

Deliberative Distortions? Homogenization, Polarization, and Domination  
in Small Group Deliberations\*

Robert C. Luskin, University of Texas at Austin

Gaurav Sood, ybar inc.

James S. Fishkin, Stanford University

Kyu S. Hahn, Seoul National University

July, 2017

## ABSTRACT

Does deliberation distort its participants' policy attitudes? It is widely believed to make them more "authentic," aligning them more closely with underlying values and interests. Yet there have also been claims that deliberation routinely and strongly "homogenizes" attitudes (decreasing their variance), "polarizes" them (moving their mean further out from the midpoint), or leads to "domination" by the socially advantaged (moving the overall mean attitude toward theirs). If so, the deliberation-induced attitude change may be stemming more from social dynamics than the merits of the arguments and may thus be making the participants' attitudes *less* authentic. This paper considers these claims both theoretically and empirically, noting the conditioning role of deliberative design. We explore the nature and normative valence of homogenization, polarization, and domination and examine their frequency and strength across 2,601 group-issue pairs in 21 Deliberative Polls. The results show no routine or pronounced homogenization, polarization, or domination.

Democrats value democracy. Realistic ones recognize its flaws and limitations, while still regarding it, in Churchillian vein, as a least-bad form of government. Many if not most also value deliberation, regarding it as something like a healthy diet for democracy. Correspondingly, the “deliberative turn in democratic theory” (Dryzek 2002) has produced a surge of theoretical and empirical work on “deliberative democracy” (e.g., Bohman and Rehg 1997, Elster 1998, Fishkin 1991, Smith 2009, Mendelberg 2002, and Gronlund et al. 2014). Among other salutary effects, deliberation is widely believed to help make citizens’ policy attitudes and electoral preferences more “authentic”—closer to those they *would* have, if they had unlimited information (facts, interpretations, perspectives) and opportunity for reflection.<sup>1</sup>

But does it? Critics have claimed, variously, that deliberating groups tend: (1) to converge on the same view (Sunstein 2000, 2002, 2009; Sunstein and Hastie 2014); (2) to become more extreme, moving further out on whichever side of the issue they started on (Sunstein 2000, 2002, 2009; Sunstein and Hastie 2014); or (3) to move toward the views of their socially advantaged (male, better educated, better-off) members (Sanders 1997, Young 2000, and, more conditionally, Karpowitz and Mendelberg 2014). We term these phenomena *homogenization, polarization, and domination*.<sup>2</sup> Each may occur, in greater or lesser degree, in any given group on any given issue in any given deliberation.

What matters for present concerns is the *pattern*. A good many groups may exhibit pronounced homogenization, polarization, or domination on a good many issues, in a good many deliberations. That much is unsurprising and untroubling. But *routine and pronounced* homogenization, polarization, or domination *would* be troubling, on account of what it would suggest—that deliberation is changing attitudes less on the basis of any rational weighing of the merits than on the basis certain social dynamics. The former should generally make the

participants' attitudes more authentic, the latter almost certainly make many of them less authentic. In this sense, routine and pronounced homogenization, polarization, or domination suggests a kind of "deliberative distortion." We expand on this argument below.

This paper examines all three phenomena, focusing on their frequency and magnitude. We consider their nature, normative valence, and dependence on deliberative design. For evidence, we corral 2,601 group-issue pairs in 21 Deliberative Polls, each containing 14-30 deliberative groups, discussing 4-12 policy issues. We examine how often and how much the deliberation actually homogenizes the groups' policy attitudes, polarizes them, and moves them toward those held by the advantaged. The results, at least for this deliberative design and others relevantly like it, are reassuring.

### **Deliberation and Democracy**

By *deliberation*, we mean elevated discussion<sup>3</sup>—discussion that is, at minimum, (1) substantive, (2) inclusive, (3) responsive, and (4) open-minded. That is, (1) the participants exchange relevant arguments and information. (2) The arguments and information are wide ranging in nature and policy implications—not all of one kind, not all on one side. (3) the participants react to each other's arguments and information. And (4) they seriously (re)consider, in light of the discussion, what their own policy views should be.<sup>4</sup>

A *discussion*, to delimit our topic further, is an extended, direct, verbal exchange. The exchange need not be face-to-face. It maybe telephonic or online. It may involve more than just words. Nonverbal and paralinguistic information may also play a part. But, to be a discussion, an exchange must be largely verbal. It must also be extended, more than just a few brief utterances. And it must be direct: the participants address and be addressed by one another. Thus the number of participants cannot be extremely large. A "discussion" involving hundreds

or millions can only figuratively be a discussion, much less a deliberation. It can contribute to “audience democracy” (Manin 1995, 218-235), but not deliberative democracy.

To be sure, a discussion’s deliberativeness is a matter of degree. The bottom of the range is vacuity: nobody says anything of substance. The top is an unattainable ideal in which every argument is made and countered, and those counter-arguments countered in turn, and so forth, and in which everyone weighs all those arguments, free of all coercion, reminiscent of Habermas’s (1990, pp. 85-87) “ideal speech situation.”<sup>5</sup> In-between, the completeness with which arguments are raised, countered, and weighed is a continuum (Fishkin 1991, pp. 36-38).

Real-world discussions inevitably fall much nearer the bottom than the top of this range. Not many people talk much about politics. They talk especially little especially with others unlike themselves, effectively limiting the range of arguments and information to which they are exposed. They shy away from topics on which those with whom they are talking may hold opposing views, engaging in “selective disclosure” (Cowan and Baldassari 2017). And they seldom give the substance of any discussion much thought. Thus studies of real-world political discussion (as in, for example, Huckfeldt and Sprague 1995, Beck et al. 2002, Mutz 2006, or Searing et al. 2007) speak only indirectly to deliberation and its effects—to what we might expect to see at substantially above-real-world-discussion levels.<sup>6</sup>

Among many other things, deliberation figures to shape the participants’ policy attitudes and preferences. We focus on attitudes (construed as continuous dispositions to favor or oppose given alternatives), but the argument is the same for preferences (discrete orderings of alternatives). The more deliberative the discussion—the larger the quantity and wider the range of the information and arguments considered and the more earnestly and open-mindedly they are considered—the more the participants should be moved by the merits of the issue as they see

them. This is something like Habermas's (1990, p. 86; 1998, pp. 305-306) "unforced force of the better argument"—subject to the non-Habermasian stipulation that what is the "better argument" may not be the same for everyone. It may be the same for everyone having exactly the same values and interests, but values and interests vary greatly.

This matters because such merits-driven thinking should make the participants' attitudes more *authentic*: closer to those they would hold in the light of unlimited (and, if truly unlimited, also perfectly balanced) information and reflection—presumably those maximally consistent with their fully considered values and interests.<sup>7</sup> That serves the individual participants. Consciously or not, most of us, most of the time, want to pursue our values and interests. We may often hold inauthentic attitudes, but usually because we have not learned or thought enough to see the conflict, only rarely because we have seen the conflict and shrugged it off. In the aggregate, more contingently, deliberation can make for a more authentic public opinion, moving the mean attitude closer to the mean authentic one.<sup>8</sup> With respect to preferences, including vote choices, it also (also contingently) can make for more frequently authentic majorities (in which the same option commands the majority of actual as of authentic preferences). That serves democracy.

### **Authenticity**

Since, as this begins to suggest, the idea of authenticity plays a key role in our argument, several points may be worth highlighting:

- A person's authentic attitude is not the same as his or her actual one. Authentic attitudes are hypothetical: the attitudes people *would* have if they had unlimited information and gave the issue unlimited consideration. Everyone, on any given issue, has an authentic attitude, even on issues they have never heard of and on which they therefore have no

actual attitude. To wit, imagine an attitude toward some policy option ranging from 0 (completely opposed) to 1 (completely supportive), with .5 representing neutrality, as in our definitions and measures below. Let the  $i$ th individual have an authentic attitude of  $A_i^*$  and an actual attitude (the inside-the-head disposition, as distinct from the questionnaire-response-based index used to measure it) of  $A_i$ , which could be any number within the  $[0, 1]$  interval.<sup>9</sup>  $A_i$  is fully authentic if it exactly =  $A_i^*$ , which rarely if ever occurs. More generally,  $A_i$ 's authenticity can be defined as  $1 - |A_i - A_i^*|$ , equaling 1 for fully authentic attitudes ( $A_i = A_i^*$ ) and 0 for fully inauthentic ones ( $A_i = 1$  and  $A_i^* = 0$ , or vice versa).

- Authenticity in this sense is promoted but not defined by the depth and inclusiveness of the information and thought underlying  $A_i$ . It is a matter of destination, not itinerary. If  $A_i^* = .71$ ,  $A_i = .69$  would be very authentic, even if it rests on little reasoning and scant or one-sided information. That will sometimes happen. A lightly-considered attitude is almost a guess about what one's attitude should be, in the vein of Converse's (1964) "non-attitudes." By chance, some such guesses will be approximately right. Not resting on much information or thought, they are apt to be unstable, but they are still, if often transiently, authentic.
- Many attitudes, on most issues, are quite *inauthentic*—especially absent or before deliberation, but even, to a lesser degree, after. No feasible deliberation ever affords unlimited or perfectly balanced information. Nobody ever gives anything unlimited reflection. Arguments are never completely countered (Fishkin 1991). But greater deliberation should tend to make policy attitudes more authentic, and more stably so.

- We can never *know* authentic attitudes. We can sometimes *estimate* them, subject to serious reservations (Luskin 2003).<sup>10</sup> But here we need not even do that. We need only *posit* them, as a conceptual touchstone. Attitude change in the direction of greater authenticity is desirable, in the opposite direction undesirable.

### **Homogenization, Polarization, and Domination**

Homogenization, polarization, and domination are species of attitude change plausibly resulting from small group discussion. Their plausibility rests on certain social dynamics: (1) Some participants may simply adopt the attitudes prevailing among their fellow group members as way of gaining approval or avoiding conflict. That should yield homogenization (Sunstein 2002; Sunstein and Hastie 2015). (2) The group members may hear more arguments on the side of the issue toward which the group initially leans. That should yield polarization (Sunstein 2002, 2009). (3) Everyday social inequalities may lead women, the less well educated, and the less well-off may speak less often, express themselves less articulately or less forcefully, or be less seriously listened-to (Fraser 1993, Sanders 1997, Young 2000). That should yield domination.

But these expectations are *ceteris paribus*. Small group discussion need not always produce homogenization, polarization, or domination. Much depends on the nature and context of the discussion—including how deliberative it is. Nor, for that matter, are such homogenization, polarization, and domination as do occur necessarily grounds for concern. The conditions under which small group discussion can be expected to produce these results and those under which they may be normatively troubling are important elements of the argument to follow.

We have informally sketched homogenization, polarization, and domination above. But



let us underline several points about these concepts and their place in our argument:

- Homogenization, polarization, and domination are statistical results, to be distinguished from the mechanisms producing them. The variance of a group's attitudes' may decrease, or its mean attitude move further out on the same side or toward the mean attitude of its socially advantaged members for various reasons—including but not necessarily confined to the social dynamics just cited. For instance, domination is presumably affected by but not the same as the quantity or quality of what the socially advantaged versus disadvantaged may say (examined, e.g., in Karpowitz et al. 2012). A group's men may say little, and of little import, but if the group's attitudes shift toward theirs that is still domination. Or they say much, and of great import, but if the group's attitudes do not shift toward theirs, that is not domination.<sup>11</sup>
- Homogenization, polarization, and domination are continuous variables, naturally centered at 0, not just discrete phenomena that may or may not obtain. The names convey only one-side of the continuum. Deliberation may not just fail to homogenize a group's attitudes but may *variegate* them (increasing rather than decreasing their variance). It may not just fail to polarize them but may *moderate* them (moving them back toward the center rather than out toward the nearer extreme). The group may not just fail to move toward the initial attitudes of the advantaged but may move away from them—may be *repelled* (in the sense of physics, not necessarily psychology) rather than dominated by them.
- The differences between homogenization and variegation, between polarization and moderation, and between domination and repulsion are matters of *sign*. The homogenization (or variegation), polarization (or moderation), or domination (or

repulsion) may also be great or small. This is a matter of *magnitude*. The binary version—whether a group homogenizes, polarizes, or exhibits domination or not—is a matter of *occurrence* (in the aggregate, of *frequency*).

For still greater clarity, we need formal definitions. Denote the time- $t$  attitude, on the  $j$ th of  $J$  issues, of the  $i$ th of  $n_g$  members of the  $g$ th of  $G$  groups by  $A_{igjt}$ , where  $A_{igjt}$  runs from 0 to 1, with .5 representing neutrality, as in our scorings, and  $t = 1, 2$  (pre- and post-deliberation). Further, denote  $A_{igjt}$ 's mean within the  $g$ th group (also running from 0 to 1) by  $\bar{A}_{gjt}$  and its standard deviation within the  $g$ th group by  $s_{gjt}$ . Finally, denote the mean time- $t$ ,  $j$ th-issue attitudes of the  $g$ th group's advantaged and disadvantaged members by  $\bar{A}_{gjt}^a$  and  $\bar{A}_{gjt}^d$ , given some mutually exclusive, exhaustive division of the group's members into advantaged and disadvantaged. In these terms, the homogenization, polarization, and domination (with respect to a given dimension of advantage) of the  $g$ th group's attitudes on the  $j$ th issue are, respectively:

$$H_{gj} = s_{gj1} - s_{gj2}.^{12}$$

$$P_{gj} = \bar{A}_{gj2} - \bar{A}_{gj1} \text{ for groups starting on the "right" } (\bar{A}_{gj1} > .5) \text{ and} \\ = \bar{A}_{gj1} - \bar{A}_{gj2} \text{ for groups starting on the "left" } (\bar{A}_{gj1} < .5),^{13}$$

and

$$D_{gj} = \bar{A}_{gj2} - \bar{A}_{gj1} \text{ for groups starting to left of their advantaged members} \\ (\bar{A}_{gj1}^a > \bar{A}_{gj1}) \text{ and} \\ = \bar{A}_{gj1} - \bar{A}_{gj2} \text{ for groups starting to right of them } (\bar{A}_{gj1}^a < \bar{A}_{gj1}).^{14}$$

$H_{gj}$ ,  $P_{gj}$ , and  $D_{gj}$  combine sign and magnitude. They are  $> 0$  for homogenization,

polarization, and domination,  $< 0$  for variegation, moderation, and repulsion. Dichotomizing—as  $H'_{gj} = 1$  if  $H_{gj} > 0$  and  $= 0$  otherwise,  $P'_{gj} = 1$  if  $P_{gj} > 0$  and  $= 0$  otherwise, and  $D'_{gj} = 1$  if  $D_{gj} > 0$  and  $= 0$  otherwise—expresses occurrence and, in the aggregate, frequency.

Figure 1 illustrates  $H_{gj}$ ,  $P_{gj}$ , and  $D_{gj}$  diagrammatically, denoting the pre- and post-deliberation group mean attitudes simply as  $\bar{A}_1$  and  $\bar{A}_2$  the pre-deliberation mean attitude of the group's advantaged members simply as  $\bar{A}_1^a$  (leaving the  $g$  and  $j$  subscripts understood), and indicating the within-group variation by more or less elongated ellipses. The diagrams are meant to underscore the following:

- $H_{gj}$  is about the within-group variance, regardless of what happens to the mean (compare Figure 1's A1a and A2a with its A1b and A2b). A decrease in the variance (as in A1a and A2a) is homogenization, an increase in the variance (as in A1b and A2b) variegation.
- $P_{gj}$  is about the group mean in relation to the midpoint. Its moving further out on the same side (as in B3a and B4a) is polarization, its moving toward or beyond the midpoint (as in B3b, B4b, and B5b) moderation. Thus while any group that polarizes is *ipso facto* becoming more extreme, not every group that becomes more extreme need be polarizing. A group that moves from .6 to .1 (as in B5b) is, in this sense, moderating, even though .1 is more extreme than .6.
- $D_{gj}$  is about the group mean in relation to the pre-deliberation mean of the group's more advantaged members. Its moving toward or beyond the latter (as in C6a, C7a, and C8a) is domination, its moving in the other direction (as in C6b and C7b) repulsion. The distinction concerns the direction of movement, not the change in proximity. Thus while any group that is repulsed is *ipso facto* increasing its distance from the advantaged, not

every group that is dominated need be decreasing it. If a group whose advantaged members are at .6 moves from .7 to .3 (as in C8a), it is dominated, even though its distance from the initial position of its advantaged members has increased, from .1 to .3.

$H_{gj}$  is at its most positive (.5) when the participants split evenly between the two polar attitudes (half at 0, half at 1) before deliberating but all have exactly the same attitude (wherever it may fall on the scale) after doing so—changing, i.e., from perfect dissensus to perfect consensus. It is at its most negative (−.5) for the opposite change, from perfect consensus to perfect dissensus.  $P_{gj}$  is undefined when the pre-deliberation mean is exactly .5 but is at its most positive (just barely under .5) when the mean is either just barely above .5 before deliberation and exactly 1 after or just barely below .5 before deliberation and exactly 0 after. It is at its most negative (−1) when all the participants and thus the mean are at one pole (1 or 0) before deliberating and at the opposite pole (0 or 1) after.<sup>15</sup>

The story for  $D_{gj}$  is slightly more complex, but suppose for simplicity that the advantaged and disadvantaged are equally numerous—as, operationally, the analyses below take them to be. Then  $D_{gj}$  is at its most positive (just barely under 1) when all the disadvantaged start at 1 or 0, all the advantaged start just barely toward the midpoint from that, and everyone ends at the opposite pole (0 or 1). In this case, all the disadvantaged move not only toward but far *beyond* the initial position of the advantaged—all the way to the opposite extreme—and the advantaged themselves move in the same direction, to the same opposite extreme. It is at its most negative (just barely above −1) when all the *advantaged* start at 1 or 0, all the *disadvantaged* start just barely toward the midpoint from that, and again everyone ends at the opposite pole (0 or 1). In this case, all the disadvantaged move as far as possible *away from* the initial position of the advantaged—all the way to the opposite extreme—and again the advantaged themselves move in the same direction,

to the same opposite extreme.<sup>16</sup>

Since our argument concerns aggregate patterns, we focus on the population means, over all possible group-issue pairs, of  $H_{gj}$ ,  $P_{gj}$ ,  $D_{gj}$ ,  $H'_{gj}$ ,  $P'_{gj}$ , and  $D'_{gj}$  (in conventional notation,  $E(H_{gj})$ ,  $E(P_{gj})$ ,  $E(D_{gj})$ ,  $E(H'_{gj})$ ,  $E(P'_{gj})$ , and  $E(D'_{gj})$ ), estimated by the sample means over the group-issue pairs in our data (in conventional notation,  $\bar{H}$ ,  $\bar{P}$ ,  $\bar{D}$ ,  $\bar{H}'$ ,  $\bar{P}'$ , and  $\bar{D}'$ ).

### Normative Valences

In any given instance (for any given group-issue pair), homogenization, polarization, or domination may be either desirable or undesirable, depending on whether the changes involved bring the distribution of actual attitudes closer to or further from the distribution of authentic ones.

Take homogenization. Some deliberative democrats prize consensus—the degenerate distribution of actual attitudes (all at the same point) resulting from complete homogenization—as a goal in its own right.<sup>17</sup> Others, more in tune with present concerns, see it as undesirable, most likely reflecting a triumph of conformism over independent thought. But, in any given instance, it could be either, depending on the distribution of authentic attitudes. If the variance of the group’s initial attitudes exceeds that of its authentic ones, we should want it to decrease, but, if it is smaller, we should want it to increase. What matters is not the attitudes’ variance *per se* but its match to the authentic attitudes’ variance, which may be higher.

Much the same is true of the other two variables. Polarization, in the sense of moving toward the nearer extreme, sounds bad, but need it be?<sup>18</sup> On many issues, the pre-deliberation distribution of opinion features a large mode at the midpoint, the safest guess (yielding the minimum mean squared error) for those essentially guessing their own “attitude.” It may simply be that deliberating (a) moves participants off the midpoint as they learn and think enough to

acquire real attitudes and (b) mostly moves them toward the side of the scale on which the bulk of those already possessed of real attitudes already sit. Surely, (a) is desirable, and (b) expectable, as a function of the discussion's "argument pool." The real test, normatively, is whether the group's mean attitude moves—in whichever direction—toward its mean authentic attitude. Sometimes that entails the mean attitude's moving toward the nearer extreme, sometimes back toward the middle.

Domination may sound still worse. But recall the distinction between it and its sources. It may sometimes occur simply because the group's attitudes are becoming more authentic. To see this, let the  $g$ th group's mean authentic attitude on the  $j$ th issue be  $\bar{A}_{gj}^*$  and consider two scenarios. First, let  $\bar{A}_{gj1} < \bar{A}_{gj1}^a < \bar{A}_{gj}^*$  or  $\bar{A}_{gj1} > \bar{A}_{gj1}^a > \bar{A}_{gj}^*$ . In this case,  $\bar{A}_{gj}^*$  is not only toward but beyond  $\bar{A}_{gj1}^a$ . Then mean authenticity increases ( $|\bar{A}_{gj2} - \bar{A}_{gj}^*| < |\bar{A}_{gj1} - \bar{A}_{gj}^*|$ ) when  $\bar{A}_{gj2}$  is toward but not too far beyond  $\bar{A}_{gj1}^*$ . That necessarily implies domination ( $\bar{A}_{gj2} > \bar{A}_{gj1}$  for  $\bar{A}_{gj1}^a > \bar{A}_{gj1}$  and  $\bar{A}_{gj1} > \bar{A}_{gj2}$  for  $\bar{A}_{gj1}^a < \bar{A}_{gj1}$ ), although, in this case, domination can also sometimes occur when authenticity decreases (when  $\bar{A}_{gj2}$  is so far beyond  $\bar{A}_{gj}^*$  that  $|\bar{A}_{gj2} - \bar{A}_{gj}^*| > |\bar{A}_{gj1} - \bar{A}_{gj}^*|$ ). Second, let  $\bar{A}_{gj1} < \bar{A}_{gj}^* < \bar{A}_{gj1}^a$  or  $\bar{A}_{gj1} > \bar{A}_{gj}^* > \bar{A}_{gj1}^a$  ( $\bar{A}_{gj}^*$  is toward but short of  $\bar{A}_{gj1}^a$ ). Then authenticity increases when  $\bar{A}_{gj2}$  is toward or beyond  $\bar{A}_{gj1}^*$ . In this case, increasing authenticity and domination are one and the same ( $\bar{A}_{gj2} > \bar{A}_{gj1}$  for  $\bar{A}_{gj1}^a (> \bar{A}_{gj}^*) > \bar{A}_{gj1}$  and  $\bar{A}_{gj1} > \bar{A}_{gj2}$  for  $\bar{A}_{gj1}^a (< \bar{A}_{gj}^*) < \bar{A}_{gj1}$ ).

### **Underlying Mechanisms, a Rough Diagnostic, and Normative Valence Redux**

At least the bulk of the attitude change from deliberation stems from some mix of two basic mechanisms, both already sketched. The first lies in the participants' weighing of the policies' merits, as they individually see them. The participants come to see more clearly how

given policies may affect their values and interests (which they may also come to see more clearly). They therefore tend to change their attitudes in the direction of greater authenticity—regardless of whether that makes for greater homogenization, polarization, or domination. The second lies in the social dynamics previously identified. They should routinely tend to produce homogenization, polarization, and domination—regardless of whether the attitude changes involved are toward or away from greater authenticity.

The all-one-or-all-the-other extremes tell the story. On the one hand, suppose the only mechanism were the weighing of the merits. Then deliberation would sometimes produce homogenization, sometimes variegation (depending on whether the group's initial attitudes have a larger or smaller variance than its authentic ones); sometimes produce polarization, sometimes moderation (depending on whether the group's initial mean attitude is less extreme than its mean authentic one); and would sometimes produce domination, sometimes repulsion (depending on whether the group's initial mean attitude lies on the same or opposite side of its mean authentic one as it does of the initial mean attitude of the advantaged). There being no reason to expect any of these parenthetical alternatives to obtain far more, or less, less than half the time, we should not expect to see homogenization, polarization, or domination much more, or less, than half the time.

On the other hand, suppose the only mechanism were the social dynamics cited above. Then the deliberation (in that case only a shell of a deliberation), would routinely produce homogenization, polarization, and domination. We should expect to see all three the vast majority of the time. But, in this case, the attitude change would, in all probability, be taking many of the participants away from their authentic attitudes. Not too far from half the time, the group's initial attitudes can be expected to have a smaller variance than their authentic ones. Not

too far from half the time, the group's initial mean attitude can be expected to be more extreme than its mean authentic one. And, not too far from half the time, its initial mean attitude can be expected to be on the opposite side of its mean authentic one as it is of the initial mean attitude of the advantaged. In these scenarios, a great many participants must be moving away from their authentic attitudes.

All this suggests a rough diagnostic. If the groups' policy attitudes often homogenize, polarize, or move toward those of the advantaged, but *also* often do the opposite, the driving considerations may be merits-based. The pattern is not enough to establish that conclusion but does not belie it. If, on the contrary, the group's policy attitudes very preponderantly homogenize, become more extreme, or move toward those of the advantaged, the driving considerations are almost certainly driven less by the merits than by the social dynamics conducting to homogenization, polarization, and domination.

As we have seen, it is neither desirable nor undesirable on its face for any individual group to homogenize, polarize, or exhibit domination on any individual issue. At the level of the individual group-issue pair, the normative valence depends on whether the attitude change is toward or away from greater authenticity, which is hard to know. But, as we now see, the aggregate pattern can be more revealing. Specifically, a pattern of routine and strong homogenization, polarization, or domination is a token of not-very-deliberative discussion, in which many of the attitude changes must be *distortions*, leading the participants away from the positions they would take strictly on the merits.<sup>19</sup>

### **Deliberative Design**

So far we have been writing of deliberation in the abstract. In the real world, it can only be approximated, by what we shall call *deliberative designs*, concretizing who deliberates, about



what, and in what fashion. Among the more familiar species are Consensus Conferences, Citizens' Juries, and Deliberative Polls (see Gastil and Levine 2005 for more on the first two and the ensuing section here for more on the third). The species differ, as do individual specimens within each species, in ways that may mute, accentuate, or even reverse the expected effects of deliberation in the abstract (as Karpowitz and Mendelberg 2014 have argued and documented for some such variations). The features having the greatest effects on homogenization, polarization, and domination are of two (not mutually exclusive) sorts. Anything strengthening the relevant social dynamics should make these statistical patterns stronger and more routine. Anything helping the participants to work their way toward their more authentic attitudes, on the other hand, should make them weaker and less routine. Non-exhaustively, the list should include:

*Topic.* Some questions are of *action* (what the government should do about ISIS or about health care), others of *fact* (whether hydraulic fracturing increases the probability of earthquakes or whether the defendant in a criminal case committed the crime). Here we examine only policy attitudes—responses to questions of action—from a design (Deliberative Polling) focused on them. The evidence for routine and pronounced homogenization and polarization, by contrast, stems mostly from responses to questions of fact—from designs focused on those.

By definition, questions of fact have a right answer. It may not be obvious. It may be disputed. People may not always reach it, even after deliberating, even in the aggregate. We have all occasionally shaken our heads at headline verdicts. Juries, like the rest of us, sometimes misjudge. We all have our biases, sometimes seeing things the way we prefer them to be (engaging in Lodge and Taber's (2013) "motivated reasoning"). But we all also experience some strain toward accuracy, toward seeing things as they are. Doing so is usually survival-positive. "Consistency" motivations coexist with "accuracy" motivations (Kruglanski 1990,

Kunda 1990, Chaiken et al., 1996). Aggregation, moreover, often increases accuracy by allowing opposing biases to cancel out (as famously proved, under certain assumptions, by Condorcet 1785; see also Galton 2007). Thus, even before or absent deliberation, the mean response to a question of fact should usually be fairly close to the reality (as in Vidmar and Hans 2007). After deliberation, it should usually be still closer. In addition, the variation around that increasingly right mean response should generally decrease.

To illustrate, consider a graduated question of fact whose possible answers form a continuum mappable onto the  $[0, 1]$  interval—say, the degree to which climate change stems from human activity, from all (1) to none (0), or the probability that a defendant is guilty, from certainly (1) to certainly not (0). Pre-deliberation, some people will have encountered enough information and given it enough thought to arrive at something close to the right answer. Others, not having done so, may distribute their responses more uniformly. Suppose the right answer is .7. Let 25% of the deliberating group start at .7, and 75% be spread evenly between 0 and 1. Suppose further that deliberation shifts these proportions to 50-50, as more of the group comes to see the reality. That changes the mean response from .55 to .6 (polarization), and the standard deviation of the responses from .2646 to .2380 (homogenization). The details—the location of the right answer and the proportions initially knowing it and moving to it—do not affect the lesson.<sup>20</sup>

Plausible elaborations of the model underlying this example also seem to leave its implications broadly intact: on questions of fact, simply gravitating toward the right answer generally homogenizes and polarizes the group's responses.<sup>21</sup> The participants need feel no urge to conform. The argument pool need not be imbalanced, nor need any imbalance have any effect. These variables may also play a part but need not. Mere movement toward the right

answer (a benign, even desirable mechanism) is enough.

Questions of action, by contrast—arguably excepting rare cases unlikely to be posed for public deliberation—have no similarly right answer.<sup>22</sup> Or, in a sense, everyone has his or her own right answer—his or her authentic attitude—a non-additive function of multiple facts and other empirical premises, on the one hand, and multiple interests and values or goals, on the other. Deliberation can be expected to move everyone toward his or her authentic attitude, but that may or may not homogenize or polarize the group’s responses. Social dynamics may still produce some, possibly a great deal of homogenization, polarization, and domination, but overall there should be homogenization, polarization, and domination than on questions of fact, which could be expected to exhibit them even if social dynamics made no difference.

*Goal/Product.* Some designs, like jury deliberations, culminate in a conscious, collective decision or verdict.<sup>23</sup> Others, while stopping short of a collective decision, may nonetheless steer their participants toward consensus. Yet others, like Deliberative Polls, only encourage the participants to refine their own individual views. Having to reach a conscious, collective decision or at least a consensus, as in a jury verdict, exposes the participants to social pressures to indulge an emerging majority, thus tending to homogenize their opinions. That in turn may foster polarization, to the extent that it is the participants initially on the other side of the midpoint or closer to it on the same side who disproportionately move toward the closer extreme. As a practical constraint, they have more room to move in that direction, and, psychologically, those “short of” the group mean may be more drawn toward those “long of” it. Another possibility is that collective decision-making increases extremity by increasing risk acceptance (Sunstein and Hastie 2015).

*Deliberators.* Deliberating groups vary in attitudinal and demographic heterogeneity.

The first should always matter, the second independently matter on topics touching relevant social identities. The more heterogeneous the group, the likelier its members are to encounter information and arguments capable of moving them toward their authentic attitudes. In any given deliberation, that may yield homogenization or variegation, either polarization or moderation, either domination or repulsion. Across deliberations, it should minimize the extent to which there is routine homogenization (or variegation), routine polarization (or moderation), or routine domination (or repulsion).

*Information provided.* Many deliberative designs give their participants a head-start—a primer on the issues to be discussed. Together, the quantity of information provided and its balance (as between sides of the argument) should have the same sort of effect as the deliberators' heterogeneity, helping the participants to find their way toward their authentic attitudes, thus thwarting routine homogenization, polarization, or domination.

*Moderation* (in the sense of moderator behavior, as distinct from the opposite of polarization). The discussions may or may not be moderated. If moderated, the moderators may intervene more or less frequently and volubly and do more or less to embolden the shy or rein in the loquacious, to encourage the airing of all sides of the issue, and to steer the group toward agreement. That should make increase homogenization and, perhaps, indirectly, polarization and domination.

*Expression of opinion.* Some designs involve or permit open voting, shows-of-hands, or other inarticulate expressions of opinion during the discussions. These convey the group's distribution of opinions without revealing anything of the reasons behind them. Seeing the distribution encourages conformism. Not hearing the reasons hinders the participants in finding their way toward their own more authentic attitudes. The upshot should be more routine

homogenization, polarization, and domination.

A companion article, in progress, attempts a fuller fledged explanation, advancing and estimating a model to account for the variation, across group-issue pairs, in homogenization, polarization, and domination. Here we seek only to convey a sense of why our results look the way they do—and quite differently from those feeding worries about deliberative distortions. These latter stem mainly from jury-like designs like those in Sunstein (2000, 2002). The deliberations mostly concern questions of fact. They aim for consensus. Their *raison d'être* is to reach a collective decision. The harvesting of opinions is often public, often inarticulate. As the foregoing should make clear, they can therefore be expected to maximize homogenization, polarization, and domination.

### **Deliberative Polling**

The design from which we take our data, Deliberative Polling (see Fishkin 1991; Luskin, Fishkin, and Jowell 2002.), is very different. A face-to-face DP draws and interviews a random sample, then invites them to attend a weekend discussing the issues at a common site. Before the weekend, they are sent carefully balanced briefing materials laying out arguments for and against policy alternatives. During the weekend, they alternate between discussions in randomly assigned small groups led by trained moderators and plenary sessions in which they put questions shaped by the small groups to panels of experts, policy makers, or politicians.<sup>24</sup> At the end, they answer the same questions as at the beginning. Frequently a separate random sample, answering the same questions at the end of the process (and sometimes also at the beginning), provides a control group.

An online variant is broadly similar. In the studies included here, the mode is voice, not merely text, so what is lost is “just” the physical presence and company of others, as well as non-

vocal communication. The discussions are spread over a period of weeks. The assignment to small groups is nonrandom, constrained by possible meeting times. But the participants still “meet” in small groups, still led by trained moderators, and still put questions to policy experts and policy makers. They still receive briefing materials. They are still randomly sampled offline, with “non-users” receiving free access and incentives, sometimes including free equipment, to get them online.<sup>25</sup>

In Deliberative Polling, the questions are of action, having no uncontested right answer. The participants make no conscious, collective decision. They only talk, listen, learn, and reflect and thus form, reinforce, qualify, or adjust their individual attitudes. The moderators are trained to intervene as little and as lightly as possible, while still encouraging equal participation and the airing of opposing arguments and counterarguments. Briefing materials provide balanced information, along with, often embedded in, arguments for and against given policy positions. The participants’ opinions are collected only by means of a confidential questionnaire at the end. These features should curtail the effects of social dynamics, while facilitating a reasons weighing of the merits.

### **Data**

Our data come from the 21 DPs summarized in Table 1. Five took place in Britain, eleven in the U.S., two in the E.U., and one each in China, Australia, and Bulgaria. Sixteen were face-to-face, five online. The topics included a referendum (in Australia), a general election (in the U.K.), and presidential nomination contest (in the U.S.), as well as policy issues ranging from foreign policy to health care. In all, the data encompass 372 small groups (containing, between them, 5,736 participants), 139 policy issues (counting each policy attitude index as tapping a somewhat different issue), and 2,601 group-issue pairs.<sup>26</sup> Appendix A describes the indices and

their ingredients in greater detail.

In the fullest accounting, we are sampling from populations of multiple sorts—from: (1) the population of time-indexed person-populations (e.g., Great Britain in 1994, which is in our sample, and Paraguay in 2011, which is not), (2) the population of the actual or potential policy issues within the sampled time-indexed person-populations (e.g., crime in Bulgaria in 2007, which is in our sample, and climate change in the U.S. in 2009, which is not), and (3) the population of individuals within them. The samples of individuals for given DPs are almost always random draws.<sup>27</sup> The samples of time-indexed person-populations or policy issues are not. Yet we hope that the relatively large number and wide range of person-populations and policy issues affords some hard-to-quantify but nonzero sense that the results are unlikely to be peculiar to just a few places, times, or issues. Although most of the DPs here were in Anglo-American democracies, the patterns of homogenization, polarization, and domination looks no different for the group-issue pairs in the DPs in other settings, as can be seen in Appendix B.

### **How Often? How Much? The Evidence from Deliberative Polling**

Table 2 shows the degree to which deliberation, as embodied in Deliberative Polling, produces homogenization, polarization, and domination (with respect to gender, education, and income), across our pooled Deliberative Polling data's 2,601 group-issue pairs. The top row reports both the means of  $H_{gj}$ ,  $P_{gj}$ , and  $D_{gj}$ , indicating how much, on average, the group-issue pairs exhibit homogenization (or variegation), polarization (or moderation), and domination (or repulsion), and the means of  $H'_{gj}$ ,  $P'_{gj}$ , and  $D'_{gj}$ , indicating the frequency with which they do so. The trailing rows add Huber-White estimates of the standard errors (White 1980),<sup>28</sup> and two-tailed  $p$ -values for the null hypotheses of 0 for the means and .5 for the frequencies.<sup>29</sup>

The null hypotheses for the means of  $H_{gj}$ ,  $P_{gj}$  and  $D_{gj}$  are  $E(H_{gj}) = 0$ ,  $E(P_{gj}) = 0$ , and  $E(D_{gj}) = 0$ —meaning that, on average, both the mean and the variance of the groups' attitudes on the issues stay the same (no homogenization, no polarization, and no domination). The claims of routine and pronounced homogenization, polarization, and domination instead suggest that  $E(H_{gj}) > 0$ ,  $E(P_{gj}) > 0$ , and  $E(D_{gj}) > 0$ —meaning that the variance of attitudes shrinks and that the mean attitude moves toward both the nearer extreme and the mean initial attitude of the more socially advantaged. It is also possible but unexpected that  $E(H_{gj}) < 0$ ,  $E(P_{gj}) < 0$ , or  $E(D_{gj}) < 0$ —that, on average, the group-issue pairs show variegation, moderation, or repulsion.<sup>30</sup>

For the frequencies  $H'_{gj}$ ,  $P'_{gj}$ , and  $D'_{gj}$ , the null hypotheses are instead  $E(H'_{gj}) = .5$ ,  $E(P'_{gj}) = .5$ , and  $E(D'_{gj}) = .5$ —meaning that, averaging across group-issue pairs, the attitudes show variegation exactly as often as homogenization, moderation exactly as often as polarization, and repulsion exactly as often as domination. The claims of routine homogenization, polarization, and domination instead suggest  $E(H'_{gj}) > .5$ ,  $E(P'_{gj}) > .5$ , and  $E(D'_{gj}) > .5$ . Or, again, the opposite sign is possible. More group-issue pairs could show variegation than homogenization, moderation than polarization, or repulsion than domination:  $E(H'_{gj}) < .5$ ,  $E(P'_{gj}) < .5$ , or  $E(D'_{gj}) < .5$ .

Although the magnitudes do not interpret themselves (see below), these results are notably weak and not always of the claimed, worrisome sign. With the exception of homogenization's, the frequencies are all very close to .5, and the means all very close to 0. Even homogenization is not much of an exception. Its frequency is only somewhat above .5, its mean only somewhat above 0. Four of the five frequencies, moreover, are *less than* .5 (indicating moderation or repulsion), and one of the five means less than 0 (indicating moderation).



The general pallor of the estimates sidelines the question of statistical significance. As the  $p$ -values indicate, half of Table 2's results are "statistically significant" ( $p < .05$ ). In these cases, we can be quite sure that, in the population from which we are sampling, the frequency is not .5 or the mean not 0. But how different are observed frequencies from .5 and the observed means from 0? Not very for homogenization, and hardly at all for anything else.

Figure 2 shows the variation across group-issue pairs. Not only do  $H_{gj}$ ,  $P_{gj}$ , and  $D_{gj}$  all average very small, but they are all packed quite tightly and symmetrically around those very small means. This is very far from routine homogenization, polarization, and domination. But let us have a closer look.

### **Homogenization**

Deliberation does appear to homogenize attitudes somewhat more often than it variegates them. The percentage of the 372 groups homogenizing on the 139 policy issues is 59.5%—distinctly (and "significantly") above the 50.0% expectable by chance, if deliberation neither homogenized nor variegated attitudes at all. But 59.5% is still entirely earth-bound—far closer to .5 than to 1, and far from "routine." On average, moreover, deliberation does not homogenize attitudes by much. The mean  $H_{gj}$  is only .013.

To get a sense of this latter number, imagine a group of 20 participants, with four group members having an initial attitude at .6 and two each at every other integer multiple of .1 from .2 to 1. This initial distribution has a mean of .6 and a middling standard deviation of .245, close to halfway between the maximum of .5, when half the group is at 0 and the other half at 1, and the minimum of 0, when every group member is at the same point (no matter which). Now let one each of the two participants initially at .2 and 1 move .1 closer to the mean (to .3 and .9). This is homogenization, but not very much. The distribution is almost completely unaltered. The mean

stays at .6, while the variation around .6 shrinks just a smidge. Specifically, the standard deviation decreases to .230. The mean  $H_{gj}$  in this unimposing scenario is thus .015. The observed mean  $H_{gj}$  of .013 less than that.

### **Polarization**

Deliberation appears not to polarize attitudes at all, but rather to moderate them slightly. Somewhat more than half (54.6%) the group-issue pairs moderate rather than polarize. And, on average, the groups' attitudes move .022 *toward* the midpoint, moderating more than they polarize. This, in contrast to homogenization, is a difference of pre- and post-deliberation means rather than of pre- and post-deliberation standard deviations. So let us consider another simple scenario to contextualize this number. Take again a group of 20 members whose initial mean attitude is .6. Decreases in attitude score thus represent moderation. If just 5 of 20 participants decrease their scores, from no matter what starting-points, by just .1 apiece, the group moderates by .025. This is not much movement. The observed mean moderation of .023 is still a bit less. But the principal lesson here is not that the mean moderation is small but that, on average, the groups moderate rather than polarize.

### **Domination**

We consider three aspects of social advantage—gender, education, and income—in each case operationally splitting the sample in two, into the relatively advantaged and the relatively disadvantaged. For gender, the dividing line is straightforward: men versus women. For education and income, we divide each group at the whole DP-sample median, recognizing that what is highly educated or affluent in one place may not be in another. Social advantage is relative.

Overall, the results show little domination. Only minorities of the group-issue pairs show

any movement toward the initial mean position of the men, the better-educated, or the higher-income (46.0%, 48.6%, and 48.5%, respectively). On average, the movement is more toward than away from these initial mean positions, but the magnitudes are quite small: .011 toward the mean position of the men, .005 toward the mean position of the better educated, and less than .0005 toward the mean position of the higher income.

These means should be read in the same light as that for polarization.  $D_{gj}$ , like  $P_{gj}$ , registers pre- to post-deliberation attitude change. They differ only in sign—a matter of whether the change is toward (+) or away from (–) the nearer extreme in the case of  $P_{gj}$ , and of whether it is toward (+) or away from (–) the mean attitude of the advantaged, in the case of  $D_{gj}$ .<sup>31</sup> We have already seen what sparse, minor attitude changes yield a mean  $P_{gj}$  of .025. The same scenario, with the movements all being in the direction of the initial mean attitude of the advantaged, rather than necessarily toward the nearer extreme, would likewise yield a mean  $D_{gj}$  of .025. But the observed mean  $D_{gj}$  is only .011. If just 2 members of a group of 20 move by just .1 each toward the initial mean attitude of the advantaged—representing almost no domination by gender—that would yield roughly the same  $D_{gj}$  (.010). The mean domination by education is only half that, and the mean domination by income just a hair above 0.

Still, a further question lurks. We have considered domination with respect to one aspect of advantage at a time. In deliberations structured this way, men do not seem to dominate women, the better educated do not seem to dominate the less well educated, and the higher-income do not seem to dominate the lower-income. But what about those advantaged in all three respects? The definition of advantage affects  $D_{gj}$  only in the pre-deliberation mean attitude  $\bar{A}_{gj1}^a$ , movement toward or away from which constitutes domination or repulsion. The question is whether the groups tend to move toward the mean attitudes those who are male, better educated,

and higher income? The final columns of Table 2 report those results. They are no different in purport from the results for the three aspects of advantage considered individually: slightly fewer than half the group-issue pairs show movement toward the initial mean attitude of the advantaged, and the average movement in that direction, though positive, is extremely faint (and statistically far from significant). There is still not much to see here.

### Probing the Results

A couple of further questions immediately suggest themselves.

#### Correlations

A first is the correlations between these patterns. *A priori*, it is not clear whether or with what sign they should be correlated, except for Sunstein's (2002) argument that homogenization plays an important part in polarization (see also Sunstein and Hastie 2015). Table 3 shows the correlations between  $H_{gj}$ ,  $P_{gj}$ , and all three versions of  $D_{gj}$ . (Predictably, the correlations between their frequencies are extremely similar.) Only two correlations break .1 in magnitude, and one of those just barely. The correlation between domination by the more educated and polarization is -.130. To some slight degree the groups moving more toward the initial mean position of their more educated members tend to polarize less. The only notable correlation is between homogenization and polarization, at .359. Evidently, groups that homogenize more also tend to polarize more, consistent with Sunstein (2002). That is in fact mathematically necessary, on the plausible premise that the group members closer to the *further* pole (e.g., closer to 0 in a group whose mean > .5) tend to move more toward the group mean, disproportionately contributing to the shrinking of the variance. But, in contrast to the jury-like groups in Sunstein (2002) and Sunstein and Hastie (2015), the DP groups we examine here exhibit little homogenization—which, to the extent that homogenization does affect polarization, may help explain why we also

find little polarization.

### **Parsing Domination**

A second question concerns the movements involved behind domination.  $D_{gj}$  concerns the whole group's movement toward or away from the mean attitude of its advantaged members, which is what matters for the post-deliberation distribution of attitudes. But this is a weighted average of the movements of its advantaged and disadvantaged members. Suppose we break  $D_{gj}$  apart, examining the movements of the advantaged and disadvantaged separately.

In particular, the mean domination could be as small as it generally is because the disadvantaged are moving more than the whole group toward the mean attitude of the advantaged (being more “dominated”), while the advantaged are at the same time moving almost as much in the opposite direction, toward the pre-deliberation mean of the disadvantaged. This has a certain plausibility—and normative appeal. In exchanging arguments and information, the advantaged may be affecting the disadvantaged's attitudes *and vice versa*, each drawing the other in their direction.

Let us see. The domination/repulsion of the disadvantaged, as distinct from the whole group, can be defined by substituting  $\bar{A}_{gj2}^d - \bar{A}_{gj1}^d$  for  $\bar{A}_{gj2} - \bar{A}_{gj1}$  in the definition of  $D_{gj}$ . The results, displayed in Table 4, remain similar for all three aspects of advantage. In each case, the pattern for the disadvantaged group members is an accentuated version of what we have seen for the whole group. Compared to the whole group, the women, the less well educated, and the lower-income move toward the mean attitude of the men, the better educated, and the higher-income a somewhat higher percentage of the time (49.5% versus 46.0%, 56.5% versus 48.6%, and 52.1% versus 48.5%), and, on average, by more (.027 versus .011, .051 versus .005, and .036 versus .000).

That the disadvantaged move further than the whole group toward or beyond the mean initial attitude of the advantaged logically requires that the advantaged be moving less in that same direction. It does not require that they be moving in the opposite direction. But in fact they are, as Table 4 also shows. With respect to every aspect of advantage, the advantaged move at least roughly as often and roughly as much toward the initial positions of the disadvantaged as vice versa (indeed somewhat more so with respect to gender, if somewhat less so with respect to education). The men move toward the initial positions of the women, the more educated toward the initial positions of the less educated, and the higher-income toward the initial positions of the lower-income. Deliberation seems to be allowing each side of these demographic divides to be influenced by the other.

### **Discussion**

Whether any given deliberating group homogenizes, polarizes, or exhibits domination on any given policy issue is of no general consequence. On any given issue, some groups will homogenize, some will polarize, and some will exhibit domination. Some will not. Some will do the opposite. What matters for generalizing about and appraising the process of deliberation, under any given design, is the pattern across groups and issues—the extent to which group-issue pairs display routine and strong homogenization, polarization, and/or domination, as Sunstein (2000, 2002, 2009), among others, has claimed they do.

Several studies of single deliberative events dispute one or more of these alleged tendencies (Luskin, Fishkin, and Jowell 2002; Fishkin et al. 2010, Fishkin et al 2011), even in “enclave deliberation” (Grönlund et al. 2015). That suffices to create doubt, but a more convincing test requires larger numbers of groups and issues. That is the distinctive value, for this purpose, of the pooled data we examine here, encompassing 372 small groups, 139 policy

issues (counting each policy attitude index as tapping a somewhat different issue), and 2,601 group-issue pairs. This provides a good window on the frequency and strength of the homogenization, polarization, and domination that occurs.

The results are generally reassuring. At least in Deliberative Polling (and presumably other, similar designs) and at least on policy issues (and presumably other questions of action), deliberation does not routinely homogenize or polarize attitudes, nor does it routinely drive them toward those of the socially advantaged. In frequency, it does homogenize attitudes somewhat more than half the time, but it polarizes them and drives them toward those of the socially advantaged less than half the time. In magnitude, averaging across group-issue pairs, it slightly homogenizes attitudes but also slightly *moderates*, rather than polarizes, them. It barely moves them at all toward the attitudes of the advantaged. Yes, the attitudes of the disadvantaged (as distinct from those of the whole group) move a bit toward those of the advantaged, but the attitudes of the advantaged move nearly as much toward those of the disadvantaged! This nearly symmetric pattern is far from domination. On the contrary, it suggests some mutual influence, a normatively pleasing result.<sup>32</sup>

But the greatest take-away is that all these patterns are faint—a far cry from the picture of routine and strong homogenization and polarization in Sunstein (2002 and 2009) and that of routine and strong domination in Fraser (1993), Sanders (1997), and Young (2000). Thus the social dynamics that should produce those patterns must not be accounting for the lion's share of the observed attitude change. Rather, the absence of pattern is consistent with the idea that the participants are deliberating on the merits and thus moving toward their authentic attitudes. Compared with their pre-deliberation attitudes, these will sometimes have a smaller variance but sometimes have a larger one, sometimes be further out from the midpoint sometimes be closer to

or on the other side of it, and will sometimes lie in the same direction as the attitudes of the advantaged but sometimes lie in the other direction. In this light, the very wan patterns we see are exactly what we should expect to see.<sup>33</sup>

Granted, these results do not establish that these deliberations *are* substantially merits-driven. Nor do they exclude the possibility of distortions from other sources—from cognitive lapses or biases or from social dynamics other than those leading to homogenization, polarization, or domination. What they do show is that these three much written- and worried-about patterns do not obtain, suggesting that the particular social dynamics that would produce them are not much in play. There is no reason, on these grounds, to fear that any deliberation-induced policy attitude change is not largely merits-driven.

Granted, too, much apparently depends on the deliberative design. Judging from other studies, the results from jury-like designs exhibit more routine homogenization, polarization, and domination. But this, in our view, is less a blanket reservation about deliberation than a basis for choosing between designs. A design that routinely homogenizes and polarizes attitudes and moves them toward those of the socially advantaged is one in which the deliberation would appear to revolve around social dynamics, rather than the merits of the arguments. Every deliberative design has its own strengths and weaknesses, but these results suggest that those sharing some of Deliberative Polling's key characteristics—being about questions of action rather than fact; requiring no conscious, collective decision, much less a consensus; providing a good deal of balanced information; and involving relatively hands-off moderation—figure to produce less and weaker homogenization, polarization, and domination.

We are keenly aware of leaving questions for future research, including our own. What accounts for the variation across groups and issues? Why do some groups homogenize, polarize,



or move toward the views of their advantaged members more than others? Fuller-fledged explanatory models should suggest some answers. But these initial results suffice to shed some important light on the normative value of deliberation and the evaluation of the designs embodying it. In particular, they suggest that the deliberative distortions some have thought inevitable need not, in fact, arise. As in all democratic practice, deliberation's virtues and possible vices hinge on the design.

## References

- Althaus, Scott. 2003. *Collective Preferences in Democratic Politics: Opinion Surveys and the Will of the People*. Cambridge: Cambridge University Press.
- Bartels, Larry M. 1996. Uninformed Votes: Information Effects in Presidential Elections. *American Journal of Political Science*, 40 (1): 194-210.
- Beck, Paul Allen, Russell J. Dalton, Steven Greene, and Robert Huckfeldt. 2002. The Social Calculus of Voting: Interpersonal, Media, and Organizational Influences on Presidential Choices. *American Political Science Review*, 96(1): 57-73
- Bohman, James and William Rehg (Eds.). 1997. *Deliberative Democracy: Essays on Reason and Politics*. Cambridge: MIT Press.
- Cohen, Joshua 1989. "Deliberation and Democratic Legitimacy," in A. Hamlin and P. Pettit eds. *The Good Polity* (Oxford: Blackwell).
- Condorcet, Nicolas (Marquis) de. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*.
- Conover, Pamela Johnston and Donald D Searing. 2005. Studying 'Everyday Political Talk' in the Deliberative System. *Acta Politica*, 40: 269-283.
- Delli Carpini, M.X. 2005. "An overview of the state of citizens' knowledge about politics". In M. S. McKinney, L. L. Kaid, D. G. Bystrom, & D. B. Carlin (Eds.), *Communicating politics: Engaging the public in democratic life* (pp. 27-40). New York: Peter Lang.
- Dryzek, John 2002. *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. Oxford: Oxford University Press.
- Elster, Jan. (Ed.). 1998. *Deliberative Democracy*. Cambridge: Cambridge University Press.
- Fishkin, James S. 1991. *Democracy and Deliberation: New Directions for Democratic Reform*.

New Haven: Yale University Press.

- Fishkin, James S., Baogang He, Robert C. Luskin, and Alice Siu. 2010. Deliberative Democracy in an Unlikely Place: Deliberative Polling in China. *British Journal of Political Science*, 40: 435-448.
- Fishkin, James S., Thad Kousser, Robert C. Luskin, and Alice Siu. 2011. Deliberative Agenda Setting: Piloting Reform of Direct Democracy in California. *Perspectives on Politics*. Vol 13, no. 4, 1030-1042.
- Fraser, Nancy 1993. "Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy." In Bruce Robbin, ed., *The Phantom Public Sphere*. Minneapolis, MN: University of Minnesota Press.
- Gastil, John, E., Pierre Dees, Philip J. Weiser, and Cindy Simmons 2010. *The Jury and Democracy: How Jury Deliberation Promotes Civic Engagement and Political Participation*. Oxford: Oxford University Press.
- Grönlund, Kimmo, Kaisa Herne, and Maija Setälä. 2015. Does Enclave Deliberation Polarize Opinions? *Political Behavior*, 37 (4): 995–1020.
- Gutmann, Amy and Dennis Thompson 2004. *Why Deliberative Democracy?* Princeton: Princeton University Press.
- Habermas, Jürgen. 1990. *Discourse Ethics: Notes on a Program of Philosophical Justification*. In S. Benhabib and F. Dallmayer eds. *The Commnicative Ethics Controversy*. Cambridge, MA: MIT Press, pp. 60-110.
- Habermas, Jürgen. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge, MA: MIT Press.

- Huckfeldt, Robert and John Sprague. 1995. *Citizens, Politics, and Social Communication*. Cambridge: Cambridge University Press.
- Karpowitz, Christopher F. and Tali Mendelberg. 2014. *The Silent Sex: Gender, Deliberation, and Institutions*. Princeton, NJ: Princeton University Press.
- Karpowitz, Christopher F., Tali Mendelberg, and Lee Shaker. 2014. Gender Inequality in Deliberative Participation. *American Political Science Review*, 106 (3): 533-547.
- Lindell, Marina, André Bächtiger, Kimmo Grönlund, Kaisa Herne, Maija Setälä, and Dominik Wyss. 2017. What drives the polarisation and moderation of opinions? Evidence from a Finnish citizen deliberation experiment on immigration. *European Journal of Political Research*, 56 (1): 23–45.
- List, Christian, Robert C. Luskin, James S Fishkin, Ian McLean. 2013. “Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls.” *Journal of Politics*, 75 (1) 80–95.
- Luskin, Robert C. 1987. Measuring Political Sophistication, *American Journal of Political Science*, 31: 856-99.
- Luskin, Robert C. 2003. The Heavenly Public: What Would the Ideal Democratic Citizenry Be Like? In Michael B. MacKuen and George Rabinowitz (eds.), *Electoral Democracy*. Ann Arbor: University of Michigan Press.
- Luskin, Robert C., Fishkin, James S., and Roger Jowell. 2002. “Considered Opinions: Deliberative Polling in Britain.” *British Journal of Political Science*, 32:455-87.
- Manin, Bernard. 1995. *The Principles of Representative Government*, (Cambridge: Cambridge University Press.
- Mansbridge, Jane 1999. “On the Idea that Participation Makes Better Citizens” in Stephen Elkin

- and Karol Edward Soltan (Eds) *Citizen Competence and Democratic Institutions*. State College, PA: Penn State University Press.
- Mendelberg, Tali. 2002. "The Deliberative Citizen: Theory and Evidence" in Michael X. Delli Carpini, Leonie Huddy, Robert Y. Shapiro eds. *Political Decision-Making, Deliberation and Participation*. Research in Micropolitics, Vol 6. Amsterdam: JAI.
- Mutz, Diana 2006. *Hearing the Other Side: Deliberative Versus Participatory Democracy*. Cambridge: Cambridge University Press.
- Myers, C. Daniel. 2017. "Interests, Information, and Minority Influence in Deliberation." *Journal of Politics* 79 (3): 804-822.
- Neuman, W. Russell. 1986. *The Paradox of Mass Politics: Knowledge and Opinion in the American Electorate*. Harvard University Press.
- Sanders, Lynn M. 1997. "Against Deliberation." *Political Theory*, 25(3): 347-376.
- Searing, Donald D., Frederick Solt, Pamela J. Conover, and Ivor Crewe. 2007. Public Discussion in the Deliberative System: Does it Make Better Citizens? *British Journal of Political Science*, 37: 587-618.
- Siu, Alice. 2009. *Look who's talking: Examining social influence, opinion change, and argument quality in deliberation*. PhD dissertation, Department of Communication, Stanford University.
- Sunstein, Cass R. 2000. Deliberative Trouble? Why Groups Go to Extremes. *The Yale Law Journal*, 110(1): 71-119.
- Sunstein, Cass R. 2002. The Law of Group Polarization. *The Journal of Political Philosophy*, 10(2): 175-195.
- Sunstein, Cass R. 2009. *Going to Extremes: How Like Minds Unite and Divide*. Oxford: Oxford

University Press.

Sunstein, Cass R. and Reid Hastie 2015. *Wiser: Getting Beyond Groupthink to Make Groups*

*Smarter*. Boston: Harvard Business Review Press.

Vidmar, Neil and Valerie P. Hans 2007. *American Juries: The Verdict*. Amherst, NY:

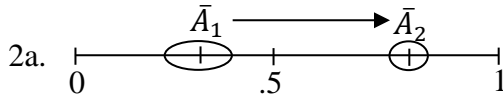
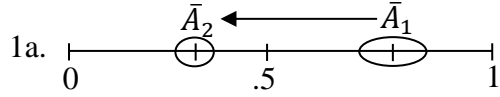
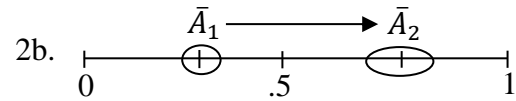
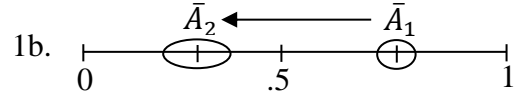
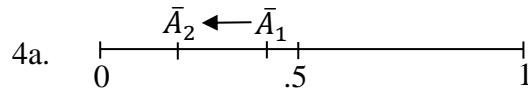
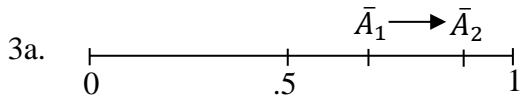
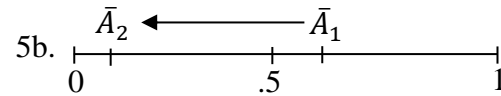
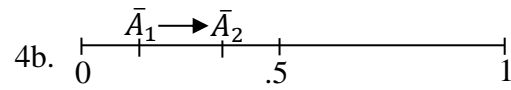
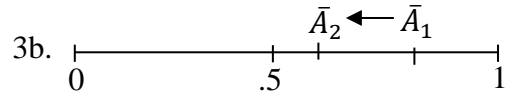
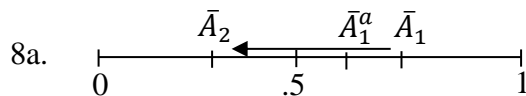
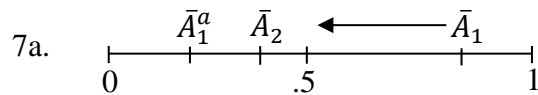
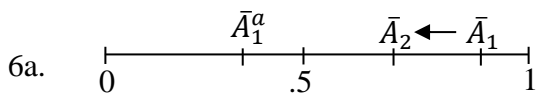
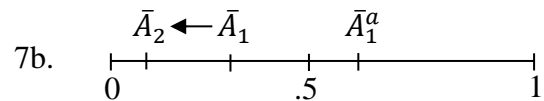
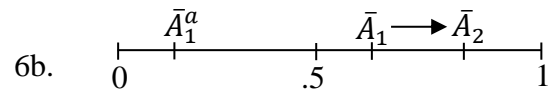
Prometheus Books.

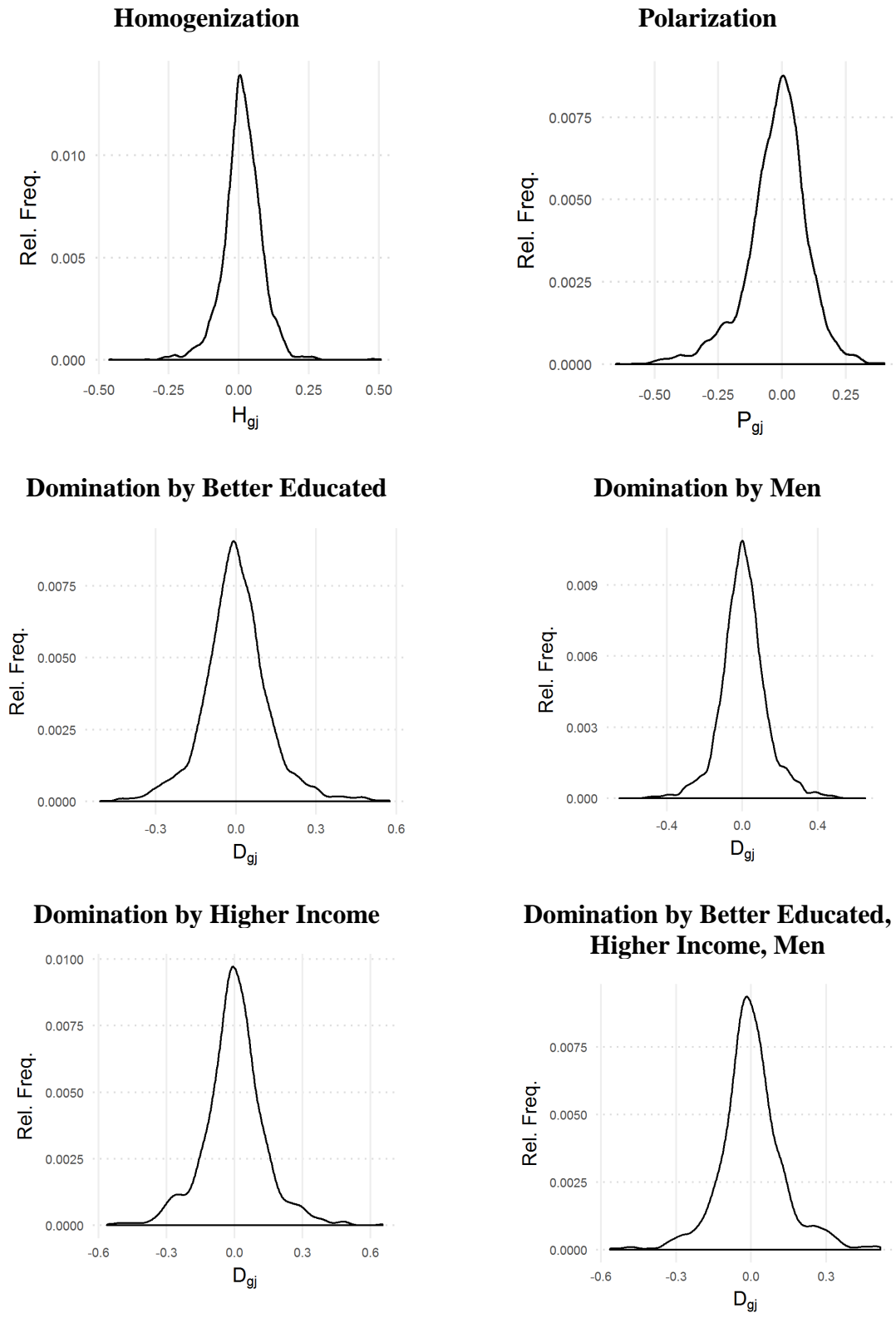
White, Halbert A. 1980. Heteroskedasticity-consistent covariance matrix estimator and a direct

test for heteroskedasticity. *Econometrica*, 817–838.

Young, Iris M. 2000. *Inclusion and Democracy*. Oxford, U.K.: Oxford University Press.

Figure 1: Illustrating the Definitions

**A. Homogenization/Variation** $H > 0$  (Homogenization) $H < 0$  (Variation)**B. Polarization/Moderation** $P > 0$  (Polarization) $P < 0$  (Moderation)**C. Domination/Repulsion** $D > 0$  (Domination) $D < 0$  (Repulsion)

**Figure 2: Distributions of Group-Issue Pairs on  $H$ ,  $P$ , and  $D$** 



**Table 1: DPs Analyzed**

<b>Broad Topic(s)</b>	<b><i>n</i></b>	<b>Country/ Countries</b>	<b>City/ Region</b>	<b>Year</b>	<b>Policy Mode</b>	<b>Indices</b>	<b>Groups</b>
Britain's Role in the EU	238	U.K.		1995	F2F	4	16
National Health System	230	U.K.		1998	F2F	11	15
British Monarchy	258	U.K.		1996	F2F	4	15
UK General Election	275	U.K.		1997	F2F	4	15
Making Australia a Republic	347	Australia		1999	F2F	5	24
Crime	299	U.K.		1994	F2F	5	20
EU Expansion, Pension Reform, Foreign Policy	344	E.U.*		2007	F2F	7	18
Meeting future electricity needs	216	U.S.	CP&L Service Area <sup>†</sup>	1996	F2F	7	16
Prioritizing public works projects	233	China	Zeguo Township	2005	F2F	9	16
Crime	278	Bulgaria		2007	F2F	5	17
Climate Change, Immigration	348	E.U.*		2009	F2F	2	25
Future of Airport, Revenue-Sharing	132	U.S.	New Haven, CT	2004	F2F	3	16
Foreign policy	340	U.S.		2003	F2F	9	24
Foreign Policy	245	U.S.		2003	OL	9	15
US General Election	246	U.S.		2004	OL	6	15
US Presidential Primaries	434	U.S.		2004	OL	3	16
Housing Policy	239	U.S.	San Mateo, CA	2008	F2F	7	26
Health Care, Education	454	U.S.		2005	OL	11	30
Meeting future electricity needs	230	U.S.	WTU Service Area <sup>†</sup>	1996	F2F	7	14
Crime, the family, foreign policy	466	U.S.		1996	OL	9	30
Meeting future electricity needs	232	U.S.	SEP Service area <sup>†</sup>	1996	F2F	7	14
<b>Total</b>	<b>6084</b>					<b>134</b>	<b>397</b>

\*All (then 27) Member-States.

<sup>†</sup>WTU = West Texas Utilities, CP&L = Central Power & Light, SEP = Southwestern Electric Power.

Table 2: Mean Homogenization, Polarization, and Domination

	Homogenization		Polarization				Domination					
	Freq	Mean	Freq	Mean	by Gender		Education		by Income		by All Three	
					Freq	Mean	Freq	Mean	Freq	Mean	Freq	Mean
<b>Mean*</b>	.595	.013	.454	-.022	.460	.011	.486	.005	.485	.000	.487	.014
<b>s.e</b>	.017	.003	.027	.010	.011	.004	.012	.005	.017	.005	.020	.004
<b><i>p</i></b>	.000	.000	.088	.032	.002	.031	.274	.325	.425	.951	.674	.368

**Table 3: Correlations between Homogenization, Polarization, and Domination****A. Magnitude ( $H_{gj}$ ,  $P_{gj}$ ,  $D_{gj}$ )**

	<i>P</i>	<i>D</i> (Gender)	<i>D</i> (Educ.)	<i>D</i> (Income)
<i>H</i>	.359	-.007	-.064	.027
<i>P</i>		.033	-.130	.042

**B. Frequency ( $H'_{gj}$ ,  $P'_{gj}$ , and  $D'_{gj}$ )**

	<i>P</i>	<i>D</i> (Gender)	<i>D</i> (Educ.)	<i>D</i> (Income)
<i>H</i>	.334	.008	.002	.044
<i>P</i>		-.003	.022	.061

**Table 4: Parsing Domination: Mean Movements of the Disadvantaged toward the Advantaged—and Vice Versa**

<b>By</b>	<b>Women</b>		<b>Men</b>		<b>Less Educated</b>		<b>More Educated</b>		<b>Lower Income</b>		<b>Higher Income</b>	
	<b>Toward</b>				<b>Men</b>		<b>Women</b>		<b>More Educated</b>		<b>Less Educated</b>	
	<b>Freq</b>	<b>Mean</b>	<b>Freq</b>	<b>Mean</b>	<b>Freq</b>	<b>Mean</b>	<b>Freq</b>	<b>Mean</b>	<b>Freq</b>	<b>Mean</b>	<b>Freq</b>	<b>Mean</b>
<b>Mean*</b>	.495	.027	.560	.033	.565	.051	.547	.030	.521	.036	.535	.043
<b>s.e.</b>	.009	.003	.015	.006	.013	.005	.014	.006	.017	.007	.016	.006
<b><i>p</i></b>	.560	.000	.000	.000	.000	.000	.001	.000	.203	.000	.033	.000

---

NOTES

<sup>1</sup>And at least a certain, presumably low threshold of cognitive ability, allowing for the possibility that some small portion of the population may not be able to reach authentic attitudes even with unlimited information and opportunity for reflection.

<sup>2</sup>“Polarization,” in this sense, is the variable popularized by Sunstein (2000, 2002). His definition (2000) refers to the median attitude, but his analyses, like our definition, use the mean. In other literatures, “polarization” refers instead to sets of entities (notably, parties) moving further apart. Any pair of groups polarizing in the first sense must also polarize in the second if they start on opposite sides, but may not if they start on the same side.

<sup>3</sup>We take individual rumination (as in Goodin’s (2000) “deliberation within”) as a related (intervening) but distinct phenomenon.

<sup>4</sup>This is, as we say, a minimum. Further stipulations may be added without affecting the argument or results to follow.

<sup>5</sup>Habermas (1996, p. 323) describes this as a “thought experiment” (applicable to all validity claims, not just political ones), while also sketching what he considers attainable conditions for actual deliberative procedures (1996, pp. 304-306; and, similarly, Cohen 1989).

<sup>6</sup>Mutz (2006) claims to be talking about “deliberation” but only examines cross-cutting exposure,” meaning contact of any sort with others holding differing views.

<sup>7</sup>Habermas’s assertion that the everyday, undeliberated “preferences ... in [standard] opinion polls do not reliably reflect” those the respondents would hold “after weighing the relevant information and arguments” (1998, p. 336) is very much in this spirit.

---

<sup>8</sup>The proviso is that everyone, regardless of authentic attitude, narrow the gap between his or her actual attitude and authentic attitudes to the same degree. If that is sufficiently far from true—if, e.g., those on one side of the issue move markedly toward their authentic preferences, while those on the other side do not—the overall mean attitude will not necessarily wind up closer to the mean authentic one.

<sup>9</sup>“Actual,” here, is thus a two-dimensional qualifier—“actual” as opposed to “authentic,” but also “actual” as opposed to “observed.”

<sup>10</sup>One approach is to use statistical models to calculate the attitude each individual would have if given the maximum score on measured knowledge (as in Bartels 1996, Delli Carpini and Keeter 1996, Althaus 2003). This “manipulates” only knowledge, but accumulated knowledge (correct, confidently held information), in natural settings, is highly correlated with thought (Neuman 1986, Luskin 1987). Well-designed deliberative fora manipulate more of the variables boosting authenticity, exposing their participants to additional information and perspectives, forcing them to confront uncongenial and unfamiliar information and perspectives, and encouraging them to think about what their attitudes should be (as in, e.g., Luskin, Fishkin, and Jowell 2002; Fishkin et al. 2012). In both approaches, however, the credibility of the estimates depends on the details—non-exhaustively, on the specification of the model and the measurement of knowledge in the first approach and on the quantity, quality, and range of the information and discussion in the second. Both sorts of estimates, moreover, are clouded by questions of linearity and monotonicity. What if the maximum knowledge score represented more knowledge than it does? What if the deliberative experience were richer or longer? We cannot know.

---

<sup>11</sup>The *perception* of influence (also examined by Karpowitz et al. 2012) is also different—presumably correlated with but not the same as actual influence, which in turn presumably affects but is not the same as domination.

<sup>12</sup>Homogenization could instead be defined as a difference or ratio of variances, but such redefinitions would tell essentially the same story, and the difference of standard deviations has the interpretive advantage of preserving the original units of measurement.

<sup>13</sup>Groups starting at .5 cannot be considered as moving further out (or in) on the same side, given that they have not started on either side.

<sup>14</sup>We exclude those few group-issue pairs for which  $\bar{A}_{gt}^a = \bar{A}_{gj1}$ . It is impossible for the group's mean attitude to move toward that of its advantaged members when they are already identical.

<sup>15</sup>The reason that  $P_{gj}$  can be larger in magnitude at its most negative (qua moderation) than at its most positive (qua polarization) is that its positive values gauge movement on the “short side of the field.” Perforce, the group mean cannot move as far toward the nearer pole as toward the further one. This does not, however, tilt the results toward moderation. It is no easier to get a  $P_{gj}$  of, say,  $-3$  than of  $+3$ , and we virtually never actually see  $P_{gj} < -.5$ . (See Figure 2 below.)

<sup>16</sup>The difference between these two scenarios lies in the ordinal relation between the means for the advantaged and disadvantaged: which starts at 0 or 1, and which is just toward the midpoint from it.

<sup>17</sup>While noting the existence of moral disagreement, Cohen (1989) and Gutmann and Thompson (1996) posit that deliberation should identify reasons “acceptable to all.” Habermas admits that pressure for decision may require majority rule, but says that “deliberations aim in general at rationally motivated agreement” (1999, p. 306).

---

<sup>18</sup>Lindell et al. (2017) raise the same question, but at the level of the individual participant and basing their answer on the different, largely empirical criterion of the effects of variables like education, knowledge, and the perceived deliberativeness of the discussion.

<sup>19</sup>Implausibly, these social dynamics could only or very disproportionately affect those whose authentic attitudes happen to be closer than their pre-deliberation attitudes to the group's pre-deliberation mean, closer than their pre-deliberation attitudes to the nearer pole, or closer than their pre-deliberation attitudes to the pre-deliberation attitudes of the advantaged. That would allow socially-driven homogenization, polarization, and domination to coexist with increased authenticity. But there is no reason to expect such selective homogenization, polarization, or domination if the attitude change is socially driven, only if it is merits-driven.

<sup>20</sup>Except when the right answer is exactly .5, in which case, the movement to it would produce homogenization but not polarization. But, then, the meaning of polarization, for groups starting at .5 is unclear. See n. 13.

<sup>21</sup>Suppose, for instance, that some participants enter believing in either of two wrong answers, on either side of the right one and, in reduced numbers, to continue doing so after deliberating. It remains hard to find plausible configurations of parameters under which simply gravitating toward the right answer does not produce homogenization and polarization. Further details available on request.

<sup>22</sup>Everyone reading this probably agrees that the question of whether allow human sacrifice has a morally right answer. It is uncontested. But is it right in the same way, and is it as incontestable (as distinct from uncontested), as the answer to the question of whether average temperatures around the world are currently rising? In the first case, any contest would at least implicitly need



---

to invoke moral principle; in the second only evidence. This is a deep question, into which we need go no further. At least among the topics posed for public deliberation, those of fact have a right answer in way that those of action do not.

<sup>23</sup>Habermas distinguishes between the “public will formation” in decision making and “public opinion formation,” referring to viewpoints formed and expressed apart from decision making, though possibly as an input to it.

<sup>24</sup>This general description holds for all but two face-to-face DPs, the exceptions being the two pan-European ones, where it was impossible to manage simultaneous interpretation for more than four or five languages per small group. Thus the assignment of participants to small groups could be random only within language.

<sup>25</sup>The first three online DPs relied on the online polling firm Knowledge Networks for (random) sampling and recruitment. The other two included in these data, administered by Polimetrix/YouGov, could only approximate random samples. In addition, as noted above, all the online DPs assigned participants to groups in less than fully random fashion, constrained by the times at which participants could “meet.”

<sup>26</sup>Not  $372 \times 139 = 51,708$ , since each index is confined to just one DP and its small groups.

<sup>27</sup>Some of the online DPs approximate random sampling by matching.

<sup>28</sup>A given group may behave—homogenize, polarize, or exhibit domination—similarly across issues, making the group-issue pairs constituting the observations on  $H_{gj}$ ,  $P_{gj}$ , and  $D_{gj}$  dependent. The Huber-White estimates cluster by policy index within each DP (as in White 1980).

<sup>29</sup>This last is especially a stretch for homogenization, defined as a difference of standard deviations, rather than of means.

---

<sup>30</sup>Thus we take  $E(H_{gj}) > 0$ ,  $E(P_{gj}) > 0$ , and  $E(D_{gj}) > 0$  as the alternative hypotheses in calculating  $p$  values.

<sup>31</sup>More precisely, both  $P_{gj}$  and  $D_{gj}$  can be expressed as the absolute mean attitude change,  $|\bar{A}_{gj2} - \bar{A}_{gj1}|$ , times a dummy variable,  $S_{gj}$ . For  $P_{gj}$ ,  $S_{gj} = 1$  for movement toward the nearer extreme and  $= -1$  for movement away from it, whereas, for  $D_{gj}$ ,  $S_{gj} = 1$  for movement toward the mean position of the advantaged, and  $-1$  for movement away from it.

<sup>32</sup>For what it is worth (the policy attitude results remain the bottom line), transcripts of the small group discussions in several DPs, analyzed by Siu (2009) tell a compatible story. The socially disadvantaged and advantaged participate roughly equally (whether gauged by the number of words spoken or the length of time speaking). The male-female differences are routinely small; those between the more and less educated and between the higher and lower income vary more but are on average quite small.

<sup>33</sup>Consistent with analyses of individual DPs peeking into the workings of policy attitude change (Luskin, Fishkin, and Jowell 2002, Fishkin et al. 2011, Gerber et al. 2014).