0.520]

Group-level Effects	**SD**
Respondent ($n_{respondents}$ = 402)	
Intercept ($\alpha_{respondent}$)	0.078
Outcome ($n_{outcomes}$ = 9)	
Individual Deliberation ($\beta_{outcome}$)	0.051
Public Deliberation ($\gamma_{outcome}$)	0.044
Intercept ($\alpha_{outcome}$)	0.131
Question ($n_{questions}$ = 41)	
Individual Deliberation ($\beta_{question}$)	0.031
Public Deliberation ($\gamma_{question}$)	0.041
Intercept ($\alpha_{question}$)	0.163
Residual (σ^2)	0.015

n_{total} = 16510. See note on Table A3 for details on estimation and inference.

Table A8. Details on Gender Model.

Main Effects	**Estimate [95% Interval]**
Individual Deliberation (β_0)	0.024 [0.001, 0.047]
Public Deliberation (γ_0)	0.072 [0.012, 0.124]
Female	–0.022 [–0.051, 0.008]
Female × *Individual*	0.019 [–0.020, 0.058]
Female × *Public*	–0.021 [–0.094, 0.052]
Prop. Women × *Public*	–0.051 [–0.156, 0.065]
Female × *Prop. Women* × *Public*	0.084 [–0.045, 0.204]
Intercept (α_0)	0.518 [0.501, 0.535]

Group-level Effects	**SD**
Respondent ($n_{respondents}$ = 402)	
Intercept ($\alpha_{respondent}$)	0.079
Outcome ($n_{outcomes}$ = 9)	
Individual Deliberation ($\beta_{outcome}$)	0.051
Public Deliberation ($\gamma_{outcome}$)	0.044
Intercept ($\alpha_{outcome}$)	0.131
Question ($n_{questions}$ = 41)	
Individual Deliberation ($\beta_{question}$)	0.031
Public Deliberation ($\gamma_{question}$)	0.041
Intercept ($\alpha_{question}$)	0.163
Residual (σ^2)	0.015

n_{total} = 16510. See note on Table A3 for details on estimation and inference.

Section A4. Alternative Regression Models

Table A9. Main Multilevel Model with Deliberation Group-Level Random Intercepts.

Main Effects	**Estimate [95% Interval]**
Individual Deliberation (β_0)	0.033 [0.015, 0.051]
Public Deliberation (γ_0)	0.059 [0.043, 0.075]
Intercept (α_0)	0.508 [0.495, 0.522]
Group-level Effects	**SD**
Respondent ($n_{respondents} = 402$)	
Intercept ($\alpha_{respondent}$)	0.077
Outcome ($n_{outcomes} = 9$)	
Individual Deliberation ($\beta_{outcome}$)	0.048
Public Deliberation ($\gamma_{outcome}$)	0.035
Intercept ($\alpha_{outcome}$)	0.098
Question ($n_{questions} = 41$)	
Individual Deliberation ($\beta_{question}$)	0.041
Public Deliberation ($\gamma_{question}$)	0.067
Intercept ($\alpha_{question}$)	0.031
Deliberation Group ($n_{deliberation\ groups} = 41$)	
Intercept ($\alpha_{deliberation\ group}$)	0.163
Residual (σ^2)	0.015

$n_{total} = 16510$. The table presents details on the following model:

$$y_i \sim N(\alpha_i + \beta_i\, Individual_i + \gamma_i\, Public_i,\ \sigma^2)$$

$$\alpha_i = \alpha_0 + \alpha_{outcome[i]} + \alpha_{question[i]} + \alpha_{respondent[i]} + \alpha_{deliberation\ group[i]}$$

$$\beta_i = \beta_0 + \beta_{outcome[i]} + \beta_{question[i]}$$

$$\gamma_i = \gamma_0 + \gamma_{outcome[i]} + \gamma_{question[i]},$$

which differs from the main model (Table A4) only insofar as it includes deliberation group-level random intercepts.

Figure A1. Replication of Figure 2 based on the Regression Model in Table A9

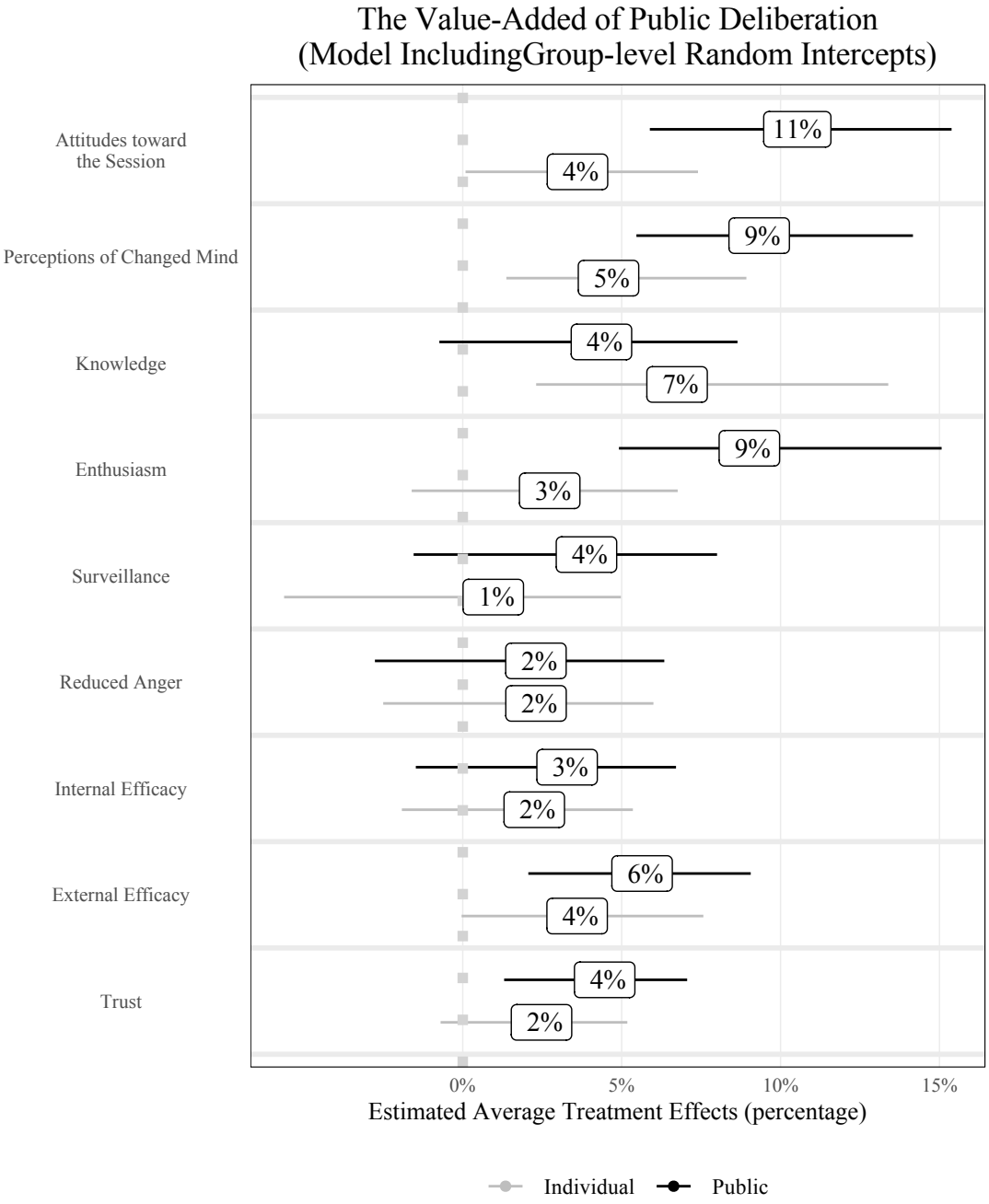


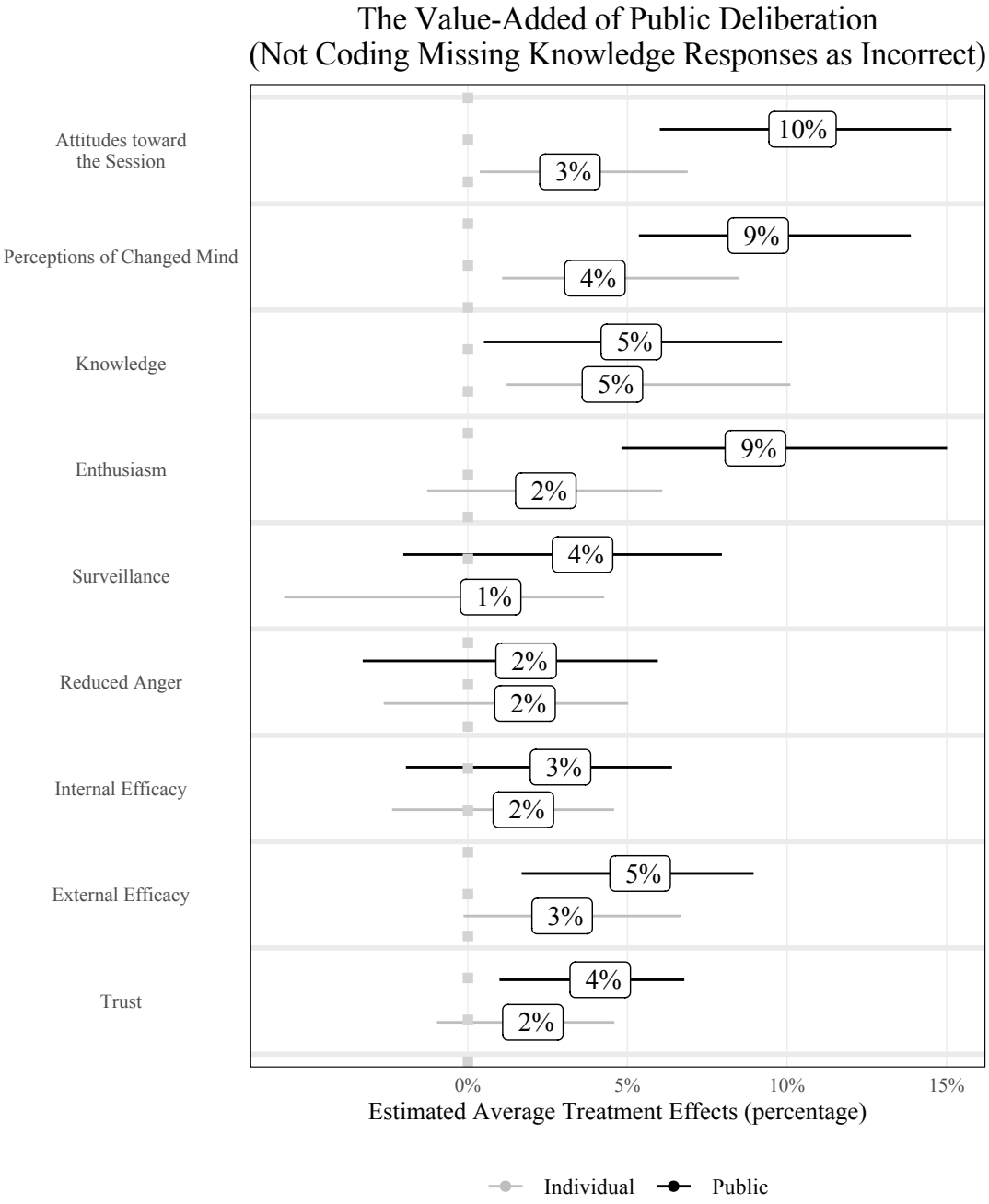
Table A10. Main Multilevel Model with Missing Knowledge Items Dropped.

Main Effects	**Estimate [95% Interval]**
Individual Deliberation (β_0)	0.026 [0.009, 0.045]
Public Deliberation (γ_0)	0.057 [0.042, 0.073]
Intercept (α_0)	0.531 [0.517, 0.543]

Group-level Effects	**SD**
Respondent ($n_{respondents}$ = 402)	
Intercept ($\alpha_{respondent}$)	0.076
Outcome ($n_{outcomes}$ = 9)	
Individual Deliberation ($\beta_{outcome}$)	0.053
Public Deliberation ($\gamma_{outcome}$)	0.040
Intercept ($\alpha_{outcome}$)	0.135
Question ($n_{questions}$ = 41)	
Individual Deliberation ($\beta_{question}$)	0.024
Public Deliberation ($\gamma_{question}$)	0.042
Intercept ($\alpha_{question}$)	0.160
Residual (σ^2)	0.015

n_{total} = 15888. The table presents the original main multilevel model, but drops all observations for which respondents did not answer a *Knowledge* question.

Figure A2. Replication of Figure 2 based on the Regression Model in Table A9



Section A5. Statistical Power and Minimum Detectable Effects

To get a sense of the statistical power of our experiment, we calculated minimum detectable effects (MDEs). That is, given our sample sizes and observed standard deviations for each of our outcomes, we calculated how big the "true" effect size would have to be for our study to have had 80% power to detect it. (SDs are calculated by averaging over all items in a question-group, effectively creating additive scales; see (2) below for more discussion of scale reliability.) This inversion of power analysis reveals whether our study demands unreasonably large effect sizes to yield highly reliable results. Our goal, then, is to evaluate whether the MDEs are overly large in magnitude, since bigger MDEs reveal limits in sample size. This approach also permits us to examine how an increase in sample size would alter these MDEs, giving us an idea of where on the "power curve" our sample size puts us.

The MDEs for *Individual* deliberation that yield 80% power given our sample sizes and observed SDs range from 4% to 10% (see Table A11). For context, observed estimates for that treatment range from 1% to 7%. Similarly, the MDEs for *Public* deliberation range from 3% to 8%, while observed estimates range from 2% to 11%. The difference in ranges of MDEs results from the difference in sample sizes in the treatment groups, with 102 assigned to *Individual* and 206 assigned to *Public*. But there is substantial overlap between the two sets of MDEs. For example, the two largest MDEs occur for *Knowledge Gains* in both treatments, owing to the higher SDs for that outcome. All other MDEs (for both treatments) were 7% or smaller. For comparison, if our sample sizes had been doubled, MDEs would only have dropped to the 2% to 7% range, with a max of 5% if we exclude the high SD *Knowledge Gains* outcome. We conclude that our study was not seriously underpowered—at least to estimate main effects. Sample sizes were sufficient to reliably detect effects of 3% to 7%.

Table A11. Minimum Detectable Effects for Additive Scales.

Outcome	*SD*	*Individual (n = 204)*	*Public (n = 300)*
Overall	0.37	0.07	0.06
Attitudes toward the Session	0.25	0.05	0.04
Perceptions of Changed Mind	0.37	0.07	0.06
Knowledge	0.50	0.10	0.08
Enthusiasm	0.33	0.06	0.05
Surveillance	0.34	0.07	0.05
Reduced Anger	0.26	0.05	0.04
Internal Efficacy	0.26	0.05	0.04
External Efficacy	0.29	0.06	0.05
Trust	0.21	0.04	0.03

Section A6: Scale Reliability and Separate Linear Models by Outcome

In this section, we first report α for each outcome (category of questions), and then report results of separate linear models by outcome. Finally, we display a version of Figure 2 based on these nine separate linear models.

Scale reliability as measured by α was as follows: *Attitudes toward the Session* (0.8), *Perceptions of Changed Mind* (0.7), *Knowledge* (0.4), *Enthusiasm* (0.7), *Surveillance* (0.8), *Reduced Anger* (0.9), *Internal Efficacy* (0.8), *External Efficacy* (0.7), and *Trust* (0.8).

To fit a separate linear model for each outcome, we first calculated the first principal component for each, scaled them to have SD = 1, and reoriented as necessary.* We listwise-delete any respondent missing at least one item per scale. We then regressed each scale on indicators for *Individual* and *Public*. The results are presented in Table A12.

Table A12. Separate Linear Models by Outcome.

Outcome Variable	*Individual* (β)	*Public* (γ)	*Intercept* (α)	*n*
Attitudes toward the Session	0.30 [-0.04, 0.64]	0.73 [0.45, 1.00]	-0.44 [-0.69, -0.22]	392
Perceptions of Changed Mind	0.34 [0.06, 0.67]	0.60 [0.35, 0.87]	-0.41 [-0.62, -0.20]	364
Knowledge	0.40 [0.09, 0.70]	0.31 [0.03, 0.59]	-0.23 [-0.44, 0.00]	400
Enthusiasm	0.14 [-0.17, 0.49]	0.44 [0.18, 0.71]	-0.28 [-0.51, -0.07]	396
Surveillance	-0.12 [-0.45, 0.20]	0.08 [-0.17, 0.34]	-0.02 [-0.23, 0.19]	396
Reduced Anger	0.04 [-0.26, 0.32]	0.03 [-0.23, 0.27]	-0.02 [-0.20, 0.19]	396
Internal Efficacy	0.00 [-0.28, 0.24]	0.08 [-0.13, 0.28]	-0.01 [-0.17, 0.14]	397
External Efficacy	0.20 [-0.19, 0.53]	0.31 [0.07, 0.55]	-0.20 [-0.41, -0.02]	397
Trust	0.02 [-0.37, 0.38]	0.15 [-0.09, 0.41]	-0.11 [-0.33, 0.08]	396

For comparison with our main multilevel model, Figure A3 replicates Figure 2 from the main text, but using the results from Table A12.

* The change from the effectively additive scales in the multilevel model to individual scales with SD = 1 means a change in the values of minimum detectable effects. However, they are now equivalent across all outcomes, within each treatment. Given our sample sizes, the MDE for *Individual* deliberation that yields 80% power is now 0.20 for all outcomes; and for *Public* deliberation, the MDE is 0.16.

Figure A3. Replication of Figure 2 based on the Regression Models in Table A12