

Can Deliberation Have Lasting Effects?

JAMES FISHKIN *Stanford University, United States*

VALENTIN BOLOTNYY *Stanford University, United States*

JOSHUA LERNER *NORC at the University of Chicago, United States*

ALICE SIU *Stanford University, United States*

NORMAN BRADBURN *NORC at the University of Chicago, United States*



Does deliberation produce any lasting effects? “America in One Room” was a national field experiment in which more than 500 randomly selected registered voters were brought from all over the country to deliberate on five major issues facing the country. A pre-post control group was also surveyed on the same questions after the weekend and about a year later. There were significant differences in voting intention and in actual voting behavior a year later among the deliberators compared to the control group. This article accounts for these differences by showing how deliberation stimulated a latent variable of political engagement. If deliberation has lasting effects on political engagement, then it provides a rationale for attempts to scale the deliberative process to much larger numbers. The article considers methods for doing so in the context of the broader debate about mini-publics, isolated spheres of deliberation situated within a largely non-deliberative society.


INTRODUCTION

There is a fundamental divide in democratic theory between “realist” approaches, which severely question the capacity of ordinary citizens to rule themselves, and deliberative approaches, which propose to rely on, or nurture, the capacity of citizens to make thoughtful and informed choices. We see this divide in Joseph Schumpeter’s contrast between the “classical” theory of democracy, in which citizens reason about the public good, and the modern “competitive” model (Posner 2005; Schumpeter 1942; Shapiro 2003), in which democracy is reduced to little more than a “competitive struggle for the people’s vote.” In that competitive struggle, advertising, manipulation of the public, and deception are all fair game as part of the competition (Schumpeter 1942, 263). The lack of meaningfulness of the public’s resulting


“volitions” is a central claim of the realist position. In this view, the point of democracy is to have peaceful transitions of power in a system that preserves rights. This division recurs throughout a vast literature, but most prominently of late, in the “realist” theory offered by Achen and Bartels (2016), as contrasted with what they call the “folk theory of democracy” in which informed citizens would supposedly make reasoned choices. For the “realist” theory, the “will of the people” is mostly chimerical or simply the product of various manipulative techniques, or the by-product of mobilizing partisan loyalties. Relying on the capacities of the public to make thoughtful decisions about the policies on offer, in this line of thinking, is a “pipe dream hardly worth the attention of a serious person” (Posner 2005, 163) and a “fairy tale” (Achen and Bartels 2016, 7).

Advocates of “deliberative democracy” can concede that the realists have a point about current democratic practices and voting behavior, but they hold out the prospect that deliberations by the people themselves, taking place under certain conditions, can be more reason-based. Is it the *capacities* of the public that are so limited or is it the nature of its *opportunities* within the current design of our democratic institutions and social practices? Perhaps voters are just not effectively motivated by the social context of being a citizen in mass society, subject to manipulative messages, incentives for rational ignorance, and a public sphere that seems to be decomposing into “filter bubbles” of like-minded “enclaves,” especially on contested issues (Chitra and Musco 2020; Dilko et al. 2017; Pariser 2011; Sunstein 2017; Spohr 2017; but see Zuiderveen et al. 2016 for skepticism). Under other possible conditions, could they perform a bit more like ideal citizens? This is not

Corresponding author: James Fishkin , Janet M. Peck Chair in International Communication and Director, Deliberative Democracy Lab, Stanford University, United States, jfishkin@stanford.edu.
Valentin Bolotnyy , Kleinheinz Fellow, Hoover Institution, Stanford University, United States, vbolotnyy@stanford.edu.

Joshua Lerner , Research Methodologist, NORC at the University of Chicago, United States, lerner-joshua@norc.org.

Alice Siu , Associate Director, Deliberative Democracy Lab, Stanford University, United States, asiu@stanford.edu.

Norman Bradburn , Senior Fellow, NORC at the University of Chicago, United States; Tiffany and Margaret Blake Distinguished Service Professor Emeritus, Irving B. Harris Graduate School of Public Policy Studies, Department of Psychology, Booth School of Business and the College at the University of Chicago, United States, Bradburn-Norman@norc.org.

Received: February 24, 2022; revised: November 05, 2022; accepted: November 30, 2023.

a utopian question. It is a question that can guide institutional designs for possible reforms. There are now experiments around the world with different designs for deliberative mini-publics that might foster reason-based decisions by random samples of citizens both on policies and on electoral choices (Fishkin 2018; Grönlund, Bächtiger, and Setälä 2014; Karpowitz and Raphael 2014).

There has long been speculation that certain kinds of political participation might make “better citizens” (Mansbridge 1999; Pateman 1970). But what kind of participation? In what respects might “better citizens” result? Most of the speculation has focused on participation in deliberative institutions such as the New England town meeting or the jury, institutions which John Stuart Mill called “schools of public spirit” (Mill [1861] 1991) when he reacted to de Tocqueville’s report of how these institutions operated in America. As de Tocqueville said, “Town meetings are to liberty what primary schools are to science; they bring it within the people’s reach” (de Tocqueville [1835] 2019, 73). Mill envisioned similar civic effects from service on juries and parish offices in England and, he speculated, from service in the deliberative institutions of ancient Athens. In all these cases, Mill argued, a citizen would be called upon “to weigh interests not his own; to be guided, in case of conflicting claims, by another rule than his private partialities; to apply, at every turn, principles and maxims which have for their reason of existence, the general good” (Mill [1861] 1991, 79). When the public is called upon to discuss what to do about public problems, they consider each other’s reasons and learn to take responsibility. This is essentially what Pateman termed the “educative effect” of political participation (Pateman 1970, 33) but focused specifically on participation in deliberative or discursive institutions.

What are the relevant dependent variables that might be affected by this kind of participation? Much of the focus has been on political efficacy (Morrell 2005), paying attention to public affairs (or knowledge gain about them) (Fishkin 2018) and on voting turnout (Gastil, Deess, and Weiser 2002; Gastil et al. 2010). For the latter, Gastil et al. found striking effects on voting turnout from participation in juries that reached a verdict (Gastil, Deess, and Weiser 2002; Gastil et al. 2010). All of these effects might plausibly fit an account of “better citizens.” Ideally, citizens will consider that they have views worth listening to (internal political efficacy), they will learn about public issues and pay attention to campaigns, and they will vote. In the ideal case, they would also make an explicit connection between their policy positions and whom they vote for.

These elements combine to fit the classical picture of voters taking their responsibilities seriously, becoming engaged and informed, having the sense of efficacy to do so and then voting based on their policy views. These are attributes of their civic capacities that speak directly to their ability to contribute to collective self-government. If deliberation were to facilitate voters behaving in this way, it would constitute a response to the realist critique, at least so far as that critique

depended on the capacities of voters rather than our current design of the institutions that engage them.

To offer a response to the realist critique, advocates of a more deliberative democracy need to face five related empirical questions:

- 1) Does deliberation in an organized setting demonstrate that ordinary citizens can come to reason-based and evidence-based judgments about what is to be done? One criterion for success on this score would be if citizens come to conclusions that clearly depart from simple partisan-based loyalties and show a judgment on the merits of the issues.¹
- 2) Can deliberation in such an organized setting have lasting effects? Or do the effects simply dissipate in the hothouse environment of political competition, campaigns, and elections?
- 3) If deliberation has lasting effects, is it primarily in the persistence of the post-deliberation policy attitudes or is it in the propensity to make reason-based choices, especially with respect to behaviors such as voting (both turnout and vote intentions)?
- 4) If deliberation has lasting effects on voting (either voting turnout and/or voting intention) are there identifiable mediators that are affected by deliberation that have these effects on voting?
- 5) Can deliberation in an organized setting, of the kind that offers encouraging answers to the first four questions, be scaled to large numbers beyond the random samples?

We approach questions 1 through 4 through a test case, a national experiment in deliberation. The experiment took place on the eve of the presidential primary season in 2019 and included a series of follow-up surveys with the deliberating sample and a control group. As for question 5, we have developed, based on this experiment, an approach that is now being piloted with new technology. We will sketch this approach at the end of the article. We report on our results in a separate paper (Fishkin et al. 2023).

AMERICA IN ONE ROOM

In collaboration with the Helena Group Foundation, we convened a national experiment in public deliberation about the major issues facing the United States in the period just preceding the 2020 presidential primary season. The event, entitled “America in One Room,” gathered a stratified random sample of 523 registered voters from around the country, recruited by NORC at the University of Chicago. A control group of 844 was also recruited by NORC and took essentially the same questionnaires in parallel with the experiment participants. The registered voter samples for both the participant and control groups were sourced from NORC’s

¹ In a separate paper we are doing qualitative and quantitative analyses of the reasoning in the transcripts that sheds light on the considered judgments.

probability-based and nationally representative AmeriSpeak panel. The recruitment and representativeness of the participant and control groups are discussed in the Supplementary material and in the Discussion. In advance of the initial survey, an advisory committee reflecting different points of view on the selected topics vetted the briefing materials for balance and accuracy. These materials serve as the initial basis for discussion when the sample is convened for its deliberations. The agenda focuses on policy options with balanced arguments for and against each option.

The first stage of the process resembles a normal public opinion poll: participants are surveyed with a standardized instrument in advance of seeing or discussing any information from the project. In the second stage, the random sample is brought together to a single place for extensive face-to-face discussions, usually over a long weekend. They are randomly assigned to moderated small group discussions, and they attend plenary sessions where they can pose questions to panels of experts or decision-makers with diverse views on a particular issue. At the end of the deliberations, participants take the same questionnaire as on the first contact, plus added questions for evaluation.

The participants gathered at a hotel in Dallas, Texas, on the weekend of September 19–22, 2019, arriving Thursday late afternoon and leaving Sunday after lunch. The agenda alternated small group discussions by issue area and plenary sessions, each lasting 90 minutes and running throughout the weekend. Each of the five issue domains (the economy, immigration, the environment, health care, and foreign policy) was discussed both in small group discussions and in plenary sessions with experts. Participants remained with the same small group (averaging about 13 persons) throughout the event, enabling them to get to know one another on a personal level over the course of the weekend. In the final questionnaire, completed just before departure, respondents were asked (as they had been in the pre-deliberation survey) to rate each specific policy proposal on a scale of 0 to 10, where 0 was “strongly oppose,” 10 was “strongly favor,” and 5 was a neutral midpoint.

Of the 47 proposals in these five issue areas, 26 can be classified as instances of extreme partisan polarization between Republicans and Democrats. The criteria are given as follows:

- a) At least 15% of respondents identifying with each party take the most extreme possible position (0 or 10) at time 1 (T1), with these Democrats and Republicans at opposite poles on the proposal.
- b) A majority of those party members who take a position at T1 are on the same side of the scale as the “extremes.”

These two criteria combine to identify extreme partisan polarization because the extremes are balanced at the two poles, with Republicans on one side and Democrats on the other.

Deliberation in this setting produced significant depolarization on 20 of 26 of the extreme partisan

issues. By depolarization, we mean that the means of the two parties move closer together. This does not necessarily mean that they both move toward the middle. They can both move in the same direction so long as they end up closer together. In a number of cases, the changes were massive, amounting to 40 percentage points for Republicans on the most hardline immigration questions and for Democrats on the most ambitious redistributive proposals. The control group changed hardly at all on these policy issues in the same period. The project included three follow-up waves: in July 2020 before the national party conventions, in late September/early October before the November 2020 Presidential election and, to capture self-reported actual voting, in the weeks following the election. All of these waves included both the treatment group (participants on the deliberative weekend) and the control group. We thus have data from T1 (before the weekend of deliberation), T2 (at the end of the weekend), T3 (10 months later, July 2020), T4 (October 2020, a year later, shortly before the presidential election), and T5 (after the election for self-reported recollection of voting. After the election, we also collected verified voting data from publicly available sources on the participants and the control group. The changes of opinion from T1 to T2 are the subject of Fishkin et al. (2021). This article focuses on the follow-up waves including T4 and T5.

To get a picture of the overall changes in policy attitudes, we use individual responses to the 26 extremely polarized issues to construct a policy-based ideology score (PBS). The PBS ranges from 0 to 10, with 0 denoting most liberal and 10 denoting most conservative. The score is constructed for each individual at T1 by averaging over their responses to each of the 26 questions. This process is similar to the weighted averaging of issues scales method advanced in Ansolabehere, Rodden, and Snyder (2008).² Before averaging, we make sure that the response to each question is converted to the PBS rubric (e.g., if the extreme Republican response to the question was 0, we flip everyone’s scores, turning 0 s into 10, 1 s into 9 s, and so forth before averaging). We do the same to construct the PBS using individual responses to the 26 questions at T2 and T3. The policy questions were not asked at T4 and T5.

It is worth emphasizing that the overall PBS is composed of policy scores in five different issue areas. The overall movement thus masks differential movement within the issue areas. For example, on the economy, participants moved significantly to the right on average. However, movements left on proposals in the other four areas outweighed that movement to the right in the

² We also explored using factor analytic methods as well as Poole’s (1998) basic space approach to estimate an analog ideology score to the PBS. When we ran those approaches, we found a Pearson correlation coefficient between each of those and the PBS to be 0.96 and 0.95 respectively. Because of the high correlations, we do not believe such differences in how the scale is constructed will impact results. We can share specifics of how we ran these robustness checks and the results upon request.

TABLE 1. Mean Policy-Based Scores (PBSs) by Issue Area across Time, Participant Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Time 1	Time 2	Time 3	Diff:	St. errors	Diff:	St. errors
Issue area	Mean	Mean	Mean	T2-T1	T2-T1	T3-T1	T3-T1
Immigration	4.34	3.17	4.01	-1.17***	0.076	-0.34***	0.057
Environment	4.04	3.91	4.12	-0.13**	0.067	0.07	0.057
Economy	4.26	4.79	4.34	0.53***	0.069	0.07	0.065
Health	4.03	3.79	3.81	-0.23***	0.075	-0.22***	0.067
Foreign policy	3.87	2.87	3.71	-1.00***	0.092	-0.16*	0.080
Overall	4.14	3.87	4.05	-0.27***	0.047	-0.09***	0.035
Observations	523	522	491				

Note: Standard errors are from paired t-tests of mean differences. The samples used for the paired t-tests are slightly different since individuals need to have taken both surveys, but this does not change the sample means reported here in meaningful ways. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

aggregation for the overall score. Depending on the issues selected, it is evident that deliberation can move people significantly in either direction.

As can be seen from Table 1, the PBS for all five issue areas moved significantly between T1 and T2. Four of the areas saw movement to the left on average, but one of the largest movements was to the right, on the economic issues (column 4). All five issue area changes between T1 and T2 were significant at the 0.01 level. For comparisons to the control group and a difference-in-difference analysis employing regressions, see Supplementary Tables A.1 and A.2.

Changes in the overall PBS are shown in a binned scatterplot in Figure 1.³ While the control group changes very little, there are large changes in the participant group especially, between T1 and T2. Clustered in the broad middle of the PBS range, individuals with more conservative initial positions showed negative movement (down in the chart and hence to the left politically), while those with more liberal initial positions showed positive movement (up in the chart and hence to the right politically). By T3, their positions reverted significantly in the direction of those they held at T1.⁴

In other words, the significant movements in the PBS between T1 and T2 reverted considerably nine months later (T3-T1). As Table 1 shows, there were still significant differences between T1 and T3 remaining but when compared to the control group movements from T1 to T3, the long-term effects of participation appear

to wash out (see Supplementary Tables A1 and A2). At the first glance, it would seem that deliberation does not produce many lasting effects. After all, these voters returned from their weekend of deliberation to the hothouse of an extremely polarizing and nasty campaign, one of the most polarizing in recent memory. It is hard to imagine sustained effects of a single weekend on participants nine months or a year after such a collective experience.

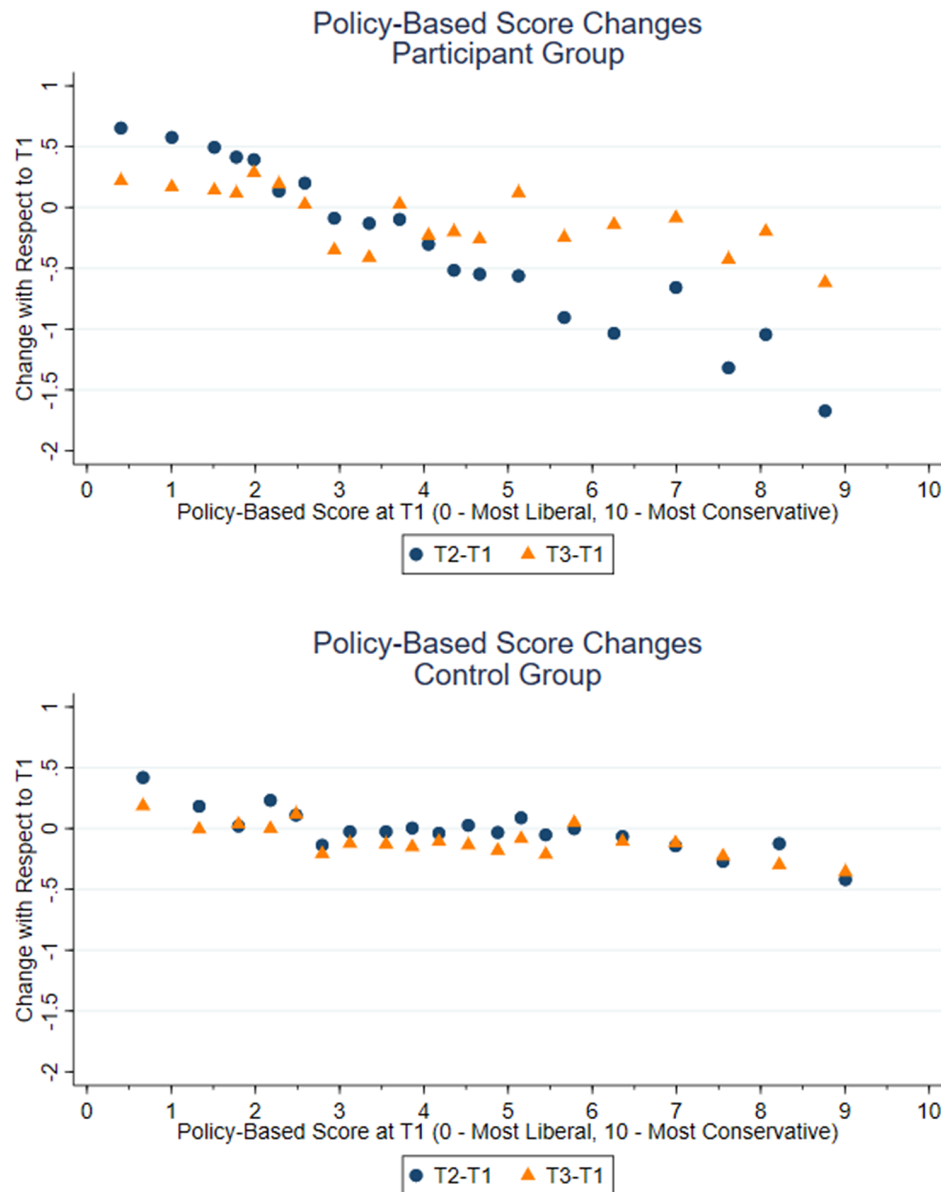
It is worth pausing to note that from the standpoint of deliberative mini-publics convened as a form of public consultation, it would not matter crucially if the results revert back nine months or a year after the deliberations have concluded. With public consultation of a stratified random sample, we are interested in what the public, in microcosm, thinks about a topic when the issues are fresh and when it has really engaged the competing arguments. Its considered judgments soon after deliberation can be taken, collectively, as a recommendation about what should be done. After that, memories fade. People return to their customary social networks and media habits. New news events offer added, and perhaps different or one-sided perspectives on the same issues discussed when the microcosm had deliberated. So reversion in whole or in part is to be expected and does not undermine the core purpose for which the microcosm was convened in the first place.

However, while reversion is not troubling for the core function of deliberative public consultation on a given set of issues, it is challenging for the broader aspiration, often shared among deliberative democrats, of somehow creating a more deliberative society. Unless a national mass deliberation were to be conducted soon before a national election (see Ackerman and Fishkin 2004 for one such scenario), it would seem to have little effect on collective self-government.

However, a more careful examination of the data collected following our national experiment offers a different picture. The follow-up surveys of treatment and control groups actually offer evidence of a significant effect on collective capacities for self-government,

³ A binned scatterplot puts the data into bins that contain equal numbers of data points. Each bin may cover different ranges of the x-axis variable. The data points are averaged on the y-axis variable and the average for each bin is displayed, making large data sets easier to visualize. Some charts also include linear fit lines or quadratic fit lines, all of which are estimated using the underlying data. See Stepner (2014) for more information.

⁴ There was a similar reversion in affective de-polarization nine months later. See the Supplementary Figures A1 and A2 showing changes in thermometer ratings in treatment and control groups. Political campaigns are known to increase affective polarization, so an intense campaign can be expected to re-engage the negative emotions on either side about the other party. See Sood and Iyengar (2016).

FIGURE 1. Policy-Based Score (PBS) Changes over Time

Note: Policy-based score (PBS) is constructed for each individual based on responses to 26 questions identified as the most polarizing. The upper chart shows the participant group, and the lower chart shows the control group. T1 is the survey wave prior to the deliberations, T2 is right after the deliberations, and T3 is 10 months after, in July 2020.

resulting in a more optimistic picture. We say “optimistic” not because of any partisan implications. In a different election, the positions of the two parties could easily be reversed. Rather, we say “optimistic” because of the effect on the civic attributes of voters that deliberation appears to stimulate.

THE PUZZLE OF LASTING EFFECTS

The follow-up surveys just before the 2020 election as well as just after the election with both the participants and the control group show significant

differences in voting behavior for these samples of registered voters.⁵

⁵ Setting aside the 5% or so who intend to vote for a third party or not vote at all, the total percentage of the control group saying they intend to vote is 86% and the total percentage of the deliberators saying they intend to vote is 91%. At first glance these numbers may seem high. For perspective, the actual percentage of registered voters in the population who turned out to vote in November 2020 was 86.3%. Thus, our vote intention numbers are within a reasonable range. Note these calculations are percentages of registered voters, not of the voter eligible population. See Reuters Staff (2021) for more information.

TABLE 2. Voting Intention for Participant and Control Groups, Time 4

Vote intention overall				
If the presidential election were held tomorrow, which candidate would you vote for (or already have voted for)?				
Sig: .000	Participants		Control	
	Frequency	Percent	Frequency	Percent
Republican candidate Donald J. Trump	145	31.3	290	40.8
Democratic candidate Joseph R. Biden	275	59.5	317	44.6
Won't vote or vote for third party or write-in candidate	24	5.2	34	4.8
Prefer not to answer	2	0.4	16	2.2
Don't know	16	3.4	48	6.8
SKIPPED ON WEB	1	0.2	5	0.7
Total	463	100	710	100

Table 2 shows the dramatic difference between the treatment and control groups in voting intention just before the election, a full year after the deliberative weekend. The control group had a gap between Joseph R. Biden and Donald J. Trump at 3.8% (the actual gap in the electorate was about 3%). But the voting intentions of the participants suggest a dramatic effect of the treatment—a gap of 28.2 percentage points between the two major candidates.⁶

How is such an effect possible? The results are surprising because the accepted wisdom in political science has long been that voting behavior, deeply rooted in group attachments, is much more stable, and is presumably much harder to change than political attitudes (Achen and Bartels 2016; Campbell et al. 1960; Green, Palmquist, and Schickler 2004). We find significant effects on two aspects of voting behavior: who one votes for and whether one votes at all. The second is just as puzzling as the first in that successful interventions on turnout tend to be soon before the election (e.g., Green and Gerber 2019). All of the effects on voting behavior discussed here occur much longer after the intervention. How can deliberation possibly have such effects almost a year later?

DELIBERATIVE DEPARTURES FROM PARTISAN LOYALTIES: WHO WAS AFFECTED?

We can begin to explore these differences between treatment and control groups first by looking at who was different in voting behavior between the two groups at election time. Second, we will do predictive modeling to indicate who departed from the voting behavior that would normally be predicted by standard demographics (including party ID). Then, we will turn to causal mediation analysis to explore the effects of a latent variable, which we term a civic awakening) that

helps to further explain the voting behavior of the treatment group.

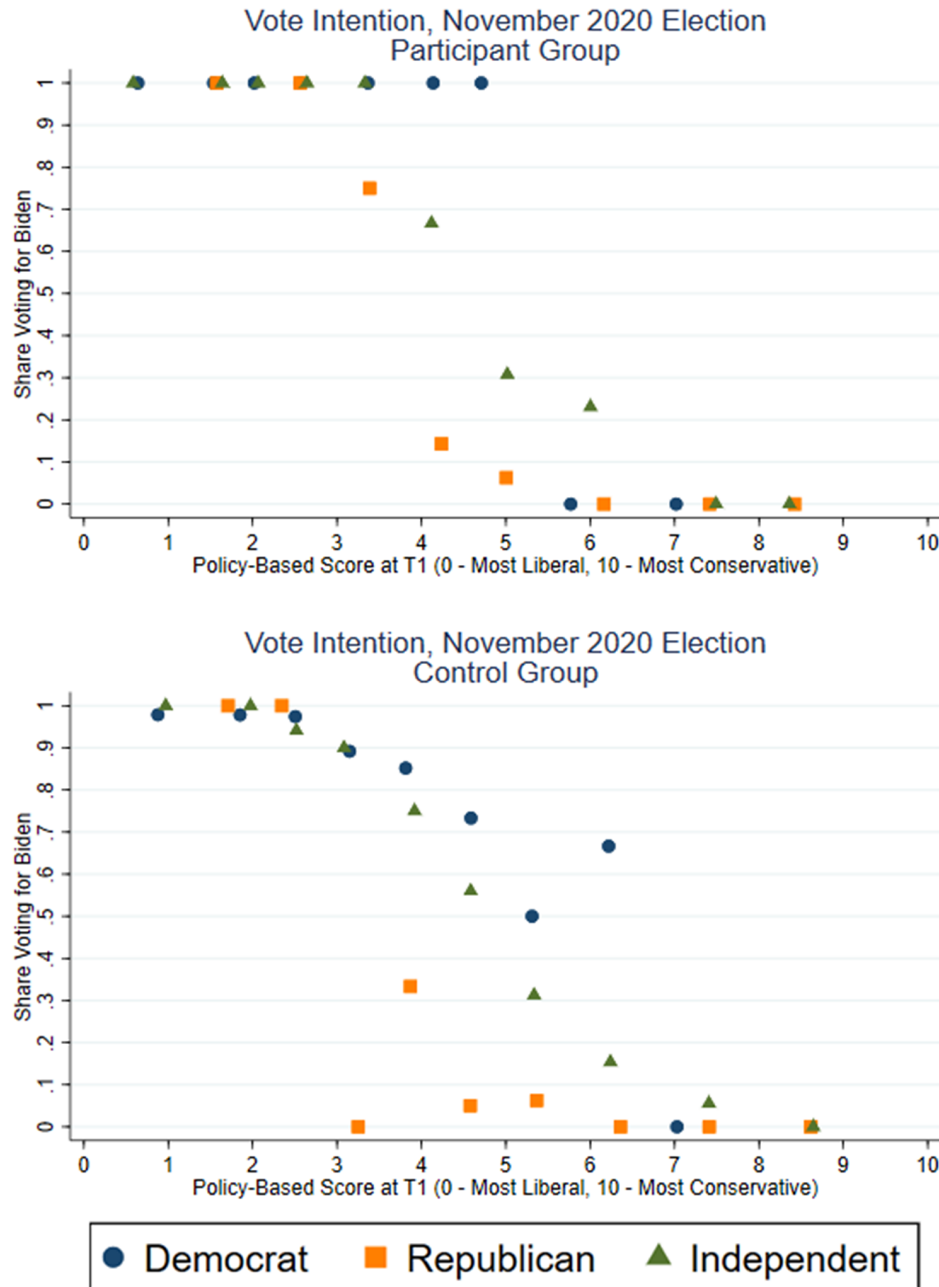
Not only are Democratic participants in the middle (roughly 3–5) PBS range more likely to intend to vote for Biden than those in the same range in the control group, but Independents and Republicans are as well. Figure 2 illustrates this finding with a binned scatterplot that breaks participants and control group members out by party. Democrats and Independents who start off in the middle group at T1 are especially more likely to intend to vote for Biden than those in the control group by T4. Self-described Republicans who, in fact, hold somewhat left-leaning policy positions at T1 are also more likely to vote for Biden at T4 than comparable individuals in the control group.

In addition to vote intention for Biden, the middle group of participants also sees the biggest effect on intention to vote at all. Figure 3 demonstrates this fact, comparing those in the middle group to those outside of the middle group at T1 and participants to control group members. The figure provides support to our view that individuals in the middle group, there are marginal voters whose political engagement has the capacity to be especially awakened by democratic deliberation.

To further investigate individuals in the middle group, we build a model designed to predict vote intention based on the control group and then see where the model performs poorly when applied to the participants.

We take our control group sample and run a probit regression of the control group's voting intentions (1 for Biden, 0 for Trump) at T4 on their characteristics at T1. The regression includes the following explanatory variables: education, gender, age, race, marital status, employment status, income level, home ownership status, metro/rural area of residence, feelings towards Republicans, feelings towards Democrats, opinion of Trump, opinion of Biden, PBS, ideological self-assessment score, and political party. The pseudo R-squared for a probit regression that includes just the demographic variables is 0.11; the pseudo R-squared for a regression that has all the variables above is 0.86.

⁶ For a similar table with recollected vote after the election, see Supplementary Table A7.

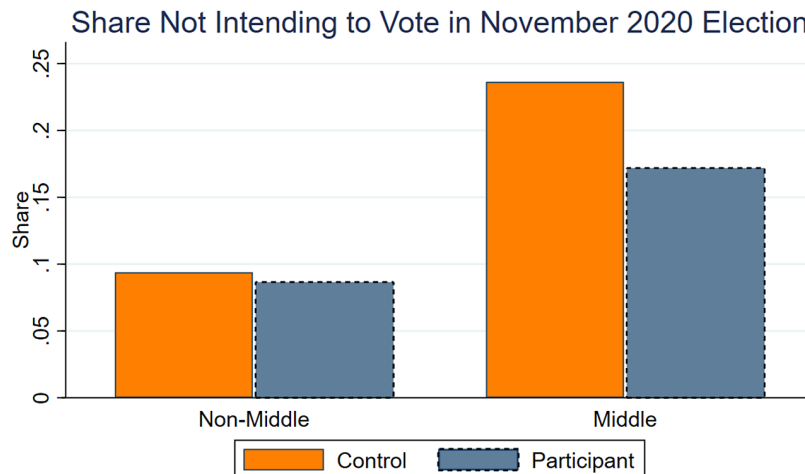
FIGURE 2. Vote Intention for Biden at T4

Note: Policy-based score (PBS) is constructed for each individual based on responses to 26 questions identified as the most polarizing. Not intending to vote for Biden means intending to vote for Trump or someone else. Vote intention data were collected in October 2020.

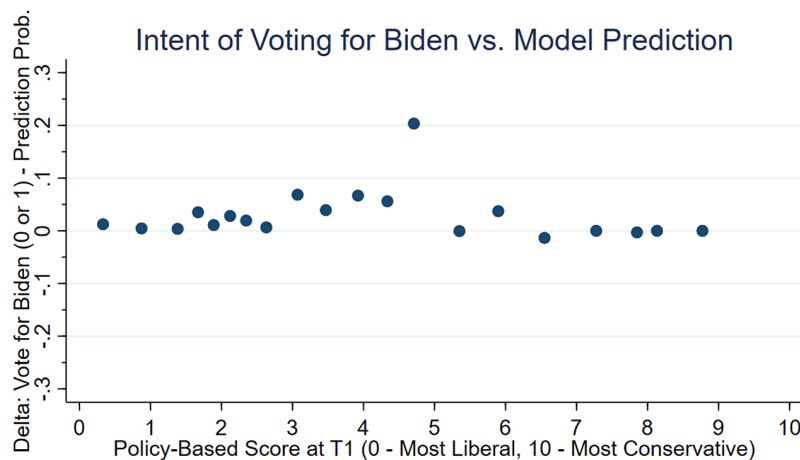
We then take the estimated model (with all the variables) and use it to predict the participants' intent to vote at T4 based on their T1 characteristics. We calculate the delta between actual vote intention and the model's predictions. To do this, we take the vote intention a participant's reports at T4 (1 = vote for Biden, 0 = vote for Trump) and subtract the model's prediction for that participant (a probability of voting for Biden that ranges from 0 to 1). Thus, if a person's delta is positive, the probability that he/she will vote for Biden is higher in

reality than the model would predict. If the delta is negative, the probability that he/she will vote for Biden is lower in reality than the model would predict.

In Figure 4, we plot the binned averages of the deltas on the y-axis against the participants' Time 1 PBS on the x-axis. The model error is close to zero for most participants, except for those who are in the middle group of the PBS, in the 3–5 score range. The errors start to pick up a bit (meaning the likelihood that the participant will vote for Biden is higher in reality than

FIGURE 3. Vote Intention by Participant Status and PBS (at T1)

Note: Middle are those who have Policy-Based Scores between 3 (inclusive) at Time 1. Non-middle are all other participants. Participants in the middle group are 6.4 percentage points (8.4%) more likely to intend to vote than control group members in the middle group. Standard error on the difference is 0.045, so difference is not statistically significant. 76.4% of control middle group intends to vote.

FIGURE 4. Effects on Vote Intention Captured by Predictive Modeling

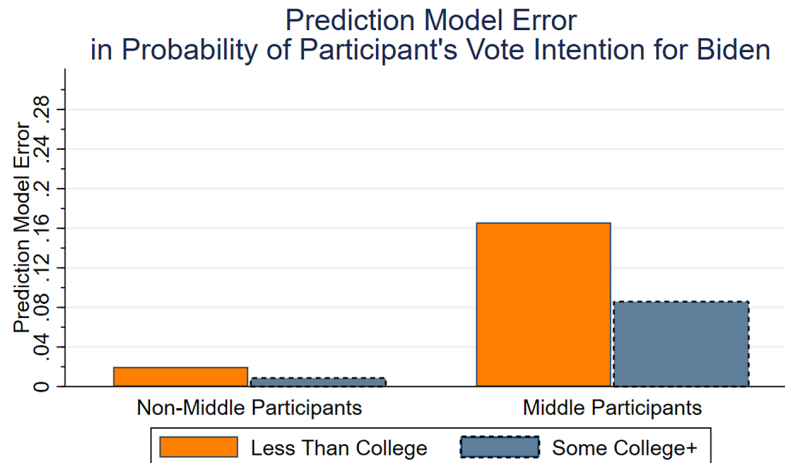
Note: Policy-Based Score is constructed for each individual based on responses to 26 questions identified as the most polarizing. A positive delta value means that the participant is more likely to vote for Biden than predicted by the model. A negative delta means that the participant is less likely to vote for Biden than predicted by the model. Probit model is estimated using Time 1 control group characteristics and predictions are made for participants based on their Time 1 characteristics. Vote intention data are collected at Time 4, in October, 2020. Full calibrated model used to construct this figure can be found in the APSR Dataverse.

predicted by the model at T1) for those with PBS between 3 and 4, and really shoot up for those with PBS between 4 and 5.

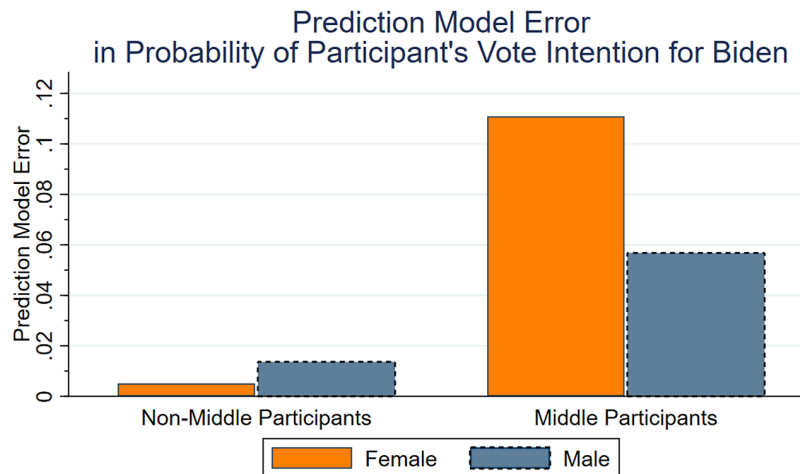
Looking at model prediction errors by demographic characteristics, we find evidence that the participants driving differences in vote intention between the participant and control groups are those without a college degree (Figure 5), especially in the middle of the PBS range. These findings corroborate our evidence that the participants who saw the largest lasting increase in

political engagement as a result of deliberation are individuals who came into the experiment with the lowest levels of political knowledge. Figure 6 also demonstrates that women in the middle range of the PBS were disproportionately affected by the deliberations.

The first issue to examine is whether or not the apparent difference is the result of differential attrition or some other distortion in the composition of the participant and control groups a year or more later (T4 and T5) compared to the way they matched up at T1.

FIGURE 5. Effects on Vote Intention Captured by Predictive Modeling, by Education

Note: Middle are those participants who have Policy-Based Scores between 3 and 5 (inclusive) at Time 1. Non-middle participants are all other participants. Positive prediction error shows that, on average, participants were more likely to vote for Biden than predicted by the model. Vote intention data are collected at Time 4, in October, 2020. Full calibrated model used to construct this figure can be found in the APSR Dataverse.

FIGURE 6. Effects on Vote Intention Captured by Predictive Modeling, by Gender

Note: Middle are those participants who have Policy-Based Scores between 3 and 5 (inclusive) at Time 1. Non-middle participants are all other participants. Positive prediction error shows that, on average, participants were more likely to vote for Biden than predicted by the model. Vote intention data are collected at Time 4, in October, 2020. Full calibrated model used to construct this figure can be found in the APSR Dataverse.

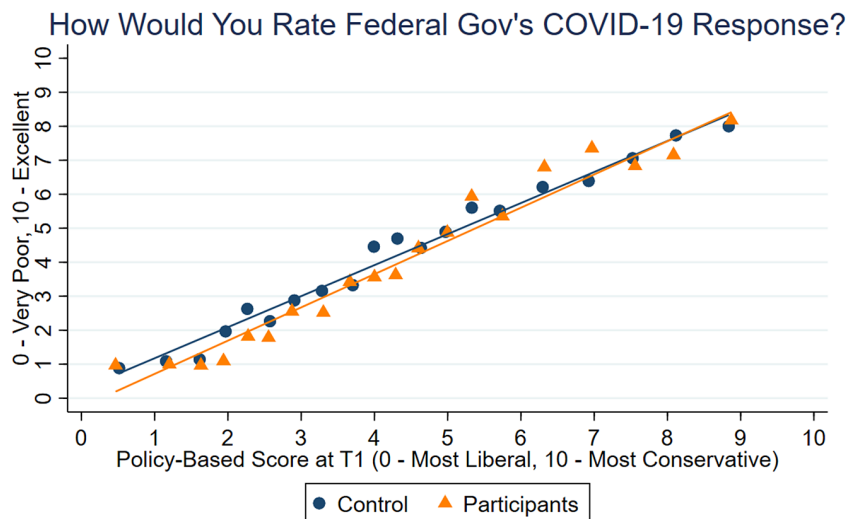
Table 3 shows that the average differences in characteristics of the control group and the participant group did not change significantly across time. Differences in the averages between participant and control groups for all the standard demographics (as well as party ID) are stable across the various waves. The dramatic differences in voting intention we saw in Table 2 thus cannot be attributed to differential attrition in either the participant group or the control group in the survey waves collected either before or after the election.

Even though the participant group did not experience differential attrition compared to the control group, one might wonder if they started out more knowledgeable and more oriented to discussion with diverse others. Perhaps starting as more amenable to diverse civic dialogue they were more easily activated by the process. However, on questions of general knowledge, there were no significant differences between the participants and the control group on five out of seven of the questions on general political knowledge at time 1 (see Supplementary Table A8).

TABLE 3. Balance Table Showing Differences in Means between Participant and Control Groups

	(1)	(2)	(3)	(4)
	Mean diff control-treatment	Mean diff control-treatment	Mean diff control-treatment	Mean diff control-treatment
	T1 sample	T3 sample	T4 sample	T5 sample
Less than HS/HS	0.11***	0.11***	0.12***	0.12***
Some college	-0.04	-0.04	-0.05*	-0.05
BA or higher	-0.06**	-0.07**	-0.06**	-0.08***
Male	0.03	0.04	0.06*	0.05
Over 60	0.02	0.02	0.02	-0.02
White	0.08***	0.08***	0.07***	0.03
Married	0.06**	0.05*	0.06**	0.02
Employed	0.01	0.01	0.02	0.02
HH Inc. <\$60,000	0.02	0.03	0.02	0.02
Homeowner	0.09***	0.08***	0.07***	0.07***
From non-metro	-0.00	0.01	-0.00	-0.01
Share Democrats	-0.03	-0.03	-0.03	-0.02
Share Republicans	0.05*	0.04	0.05	0.04
Share Independents	-0.02	-0.01	-0.02	-0.02
Policy-based score, T1	0.33**	0.34**	0.36***	0.27*
Observations	1367	1267	1173	1150

Note: Table shows differences in average characteristics of the control group and the treatment group at each time period of the study. T1 is just before September 2019, before the deliberations; T2 is just after the deliberations (not shown because the sample is the same as at T1); T3 is 10 months later, in July 2020; T4 is October 2020; T5 is November–December 2020, after the general election. All characteristics are based on data collected at T1. The number of observations in each sample includes the participant and control group samples. For full balance tables for each time period of the study, please see the Appendix Tables A3–A6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

FIGURE 7. Evaluating the COVID-19 Pandemic Response

Note: Policy-Based Score is constructed for each individual based on responses to 26 questions identified as the most polarizing. Question assessing federal government's response to the pandemic was asked at Time 4 (October, 2020).

On questions about others “who disagree with you strongly,” such as whether they “have good reasons” or whether “they are thinking clearly,” the participant and control groups show virtually no difference at time 1. This is also true for the time 1 views of the sample who answered the voting intention questions at time 4 (see Figure A2 in the Appendix). We take these questions as an approximate measure of the

predisposition to engage with those with whom you most strongly disagree. There does not seem to be any difference in this predisposition between treatment and control groups nor any strong indication that the participants started out as more informed citizens.

Another potential explanation stems from the global COVID-19 pandemic and the Trump administration's

response, both of which occurred during the T3-T5 survey waves. One might anticipate that participants in the deliberations might have perceived the impact of the pandemic—and the public policy responses to the pandemic—differently than the general populace, which would have given the participants a more negative view of the Trump administration. To explore this possibility, Figure 7 looks at respondent PBS against their assessment of the federal government's COVID-19 response, broken down by the participant and control groups. What we see in the figure is that when accounting for pretreatment policy positions, there is no difference at all between the two groups. This would indicate that, whatever effect the deliberations had, it was not primarily through differential changes in perception of the Trump administration's performance during the COVID-19 pandemic. This makes sense in that the effects of the pandemic should not have been localized or uniquely felt by participants; any effects that COVID-19 had on the election were likely homogeneous between the participant and the control group. COVID-19 was everywhere.

SOLVING THE PUZZLE

Our proposed solution to the puzzle of the delayed effect is that the deliberations gave rise to a latent variable, which might be termed *an awakening of civic capacities*, that has an effect, in turn, on voting (whether or not one votes at all) and on vote choice (whom one intends to vote for). The people who deliberated over the weekend, as compared to the control group, became more politically engaged. We take significant movement on the PBS over the weekend as an indicator that they were deeply involved in the deliberations. Those who deliberated were also more likely to follow the campaign, have a greater sense of internal efficacy (belief that their political views were “worth listening to”),⁷ and acquire (and continue to acquire) general political knowledge. Deliberative change on the issues, following the campaign, feeling that you have views worth listening to, and becoming more knowledgeable are all elements of a coherent civic awakening—a picture of more engaged citizens.

These elements of a civic awakening are roughly similar to those found by Gastil, Deess, and Weiser (2002) in their study of the indirect effects on voting from serving on a jury that reached a verdict.⁸ They

found that the depth of deliberation (which they measured by the number of counts considered at a trial that reached a verdict) was one of the mediators in increasing the likelihood of voting. Our deliberators all considered the same number of policy issues, but we measure “depth of deliberation” through the opinion changes on the PBS score on the deliberative weekend. Gastil et al. (2010) also found public affairs media use as measured by “following the campaign,” political efficacy and satisfaction with the deliberative process⁹ all connected to the civic awakening from jury service. We use “following the campaign” and gain in general political knowledge as mediators along with internal political efficacy.

In the jury case, the dependent variable was limited to whether or not one voted. In our analysis, we are interested both in turnout and in *how* one voted and whether that voting behavior has a connection to one's policy positions. The latter is essential for considering the broader question of the impact of civic engagement on collective self-government.

These elements of a civic awakening are most simply captured graphically. First, we saw that the treatment stimulated significant policy change on the issues (Figure 1 and Table 1). These changes indicate who engaged in the deliberations to the point of changing their opinions significantly on the most contested issues. Second, the participants were more likely than the control group to say at T3 that they are “closely following the campaign”. Figure 8 shows how this difference is mostly (but not entirely) clustered around the moderate middle range of the policy score (based on the T1 scores).

Third, the deliberators show an increase in internal efficacy or self-efficacy between T1 and T3. They are more likely to think, at T3, that their opinions are “worth listening to” compared to the responses from the control group. Again, as pictured in Figure 9, these differences are clustered mostly, but not entirely, in the broad middle of the policy score range (based on their scores at T1).

Fourth, we have a measure of general political knowledge on items that were not explicitly the subject of the deliberations (who controls the House and who controls the Senate). This measure did not increase right after deliberation, at T2, but was significantly higher for participants at T3 (Figure 10). This suggests that 10 months after deliberation, the participants were obtaining general political knowledge *on their own*, a sign that the civic awakening that occurred during the course of deliberation is manifesting itself in lasting ways.

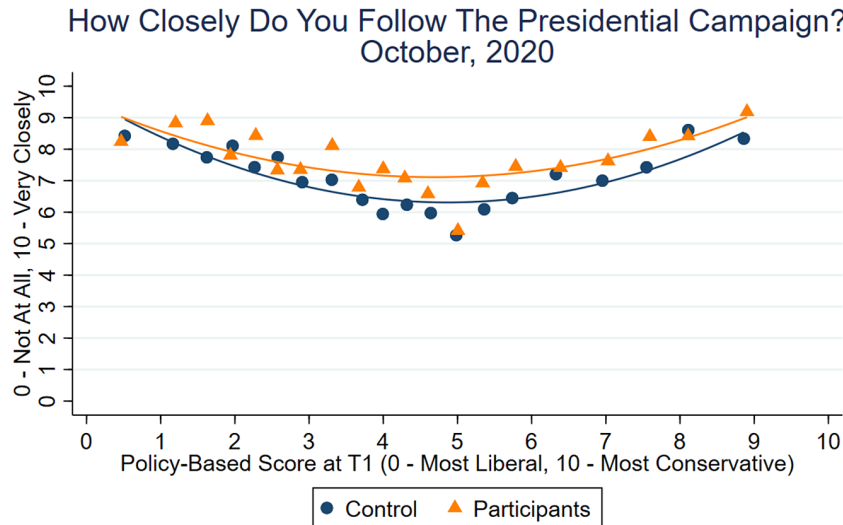
Let us review these aspects of the civic awakening and note how they are distributed in the policy space.

⁷ We have focused on internal or self efficacy rather than external efficacy for a long-term effect as one's sense of developing “opinions worth listening to” is not dependent on the contested and changing political context of a hotly contested campaign. After a year of no-holds-barred campaigning, it is hard to imagine citizens agreeing that “Public officials care about what people like me think” (the standard external efficacy question that we included.) But it is possible to imagine that deliberators might continue to believe “I have opinions about politics that are worth listening to” (a standard measure of internal efficacy that we included.)

⁸ These are also suggestive connections to the construct of “political capital” in Jacobs, Cook, and Delli Carpini (2009) which includes political efficacy, political attention and general political knowledge

among other variables. But this is from a single cross-sectional survey, not an experiment (exploring the effects of attending public meeting).

⁹ Our battery of evaluation questions of the deliberation has a median close to 10 on a 0 to 10 scale. There is not enough meaningful variation to employ them in the analysis. Furthermore, we only have evaluation questions from those who took part in the deliberations. We have no such data from the control group (which would have been required to include them in the causal mediation analysis.)

FIGURE 8. Following the Campaign

Note: Policy-Based Score is constructed for each individual based on responses to 26 questions identified as the most polarizing. Responses to the question “How closely do you follow the presidential election campaign?” were collected in October, 2024 (T4).

First, as we see in Figure 8, there are clear differences between the participant and control groups in how closely the respondents are following the 2020 election, regardless of the PBS. We also see a similar relationship in Figure 9 for respondent’s self-reported beliefs in the value of their own political opinions: again, participants were more likely to believe that their political opinions were “worth listening to” even at T3, a persistent change long after the deliberative weekend. The deliberative treatment also affected general political knowledge. It increased between T2 and T3, throughout the course of the election. Once again, the increases are clustered in the broad middle of the policy space and show a large difference between participant and control groups. For details on the two general political knowledge questions (as well as the policy-specific knowledge questions) and how they compared to responses from the control group at T1, T2, and T3, see Supplementary Table A.8, Panels A, B, and C.

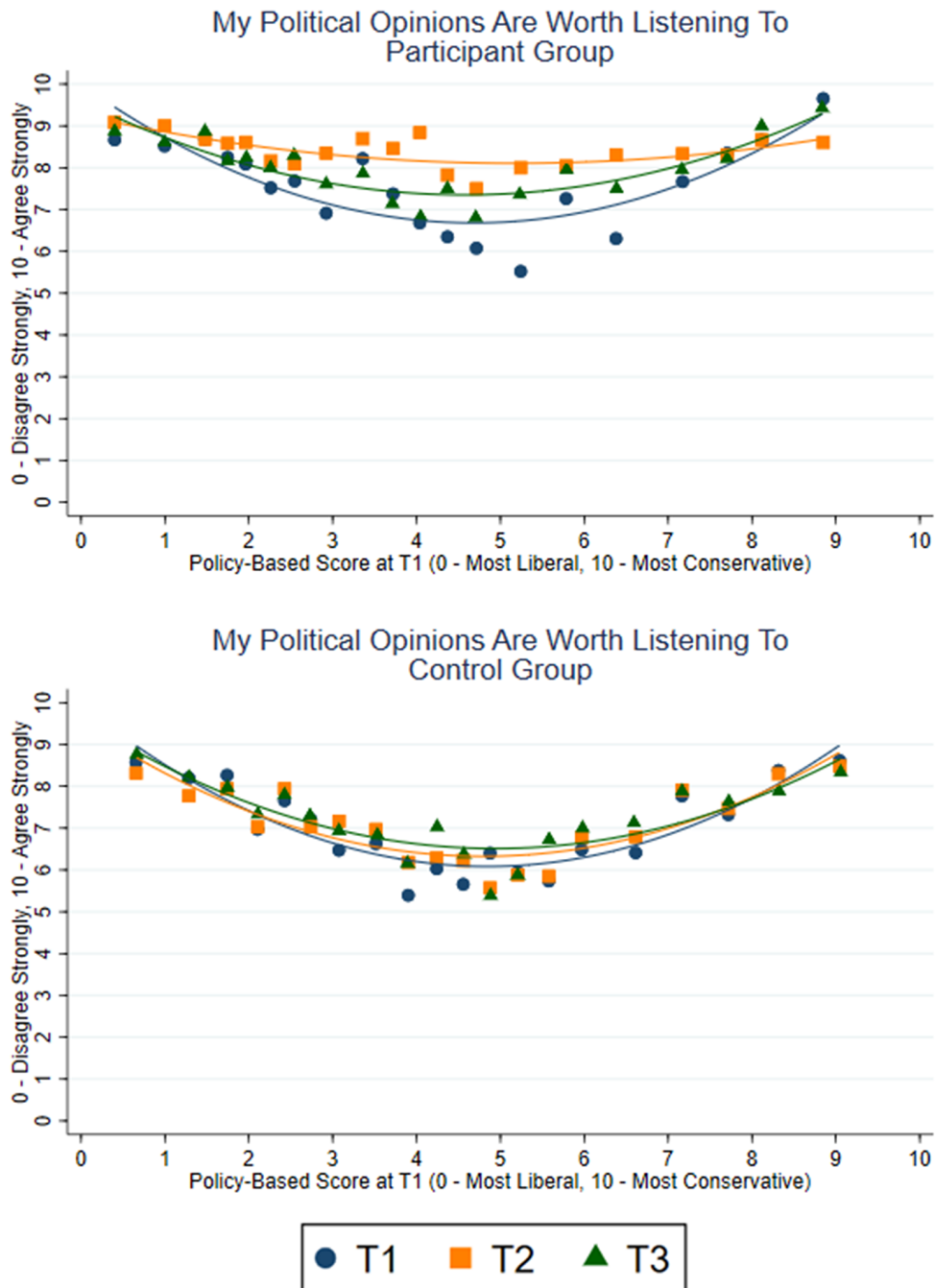
To summarize, we believe the following elements of civic awakening serve as mediators for whether citizens will vote at all and whom they intend to vote for: a) changes in the PBS over the weekend of deliberation (Table 1 and Figure 1); b) closely following the campaign (Figure 8); c) having “opinions worth listening to” (Figure 9); and d) general political knowledge (Figure 10). We will employ causal mediation analysis to demonstrate the effect of these variables on voting.

First, a few words on how we conceptualize the civic awakening. Our theory of measurement requires us to differentiate between two types of measures: “reflective” and “formative” (Sokolov 2018; Stenner, Burdick, and Stone 2008; Trochim 2001). A formative measure requires knowing all of the factors that make up a construct and including measures for all of these components.

A classic example of a formative measure is socioeconomic status, which is defined as the combination of education, income, and occupational prestige (Auerbach, Lerner, and Ridge 2022). If one part is not included, then the index would be incomplete and not measure socioeconomic status. Reflective measures examine multiple outputs of a force and use latent trait modeling to identify this force in these results. For instance, intelligence is a latent ability assessed through various types of tests. IQ tests take test question responses as reflections of an individual’s underlying ability. In this case, there is no complete corpus of intelligence components to be assembled (Coltman et al. 2008).

We view the civic awakening as a formative construct: we believe that it is a combination of individuals propensity to follow the campaign, to feel their opinions are worth listening to, to become knowledgeable about politics, and to have deliberated in depth (measured by whether attending the deliberation caused a general shift in their underlying political attitudes [$PBS_{T1} - PBS_{T2}$]). If we treated this as a singular reflective measure, we would want to study their underlying correlation matrix and use methods that exploit similar covariance between the variables (like factor analysis). Because we believe civic awakening is a latent combination of these observable factors, such tests are inappropriate, though there are modest correlations between “follows the campaign,” “having opinions worth listening to,” and knowledge (with Pearson’s correlation coefficient ranging from 0.24 to 0.37 between the three).

Fundamentally, we believe that these four indicators are evidence of the formative construct of an individual’s unobservable civic engagement. We choose to keep these as separate indicators, rather than utilize

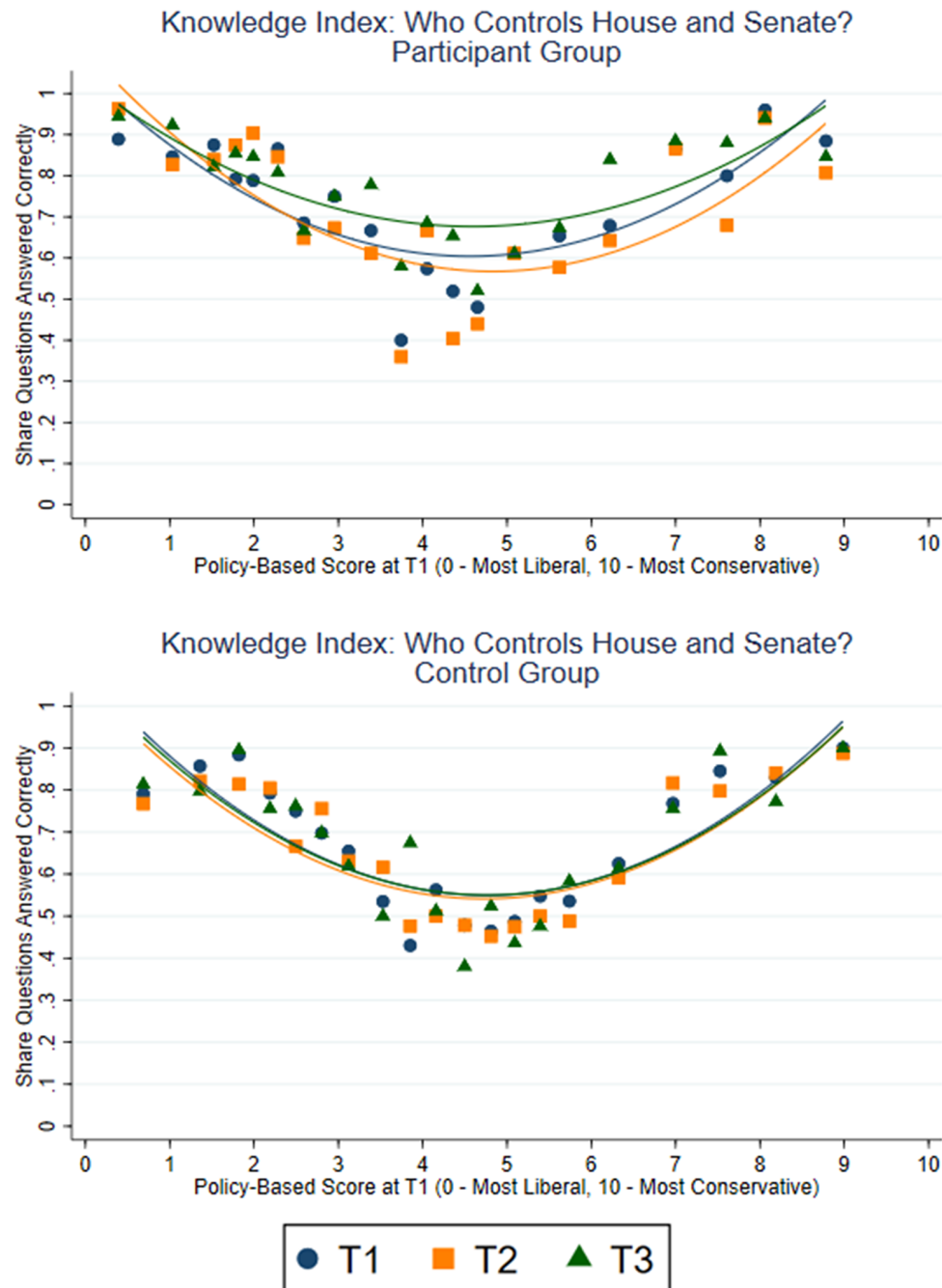
FIGURE 9. Having “Political Opinions Worth Listening to”

Note: Policy-based score is constructed for each individual based on responses to 26 questions identified as the most polarizing. Responses to the question “How strongly would you disagree or agree with the following statement?” [I have opinions about politics that are worth listening to.] were collected at T1 (just before deliberations), T2 (just after), and T3 (10 months later, July 2020).

an aggregation strategy, because keeping these separate makes the results of our treatment on each individual indicator clear, showing that certain indicators only affect certain outcomes. Aggregation would lose specificity that makes our overarching story much clearer.

CAUSAL MEDIATION ANALYSIS: ESTIMATING DIRECT AND INDIRECT EFFECTS OF DELIBERATION

The traditional method of exploring relationships between a treatment and outcomes is by using a

FIGURE 10. General Political Knowledge

Note: Policy-based score (PBS) is constructed for each individual based on responses to 26 questions identified as the most polarizing. Y-axis reports the average share of people correctly answering the questions: “Which political party holds the majority in the Senate?” and “Which political party holds the majority in the House?” Those who select Democrats, Independents, or say they do not know for the Senate are coded as not knowing the correct answer; those who select Republicans, Independents, or say they do not know for the House are coded as not knowing the correct answer. T1 is just before the deliberations (September 2019), T2 is just after, and T3 is 10 months later, in July 2020. The upper chart shows the participant group, and the lower chart shows the control group.

regression model. However, this method fails to disentangle underlying causes and effects that are indirect, rather than direct. In our case, we know that there is an effect of participating in the deliberations on an

individual’s propensity to vote; this effect is perceivable even a year after the event. It is, however, unsatisfactory to state that participating in the deliberations is the direct cause of an increased propensity to vote and to

vote in particular ways: surely, there were intermediate steps caused by the deliberations that, when taken together, affect these outcomes.

When faced with the possibility of indirect effects, investigators may have prior knowledge that an explanatory variable plausibly exerts its effect on an outcome via direct and indirect pathways. In the indirect pathway, there exists a *mediator* that transmits the causal effect.

Suppose we have variables T and Y indicating the treatment variable and outcome variable, respectively. Mediation in its simplest form involves adding a mediator M between T and Y . The sequential ignorability assumption, critical to causal mediation analysis, states that the treatment (explanatory variable T) is first assumed to be ignorable given the pretreatment covariates, and then the mediator variable (M) is assumed to be ignorable given the observed value of the treatment as well as the pretreatment covariates (Imai, Keele, and Tingley 2010; Imai et al. 2011).

The first part is often satisfied by randomization, while the second part implies that there are no unmeasured confounding variables between the mediator and the outcome. The standard mediation analysis starts with three equations, usually modeled with continuous outcomes (though advances in methods now allow for most parametric modeling approaches for either stages of the mediation):

$$Y = i_1 + cT + e_1 \quad [1]$$

$$Y = i_2 + c'T + bM + e_2 \quad [2]$$

$$M = i_3 + aT + e_3 \quad [3]$$

where i_1 , i_2 , and i_3 denote intercepts; Y is the outcome variable; T is the treatment variable; M is the mediator; c is the coefficient linking T and Y (total causal effect); c' is the coefficient for the effect of T on Y adjusting for M (direct effect); b is the effect of M on Y adjusting for explanatory variables; and a is the coefficient relating to the effect of T on M . e_1 , e_2 , and e_3 are residuals that are uncorrelated with the variables in the right-hand side of the equation and are independent of each other. Under this specific model, the causal mediation effect (CME) is represented by the product coefficient of ab . Of note, Equation 3 can be substituted into Equation 2 to eliminate the term M :

$$Y = i_2 + bi_3 + (c' + ab)T + e_2 + be_3 \quad [4]$$

It appears that the parameters related to direct (c') and indirect effects (ab) of T on Y are different from those of their total effect. That is, testing the null hypothesis $c = 0$ is unnecessary since CME can be nonzero even when the total causal effect is zero (i.e., direct and indirect effects can be opposite), which reflects the effect cancellation from different pathways.

This standard setting for mediation analysis was refined and brought into the potential outcomes framework in Imai et al. (2011). The authors propose a set of methods that unifies the approach to identifying direct and indirect effects, relying on a set of assumptions that

are more readily testable than classical mediation analysis provides.

MEDIATION AND DELIBERATIVE POLLING

We believe we have identified four effects that may represent a mediated effect of our treatment on voting (both whether to vote and whom to vote for). The four mediators are change in PBS immediately following the deliberative weekend, a self-reported measure of following the campaign, a self-reported measure of whether one's political opinions are worth listening to, and general political knowledge. We view these four collectively as latent indicators of an underlying civic awakening that made participants more politically and civically engaged. We believe that the effect of the treatment on these mediators caused eventual changes in two key dependent variables: a respondent's propensity to vote at all and a respondent's propensity to vote for Biden.

For each mediator, we estimated as follows:

1st stage:

$$\text{Follows Campaign}_{T3} \sim c^* \text{Treatment} + \text{PBS}_{T1}$$

$$+ \text{Demographic Controls}_{T1}$$

$$\text{Worth Listening To?}_{T3} \sim c^* \text{Treatment} + \text{PBS}_{T1}$$

$$+ \text{Demographic Controls}_{T1}$$

$$\text{KnowledgeIndex}_{T3} \sim c^* \text{Treatment} + \text{PBS}_{T1}$$

$$+ \text{Demographic Controls}_{T1}$$

$$\text{PBS}_{T1} - \text{PBS}_{T2} \sim c^* \text{Treatment} + \text{PBS}_{T1}$$

$$+ \text{Demographic Controls}_{T1}$$

2nd stage:

$$\text{Outcome}_{1,2} \sim \text{Follows Campaign}_{T3} + c^* \text{Treatment}$$

$$+ \text{PBS}_{T1} + \text{Demographic Controls}_{T1}$$

$$\text{Outcome}_{1,2} \sim \text{Worth Listening To?}_{T3} + c^* \text{Treatment}$$

$$+ \text{PBS}_{T1} + \text{Demographic Controls}_{T1}$$

$$\text{Outcome}_{1,2} \sim \text{KnowledgeIndex}_{T3} + c^* \text{Treatment}$$

$$+ \text{PBS}_{T1} + \text{Demographic Controls}_{T1}$$

$$\text{Outcome}_{1,2} \sim (\text{PBS}_{T1} - \text{PBS}_{T2}) + c^* \text{Treatment}$$

$$+ \text{PBS}_{T1} + \text{Demographic Controls}_{T1}$$

where $\text{Outcome}_{1,2}$ refers to vote at all ($T4$) and vote for Biden ($T4$), respectively. To estimate the mediation effects, we utilized a mixed effects regression framework, with demographic controls and random intercepts at the state level.¹⁰ Demographic controls include education, gender, age, race, marital status, employment status, income level, home ownership status, metro/rural area of residence, and party ID. For consistency, we use the same sets of controls and

¹⁰ State level random intercepts allow for heterogeneities in voting propensities by state level characteristics, similar to a regression strategy discussed in Gelman and Hill (2006).

TABLE 4. Average Causal Mediated Effect (ACME) of Participation in A1R (95% CI) on Vote Intention

		Dependent variable:	
		Vote at all (t4)	Vote for Biden (t4)
Mediator	Follows campaign	0.0249*** (0.01, 0.04)	0.0012 (-0.002, 0.01)
	Worth listening to?	0.0115*** (0.01, 0.02)	0.0037* (-0.001, 0.01)
	Knowledge index	0.00859*** (0.003, 0.020)	-0.0006 (-0.003, 0.002)
	PBS _{T1} - PBS _{T2}	-0.0012 (-0.009, 0.01)	0.0184*** (0.01, 0.03)
Observations		1145	1000

Note: Each model is fit using a generalized linear mixed effects model for both the mediators and the dependent variables—linear models for each of the mediators and logistic regression models for the dependent variables. The dependent variables are Vote at All and Vote for Biden, as indicated at the top of the table. Random intercepts were fit at the level of the respondents' home state. We include a set of respondent demographic controls (age, gender, race, education poverty, and party ID) in each model, as well as respondent PBS. Observations include participants and control groups members. Models are fit using the "mediation" package in R with 95% CI included in the parenthesis. Complete model results are available in the Supplementary Tables A14-1, A14-2, A14-3, and A14-4. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

regression modeling specifications for each of the models.

Participation in the deliberations significantly increased campaign interest, self-efficacy, general political knowledge, and movement to the left overall for the PBS between T1 and T2, as demonstrated earlier. Because of this, we know that it is possible that these four mediators will have significant indirect effects on outcomes, even if there is weaker evidence for a direct effect of these mediators. For the entire sample, these effects are shown in Figures 3 through 5. "Following the campaign," having opinions "worth listening to," and general political knowledge are significant mediators for whether or not one will vote. Table 4 shows the causal mediation effects over the full range of the PBS.

Table 4 presents the results of this mediation analysis, with the mediators on the left-hand side of the table, and the dependent variables on top. The effect listed is the average causal mediated effect with 95% confidence intervals presented. The effects in this analysis are estimated using the "mediation" package in R (see Imai et al. 2011 for a discussion and Tingley et al. 2014 for an overview of the features of the package). 95% credible intervals are estimated with a parametric bootstrap with 1000 intervals, estimated with robust standard errors.

Indirect effects of the treatment were significant for voting at all if mediated through increases in respondent attention to the campaign, self-efficacy, and general political knowledge, though the relationship is not significant between these mediators and voting for Biden. However, there were significant indirect effects in intention to vote for Biden if mediated through changes in their PBS before and after treatment (movement to the left). For the group as a whole, there were significant indirect effects on voting for Biden from the treatment if the treatment induced a movement to the left along the PBS. But as Figures 8–10 suggest, these effects are likely to be more strongly felt among those in the middle range of the PBS (defined as respondents with scores 3–5 in their T1 PBS). As those participants start off at particularly low levels of civic engagement and as they lean slightly left in their policy positions, we believe they are likely to decide whether to vote and whom to vote for on the margin. Table 5 looks at the same set of models as Table 4 but restricts the analyses to just this middle group.

Here, we see a much larger effect from changes in the weekend on voting for Biden. However, with the smaller N of the middle range only, the other effects, except for following the campaign, are no longer significant. Before, for the full range of the PBS, the ACME was a rounded 0.02—roughly a 2% increase in the probability of voting for Biden. Now the effect is a rounded 0.06—roughly a 6% increase in the probability of voting for Biden. This suggests that the indirect effect of participating in the deliberations, mediated through short run changes to respondent PBS and thus an openness to moving one's average position left on policy issues, was responsible for a 6% increase in the likelihood a respondent would vote for Biden—even after conditioning on a variety of demographic controls as well as state fixed effects.

MEDIATION AND VOTE RECOLLECTION

In the interest of assessing whether the long run effect of deliberation on voting is not simply a function of intention, we also ran our same models from the previous section on a slightly different version of respondent voting behavior: voting recollection. Unlike in the previous section, where respondents were asked if they intended to vote and how they planned to vote in the upcoming election (T4), we now rely on retrospective descriptions of respondent voting behavior (T5) for our dependent variables.

We perform this analysis using the mediation analysis framework from the previous section, simply substituting out the dependent variables. We are still interested in the indirect effect of deliberation on following the campaign, self-efficacy, political knowledge, and changes in ideology pre- and posttreatment. We use reported voting at all as our first dependent variable and reported voting for Biden as our second dependent variable.

In Table 6, we see two important trends. First, the mediated effects of deliberation through knowledge

TABLE 5. Average Causal Mediated Effect (ACME) of Participation on Vote Intention Middle Group Only

		Dependent variable	
		Vote at all (t4)	Vote for Biden (t4)
Mediator	Follows campaign	0.0379*** (0.012, 0.06)	0.0045 (-0.011, 0.02)
	Worth listening to?	0.0083 (-0.004, 0.03)	0.0061 (-0.009, 0.02)
	Knowledge index	0.0063 (-0.004, 0.02)	-0.0022 (-0.013, 0.01)
	PBS _{T1} - PBS _{T2}	0.0043 (-0.0130, 0.02)	0.0487*** (0.0234, 0.10)
Observations		338	339

Note: Each model is fit using a generalized linear mixed effects model for both the mediators and the dependent variables—linear models for each of the mediators and logistic regression models for the dependent variables. The dependent variables are Vote at All and Vote for Biden, as indicated at the top of the table. Random intercepts were fit at the level of the respondents' home state. We include a set of respondent demographic controls (age, gender, race, education, poverty, and party ID) in each model, as well as respondent PBS. Observations include participants and control groups members. Models are fit using the "mediation" package in R with 95% CI included in the parenthesis. Complete model results are available in the Supplementary Tables A15-1, A15-2, A15-3, and A15-4. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

and change in ideology are basically the same between voting intention and voting recollection; deliberation increases knowledge, which increases an individual's propensity to vote. Furthermore, the shift in the PBS from deliberation between T1 and T2 had a bigger effect on an individual's propensity to report actually voting for Biden. The primary change is with the first two mediators: following the campaign and respondent self-efficacy. The effect that deliberation has on respondents' self-reported following of the campaign still has an effect on their propensity to vote. What has changed is that this same effect also makes participants more likely to self-report voting for Biden. The opposite is true for respondent self-efficacy; there is now no effect from deliberation to self-efficacy to any change in self-reported behavior. This may represent an interesting illustration of the potential difference in how prospective versus retrospective assessments of voting behavior tracks respondent self-assessment.

Table 7 shows the results of causal mediation analysis of the middle group, but using measures of reported

voting after the November 2020 election. Similar to the relationship between Tables 4 and 6, Table 7 tells largely the same story as Table 5, with the sole exception being the emergence of an indirect effect of deliberation through following the campaign on voting for Biden. The results are otherwise largely unchanged.

There is a long-standing discussion about overreporting of voting in the literature (Belli, Traugott, and Beckman 2001; Bernstein, Chadha, and Montjoy 2001). But the fact that our results are essentially unchanged, whether we measure voting outcomes before or after the election suggests a robustness of the effects of deliberation on this array of mediators. Thus far, the causal mediation analyses have employed self-reported voting.

But we also collected verified votes for the participant and control samples after the election. Of course, there are well-known challenges with voter verification (Katosh and Traugott 1981; Miller et al. 2021). Some voters exaggerate whether they have voted, some move, some have different spellings of their names, or change their names. Despite these issues, we collected verified voter information for our sample and then redid the causal mediation analysis for those for whom we could definitely verify that they voted. The results are presented in Supplementary Table A12 for the whole range of the PBS and in Supplementary Table A13 for the middle range only. They show that the causal mediation results for voting at all and for voting for Biden remain essentially unchanged. They are not the result of people overreporting that they had voted because the same causal relations hold for those for whom we could definitely verify whether or not they voted.

Thus far, we have traced elements of a civic awakening—greater efficacy, increased knowledge, and closer attention to the campaign among the deliberators. We have also seen indirect effects of the civic awakening on voting at all and voting for Biden. However, we can also explore whether there is a direct effect of their time 3 PBS scores on how they voted. Once awakened, are the deliberators more likely to take their policy preferences into account in deciding whom to vote for?

We can see this relationship in Table 8. In this set of regressions, we seek to compare how policy positions (PBS score) measured at different times predicts voter behavior in the 2020 election. We focus on three separate models which are identical in all indications, except they use a respondents' PBS measured at three separate times. If, as we believe, participants become better spatial voters—voters who are better able to transform policy preferences into voting behavior—we should see a strong negative correlation between respondent PBS and voting for Biden (negative since positive scores mean more conservative), and we should see a significant coefficient on the interaction between participant status and the PBS score. We compare ideology pretreatment (time 1), immediately posttreatment (time 2), and in the year follow-up (time 3). We find results that confirm our expectations: respondents are more likely to vote for Biden than

TABLE 6. Average Causal Mediated Effect (ACME) of Participation (95% CI) on Recollected Vote

		Dependent variable	
		Did vote at all	Did Vote for Biden
		(t5)	(t5)
Mediator	Follows campaign (t3)	0.0108*** (0.006, 0.02)	0.0089*** (0.0017, 0.02)
	Worth listening to? (t3)	0.003 (0.001, 0.01)	0.0046 (-0.0015, 0.01)
	Knowledge index (t3)	0.0055*** (0.002, 0.01)	0.0001 (-0.00004, 0.0001)
	PBS _{T1} – PBS _{T2}	-0.0026 (-0.01, 0.00)	0.0256*** (0.015, 0.04)
	Observations	1047	1113

Note: Each model is fit using a generalized linear mixed effects model for both the mediators and the dependent variables—linear models for each of the mediators and logistic regression models for the dependent variables. The dependent variables are vote at all and vote for Biden, as indicated at the top of the table. Random intercepts were fit at the level of the respondents' home state. We include a set of respondent demographic controls (age, gender, race, education, poverty, and party ID) in each model, as well as respondent PBS. Observations include participants and control groups members. Models are fit using the “mediation” package in R with 95% CI included in the parenthesis. Complete model results are available in the Supplementary Tables A16–1, A16–2, A16–3, and A16–4. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 7. Average Causal Mediated Effect (ACME) of Participation on Recollected Vote: Middle Group Only

		Dependent variable	
		Did vote at all	Did vote for Biden
		(t5)	(t5)
Mediator	Follows campaign (t3)	0.012* (-0.00002, 0.008)	0.0179* (-0.0004, 0.05)
	Worth listening to? (t3)	0.001 (-0.009, 0.01)	0.0115 (-0.006, 0.03)
	Knowledge index (t3)	0.0031 (-0.0037, 0.01)	-0.0042 (-0.018, 0.001)
	PBS _{T1} - PBS _{T2}	0.0058 (-0.0093, 0.02)	0.0507*** (0.0221, 0.09)
	Observations	303	324

Note: Each model is fit using a generalized linear mixed effects model for both the mediators and the dependent variables—linear models for each of the mediators and logistic regression models for the dependent variables. The dependent variables are vote at all and vote for Biden, as indicated at the top of the table. Random intercepts were fit at the state level. We include a set of respondent demographic controls in (age, gender, race, education poverty, and party ID) in each model, as well as respondent PBS. Observations include participants and control groups members. Models are fit using the “mediation” package in R with 95% CI included in the parenthesis. Complete model results are available in the Supplementary Tables A17–1, A17–2, A17–3, and A17–4. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

members of the control group, having a higher PBS makes one less likely to vote for Biden, and, most importantly, being a participant makes this relationship statistically *stronger*.

The relationship between voter policy positions at each time and whether or not they ultimately vote for Biden gets stronger for participants given their time 3 PBS, rather than their time 1 or time 2 PBS; the

TABLE 8. Voting for Biden by Participant Status and Policy Score over Time

	Model 1	Model 2	Model 3
Participant	4.652** (1.671)	2.761* (1.354)	5.420** (2.073)
PBS _{T1}	-1.551*** (0.139)		
Participant X PBS _{T1}	-0.921* (0.361)		
PBS _{T2}		-1.623*** (0.149)	
Participant X PBS _{T2}		-0.781* (0.319)	
PBS _{T3}			-1.757*** (0.170)
Participant X PBS _{T3}			-1.075* (0.447)
Num.Obs.	966	965	944
Adj. <i>R</i> ²	0.886	0.858	0.907
AIC	427.3	455.9	353.5
BIC	476.1	504.6	402.1

Note: Dependent variable is a binary on “voting for Biden,” conditional on having voted. The model is a logit regression with random intercepts for state. Each model includes demographic controls and party ID (which are the same as in the mediation results). The model uses PBS scores for each respondent measured at different times. Full regression table is available in the Supplementary Table A18. **p* < 0.05, ***p* < 0.01, ****p* < 0.001.

coefficient on the interaction goes from -0.921 for time 1 PBS interacted with treatment status, to -0.781 in time 2, and to -1.075 in time 3. This suggests that participant policy position is becoming a better predictor of voting for Biden over time; participants are voting more in line with their spatial preferences as their spatial preferences shift over time.

We can speculate how these results might apply to the broader universe of eligible rather than just registered voters. Might similar effects have been found among non-registered but eligible voters? The non-registered voters are likely to be less knowledgeable and less educated. Do we think deliberation would have comparable effects on them? Figure 10 shows the biggest effects of the treatment on the less knowledgeable in the middle of the policy space (PBS). Figure 5 shows that the biggest effects on voting intention came from those who lacked a college degree. So it is worth speculating that if registration as a barrier to voter participation were somehow to disappear, deliberation could be expected to have comparable or even greater effects among those currently non-registered.

CAN THESE DELIBERATIVE EFFECTS BE SCALED?

The picture that emerges from these analyses is that deliberation in an organized setting, on the model of

Deliberative Polling (Fishkin 1991; 2018), fostered elements of a civic awakening, particularly in the moderate and less politically engaged middle of the policy space. Those who were most affected by the deliberations during the weekend (as indicated by the changes in their PBS), those who subsequently followed the campaign more closely, those who thought they had opinions “worth listening to,” and those who gained knowledge over the course of the campaign were also more likely to vote and most particularly, more likely to vote for Biden in the 2020 Presidential election. In short, by intensively deliberating on the issues, becoming more aware of the campaign, having greater self-efficacy, and becoming more knowledgeable, they brought to life many of the elements of the “folk theory of democracy.” This is not a myth beyond the competence of ordinary citizens. It is a set of capacities that can be stimulated by institutional design. We think it is remarkable that such a short intervention can have a lasting effect a year later via the mediating variables in this civic awakening that led them to process the campaign and their voting decisions differently than the control group.

Think of the changed distribution of this political engagement. Before deliberation, our civic mediators tended to be distributed in the policy space in a kind of sunken parabola (a U shape) bottoming in the middle range (see Figures 3–5). Those in the broad middle range were left out—less likely to “follow the campaign,” less likely to think they had “opinions worth listening to,” and less likely to gain general political knowledge. But deliberation brought up the middle ranges and created a distribution on these variables more like a plateau, putting everyone on a more equal footing. This is a more inclusive form of democracy, where so many are not simply left out and where deliberation is an intrinsic part of participation.

This is very much like the vision in “Deliberation Day,” the idea of a national holiday in which the whole country deliberates on the issues in many organized small groups and comes to a considered judgement during the Presidential campaign (Ackerman and Fishkin 2004). In anticipation of such informed voters being energized *en masse*, the book argues that there would be rational incentives for candidates to adjust their campaign strategies to appeal to voters in more thoughtful and nuanced ways. Perhaps this would disincentivize at least some of our more manipulative campaign practices. Whether or not this latter claim is correct, it is surely true that a scaling of the deliberative process would take voters out of their filter bubbles and engage them with diverse others as they determine their views on the issues. Activating the broad middle of the policy spectrum would change the incentives for candidates (and their allies via independent expenditure groups) to do more than simply address the base of their parties to stimulate turnout. The overall electorate might depolarize because the universe of more moderate and potentially persuadable voters would be enlarged by bringing those in the broad middle of the policy space back into the political arena.

If we are correct in this picture, can deliberation actually be scaled? We believe this is an area ripe for

creative experimentation. One approach is through the Stanford Online Deliberation Platform, which reproduces the experience of Deliberative Polling for innumerable small group discussions. In fact, it has already been successfully employed as the mode of deliberation with stratified random samples in Japan, Hong Kong, Chile, Canada, and the United States, with up to 1,000 deliberators in 104 small groups (plus a separate control group).¹¹ In theory, the automated platform can handle any number of deliberative participants randomly assigned (with stratification) to small groups of ten or twelve. Further projects are planned to continue to expand scaling to much larger numbers, and study effects on participants. If eventual aspirations for mass participation in such processes succeed, this work suggests that we can achieve a more deliberative society.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <https://doi.org/10.1017/S0003055423001363>.

DATA AVAILABILITY STATEMENT

Research documentation and data that support the findings of this study are openly available at the American Political Science Review Dataverse: <https://doi.org/10.7910/DVN/ERXBAB>.

ACKNOWLEDGEMENTS

We would especially like to thank Henry Elkus, Sam Feinburg, and Jeff Brooks of Helena for their vision and collaboration. We also thank Larry Diamond for his invaluable contributions throughout the A1R project. We also want to thank Michael Dennis, Jennifer Carter, and their superb team at NORC at the University of Chicago. In addition, we want to thank Siddharth George for his valuable suggestions. This paper benefited from a presentation at the meetings of the American Political Science Association, September 2021. We would like to thank Helene Landemore, Jonathan Collins, Kimmo Grönlund, and John Gastil for their insights.

FUNDING STATEMENT

This research was funded by the Helena Group Foundation.

CONFLICT OF INTEREST

The authors declare no ethical issues or conflicts of interest in this research

¹¹ Subject of a separate paper, Fishkin et al. (2023).

ETHICAL STANDARDS

The authors declare the human subjects research in this article was reviewed and approved by NORC IRB: #21-07-386 and by the Stanford University IRB 35343. The authors also affirm that this article adheres to the APSA's Principles and Guidance on Human Subject Research.

REFERENCES

- Achen, Christopher H., and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press.
- Ackerman, Bruce, and James S. Fishkin. 2004. *Deliberation Day*. New Haven, CT: Yale University Press.
- Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102 (2): 215–32.
- Auerbach, Kiran, Joshua Lerner, and Hannah Ridge. 2022. "Measuring State Capacity in the U.S. States." Paper presented at the Annual Meeting of the American Political Science Association, Montreal.
- Belli, Robert F. Michael W. Traugott, and Matthew N. Beckman. 2001. "What Leads to Voting Overreports? Contrasts of Overreporters to Validated Voters and Admitted Nonvoters in the American National Election Studies." *Journal of Official Statistics* 17 (4): 479–98.
- Bernstein, Robert, Anita Chadha, and Robert Montjoy. 2001. "Overreporting Voting: Why It Happens and Why It Matters." *Public Opinion Quarterly* 65 (1): 22–44.
- Campbell, Angus, Philip Converse, Warren Miller, and Donald Stokes. 1960. *The American Voter*. New York: John Wiley and Sons.
- Chitra, Uthsav, and Christopher Musco. 2020. "Analyzing the Impact of Filter Bubbles on Social Network Polarization." In *WSDM '20: Proceedings of the 13th International Conference on Web Search and Data Mining*, 115–23. <https://doi.org/10.1145/3336191.3371825>.
- Coltman, Tim, Timothy M. Devinney, David F. Midgley, and Sunil Venaik. 2008. "Formative Versus Reflective Measurement Models: Two Applications of Formative Measurement." *Journal of Business Research* 61 (12): 1250–62.
- de Tocqueville, Alexis. [1838] 2019. *Democracy in America*, trans. Henry Reeve. Clark, NJ: The Lawbook Exchange.
- Dilko, Ivan, Igor Dolgov, William Hoffman, Nicholas Eckhart, and Maria Molina. 2017. "The Dark Side of Technology: An Experimental Investigation of the Influence of Customizability Technology on Online Political Selective Exposure." *Computers in Human Behavior* 73: 181–90.
- Fishkin, James S. 1991. *Democracy and Deliberation*. New Haven, CT: Yale University Press.
- Fishkin, James S. 2018. *Democracy When the People Are Thinking: Revitalizing Our Politics through Public Deliberation*. Oxford: Oxford University Press.
- Fishkin, James, Valentin Bolotnyy, Joshua Lerner, Alice Siu, and Norman Bradburn. 2023. "Scaling Dialogue for Democracy: Can It Create More Deliberative Voters?" Working Paper.
- Fishkin, James, Alice Siu, Larry Diamond, and Norman Bradburn. 2021. "Is Deliberation an Antidote to Extreme Partisan Polarization: Reflections on America in One Room." *American Political Science Review* 115 (4): 1464–81.
- Fishkin, James, Valentin Bolotnyy, Joshua Lerner, Alice Siu, and Norman Bradburn. 2024. "Replication Data for: Can Deliberation Have Lasting Effects?" Harvard Dataverse. Dataset. <https://doi.org/10.7910/DVN/ERXBAB>.
- Gastil, John, E. Pierre Deess, and Philip J. Weiser. 2002. "Civic Awakening in the Jury Room: A Test of the Connection between Jury Deliberation and Political Participation." *Journal of Politics* 64 (2): 585–95.
- 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: Oxford University Press.
- Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Green, Donald, and Alan Gerber. 2019. *Get Out the Vote*, 4th edition. Washington, DC: Brookings Institution Press.
- Green, Donald, Bradley Palmquist, and Eric Schickler. 2004. *Partisan Hearts and Minds*. New Haven, CT: Yale University Press.
- Grönlund, Kimmo, Andre Bächtiger, and Maija Setälä. 2014. *Deliberative Mini-Publics: Involving Citizens in the Democratic Process*. Colchester, UK: ECPR Press.
- Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. "A General Approach to Causal Mediation Analysis." *Psychological Methods* 15 (4): 309.
- Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105 (4): 765–89.
- Jacobs, Lawrence R., Fay Lomax Cook, and Michael X. Delli Carpini. 2009. *Talking Together: Public Deliberation and Political Participation in America*. Chicago, IL: University of Chicago Press.
- Karpowitz, Christopher, and Chad Raphael. 2014. *Deliberation, Democracy, and Civic Forums: Improving Equality and Publicity*. Cambridge: Cambridge University Press.
- Katosh, John P., and Michael W. Traugott. 1981. "The Consequences of Validation and Self-Reported Voting Measures." *Public Opinion Quarterly* 45 (4): 519–35.
- Mansbridge, Jane. 1999. "On the Idea That Participation Makes Better Citizens." In *Citizen Competence and Democratic Institutions*, eds. Stephen L. Elkin and Karol Edward Soltan, 291–328. University Park: Penn State University Press.
- Mill, John Stuart. [1861] 1991. *Considerations on Representative Government*. Amherst, NY: Prometheus Books.
- Miller, Jon D., Jason Kalmbach, Logan T. Woods, and Claire Cepuran. 2021. "The Accuracy and Value of Voter Validation in National Surveys: Insights from Longitudinal and Cross-Sectional Studies." *Political Research Quarterly* 74 (2): 332–47.
- Morrell, Michael E. 2005. "Deliberation, Democratic Decision-Making and Internal Political Efficacy." *Political Behavior* 27 (1): 46–69.
- Pariser, Eli. 2011. *The Filter Bubble*. New York: Penguin.
- Pateman, Carole. 1970. *Participation and Democratic Theory*. Cambridge: Cambridge University Press.
- Poole, Keith T. 1998. "Recovering a Basic Space from a Set of Issue Scales." *American Journal of Political Science* 42 (3): 954–93.
- Posner, Richard. 2005. *Law, Pragmatism and Democracy*. Cambridge, MA: Harvard University Press.
- Reuters Staff. 2021. "Fact Check: '133 Million Registered Voters' Argument Uses Flawed Logic." *Reuters*, January 1. <https://www.reuters.com/article/uk-factcheck-voters-133-million/fact-check-133-million-registered-voters-argument-uses-flawed-logic-idUSKBN296284>.
- Schumpeter, Joseph A. 1942. *Capitalism, Socialism and Democracy*. New York: Harper and Row.
- Shapiro, Ian. 2003. *The Moral Foundations of Politics*. New Haven, CT: Yale University Press.
- Sokolov, Boris. 2018. "The Index of Emancipative Values: Measurement Model Misspecifications." *American Political Science Review* 112 (2): 395–408.
- Sood, Gaurav, and Shanto Iyengar. 2016. "Coming to Dislike Your Opponents: The Polarizing Impact of Political Campaigns." Working Paper. doi:10.2139/ssrn.2840225.
- Spohr, Dominic. 2017. "Fake News and Ideological Polarization: Filter Bubbles and Selective Exposure on Social Media." *Business Information Review* 34 (3): 150–60.
- Stenner, Alfred, and Donald Burdick, and M. H. Stone. 2008. "Formative and Reflective Models: Can a Rasch Analysis Tell the Difference?" *Rasch Measurement Transactions* 22: 1152–53.
- Stepner, Michael. 2014. "Binscatter: Binned Scatterplots in Stata." PowerPoint Presentation. <https://michaelstepner.com/binscatter/binscatter-StataConference2014.pdf>. Accessed February 4, 2022.
- Sunstein, Cass R. 2017. *Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press.
- Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. 2014. "Mediation: R Package for Causal Mediation Analysis." *Journal of Statistical Software* 59 (5): 1–38.
- Trochim, William M. K. 2001. *Research Methods Knowledge Base*, 2nd edition. Cincinnati, OH: Atomic Dog Publishing.
- Zuiderveen Borgesius, Frederik J., Damian Trilling, Judith Möller, Balázs Bodó, Claes H. de Vreese, and Natali Helberger. 2016. "Should We Worry about Filter Bubbles?" *Internet Policy Review* 5 (1). doi:10.14763/2016.1.401.