


ARTICLE

# Deliberative Distortions? Homogenization, Polarization, and Domination in Small Group Discussions

Robert C. Luskin<sup>1\*</sup> , Gaurav Sood<sup>2</sup>, James S. Fishkin<sup>3</sup> and Kyu S. Hahn<sup>4</sup>

<sup>1</sup>University of Texas at Austin, Austin, TX, USA, <sup>2</sup>Convoy, Inc., Seattle, WA, USA, <sup>3</sup>Stanford University, Palo Alto CA, USA and <sup>4</sup>Seoul National University, Seoul, Republic of Korea

\*Corresponding author. E-mail: [rluskin@mail.utexas.edu](mailto:rluskin@mail.utexas.edu)

(Received 19 July 2017; revised 15 May 2020; accepted 27 March 2021; first published online 7 April 2022)

## Abstract

Deliberation is widely believed to enhance democracy by helping to refine the ‘public will’, moving its participants’ policy attitudes closer to their ‘full-consideration’ policy attitudes – those they would hypothetically hold with unlimited information, to which they gave unlimited reflection. Yet there have also been claims that the social dynamics involved generally ‘homogenize’ attitudes (decreasing their variance), ‘polarize’ them (moving their means toward the nearer extreme), or engender ‘domination’ (moving their overall means toward those of the attitudes held by the socially advantaged) – attitude changes that may often be *away* from the participants’ full-consideration attitudes and may thus distort rather than refine the public will. This article uses 2,601 group-issue pairs in twenty-one Deliberative Polls to examine these claims. Reassuringly, the results show no routine or strong homogenization, polarization, or domination. What little pattern there is suggests some faint homogenization, but also some faint *moderation* (as opposed to polarization) and *opposition* (as opposed to domination) – all as is to be expected when the outside-world forces shaping pre-deliberation attitudes are slightly more centrifugal than centripetal. The authors lay out a theoretical basis for these expectations and interpretations and probe the study’s results, highlighting, among other things, deliberation’s role in undoing outside-world effects on pre-deliberation attitudes and the observed homogenization’s, polarization’s, and domination’s dependence on deliberative design.

**Keywords:** deliberation; discussion; deliberative democracy; public opinion; policy attitudes; centrifugal and centripetal forces

Deliberation, in the sense of elevated discussion (roughly, an open-minded weighing of the arguments and evidence for and against competing alternatives), is generally seen as enhancing democracy (Bohman and Rehg 1997; Dryzek 2002; Elster 1998; Fishkin 1991) – perhaps most centrally by refining the ‘public will’ that democracy translates into policy choices. The participants acquire a better sense of what policies they *should* favor, in light of their own values and interests, moving their policy attitudes toward those they *would* hold with the benefit of unlimited information and thought.<sup>1</sup> This effect stems from deliberation’s defining properties, those making it more than just any discussion.

But discussions also have non-deliberative effects, products of their social dynamics. These may plausibly lead a deliberating group’s participants to: (1) converge on the same attitude (what we shall term *homogenization*), (2) adopt more extreme attitudes on whichever side of the issue the group started on (commonly termed *polarization*), or (3) adopt attitudes closer

<sup>1</sup>We focus on ‘attitudes’, but our arguments also apply, with minor tweaks, to ‘preferences’.

to those of their more socially advantaged (male, better-educated, more affluent) co-deliberators (to be termed, more hesitantly, *domination*). These are all varieties of group-level attitude change – as distinct from their sources and consequences, a point particularly worth stressing for ‘domination’, which in other, nearby usage (for example, Squires 2008) denotes underlying dialogic inequalities. Here it simply means attitude change toward the attitudes of the advantaged.

A widely held, if not always fully articulated, concern about these particular attitude changes is that they may be mostly *away* from the attitudes the participants would hold with the benefit of unlimited information and thought.<sup>2</sup> They are, in that case, our title’s ‘distortions’. Discussions routinely producing them would be warping rather than refining the public will. But, as the title’s punctuation suggests, this leaves some questions. To what extent do these putative distortions actually prevail? And to what extent is it the deliberation in the discussion that produces them – making them *deliberative* distortions?

Here we tackle these questions with data from twenty-one Deliberative Polls (DPs), on various issues, in various contexts, encompassing 2,601 group-issue pairs. The core empirical analysis addresses the first question directly and affords inferences addressing the second. For these purposes, size (the dataset’s) matters. Any one group, discussing any one issue, in any one context, may or may not exhibit homogenization, polarization, or domination, no matter what the discussion is like. It is the distributions across groups, issues, and contexts that are revealing. We preface this analysis by laying out the underlying theory and follow it by considering nuances, possible objections, and implications for deliberative design.

## The Conceptual Terrain

We begin by sketching some key concepts.

### *Deliberation, Policy Attitudes, and Associated Cognition*

#### *Deliberation*

What lifts *deliberation* above mere discussion is its being (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 the policy implications neither all of one kind nor all on one side. (3) The participants react to each other’s arguments and information. And (4) they seriously and even-handedly (re)consider, in light of the discussion, what their policy attitudes should be. In short, deliberation requires that its participants engage in a serious, open-minded, even-handed weighing of the merits.<sup>3</sup> It does *not* require consensus-seeking or conscious, collective decision-making (cf. Cohen 1989; Gutmann and Thompson 1996; Gutmann and Thompson 2004). It may or may not yield a consensus. It may but need not affect subsequent decision-making by other bodies (as many DPs have done). It may even – optionally – involve conscious, collective decision-making itself, although that may alter the discussion’s effects, in ways we consider below.

Realistically, ‘deliberation’ is not a discrete property – something that does or does not occur – but the high end of a continuum (Fishkin 1991). Some discussions are highly deliberative, others hardly at all. The unattainable top of the range is something like Habermas’s (1990) ‘ideal speech situation’: a thought experiment in which every argument is made and countered, and in which everyone weighs all the arguments and counterarguments, free of all coercion. The bottom is

<sup>2</sup>Construed as agreement on a *course of action*, the homogenization resulting from compromise may be seen as *desirable* (as per Cohen 1989; Gutmann and Thompson 1996; Habermas 1996). It may minimize post-decision grievance and strife. But the underlying *policy attitudes* – our present concern – need not have converged to achieve compromise. Indeed, the notion of *compromise* presumes they have not.

<sup>3</sup>A discussion’s deliberativeness is thus an aggregative property – a function of its participants’ mentation and communication. Solitary deliberation is a degenerate case, in which the group size is 1, and the ‘dialogue’ internal.

vacuity: nobody says anything of substance. In these terms, the great majority of naturally occurring discussions fall much nearer the continuum's bottom than its top. By and large, they involve little focus on seriously weighing the merits, and the participants have little knowledge to share (Bennett, Flickinger and Rhine 2000; Kinder and Kalmoe 2017; Luskin 1987), are demographically similar and attitudinally like-minded (Bennett, Flickinger and Rhine 2000; Butters and Hare 2020; McPherson, Smith-Lovin, and Cook 2001), circumnavigate whatever few areas of disagreement exist (Bennett, Flickinger and Rhine 2000; Cowan and Baldassarri 2018; Gerber et al. 2012), and discount whatever little counterattitudinal information may nevertheless poke through (Lodge and Taber 2013). This is a far cry from 'deliberation' (Mansbridge 1999a; Mansbridge 1999b).

It would therefore be a mistake to regard studies of naturally occurring discussions (as in, for example, Beck et al. 2002; Huckfeldt and Sprague 1995; Mutz 2006, or Searing et al. 2007) as saying much about deliberation. For that, we need *deliberative designs*: discussions organized to be more much deliberative than the vast majority of those in everyday life. Examples include Consensus Conferences and Citizens' Juries (thumbnailed by Ozanne, Corus and Saatcioglu 2009), as well as DPs. In varying ways, and with varying success, these all get their participants to talk more about policy issues, to learn and think more about them, and to do so in a more earnest, open-minded way.

### Policy Attitudes

We take *policy attitudes* to be evaluations of policy options: how much one favors or opposes  $X$  or favors or opposes  $X$  over  $Y$ .<sup>4</sup> A policy attitude is thus 'positional' – expressible as a point on a numerical continuum (taken here to run from 0 to 1, with 0.5 representing neutrality). We denote the  $i$ th individual's time- $t$  attitude on the  $j$ th issue by  $A_{ijt}$ , where  $t = 1, 2$  (pre- versus post-deliberation).

### Policy-Attitude-Associated Cognition

Assorted *cognitions* (perceptions, beliefs, perspectives) may underpin  $A_{ijt}$ . Collectively these may be more or less *complex* (numerous and cognitively interconnected), more or less *factually accurate*, and more or less *balanced* (congenial, in equal or representative proportions, to opposing sides). Note that these cognitive variables – call them  $C_{ijt}$ ,  $F_{ijt}$ , and  $B_{ijt}$  – are conceptually distinct from the attitudes they support. The time- $t$  attitude is just  $A_{ijt}$ , no matter the  $C_{ijt}$ ,  $F_{ijt}$  or  $B_{ijt}$  behind it. Individuals 1 and 2 may have the same attitude ( $A_{1jt} = A_{2jt}$ ) even if, for example,  $C_{1jt} > C_{2jt}$ , making 1's attitude better 'developed' or 'crystallized' (further from what Converse (1970) called a 'non-attitude'). Similarly, the time-1 to time-2 attitude change is just  $A_{ij2} - A_{ij1}$ , however much or little  $C_{ijt}$ ,  $F_{ijt}$ , or  $B_{ijt}$  may have changed.

### Full-Consideration Attitudes

The attitudes people have are not necessarily those they *would* have with the benefit of unlimited information and reflection. Denote the  $i$ th individual's *full-consideration attitude* on the  $j$ th issue as  $A_{ij}^*$ . Axiomatically, we take this to be the attitude most closely aligned with his or her values and interests.  $A_{ij}^*$  is thus close kin to Lau and Redlawsk's (1997) 'correct votes', Mansbridge's (1983) 'enlightened preferences', and the 'full-information' votes and policy attitudes simulated by Bartels (1996), Delli Carpini and Keeter (1996), and Althaus (2003), among many others. Now, we never *know*  $A_{ij}^*$ . Even estimating it can be tricky (Luskin 2003). But here we need it only as a conceptual touchstone, for which we need only posit its existence.

<sup>4</sup>Consistent with prevailing definitions and usage (see Albarracín and Shavitt 2018; Bohner and Dickel 2011). Usage has long been in this spirit, sensibly unidimensional. Definitions, which often used to confound 'attitude' with cognition and/or behavior, have belatedly caught up.

*Appropriateness*

In similar vein, we may define an attitude’s *appropriateness* (for the individual holding it) as its proximity to its holder’s full-consideration attitude:  $\alpha_{ijt} \equiv 1 - |A_{ijt} - A_{ij}^*|$ . Thus  $\alpha_{ijt} = 1$  when  $A_{ijt} = A_{ij}^*$  and  $= 0$  when  $A_{ijt} = 1$  and  $A_{ij}^* = 0$ , or vice versa.

**Homogenization, Polarization, and Domination**

Our focal variables are all species of group-level attitude change. It sometimes makes sense to treat them as dichotomies, simply distinguishing cases in which they occur from those in which they do not. More informatively, however, they can be treated as continua centered at 0. Their names express their worried-about sides (taken to be numerically positive), but the opposite sides (taken to be numerically negative) also exist: a group’s attitudes may exhibit *homogenization* or *variegation* (decreasing or increasing variance); *polarization* or *moderation* (movement toward or away from the nearer extreme), and *domination* or *opposition* (movement toward or away from the attitudes of the group’s socially advantaged members).

To formalize these variables, let  $A_{ijt}$ ’s time- $t$  mean within the  $g$ th group be  $\bar{A}_{gjt}$ ; its time- $t$  standard deviation within the  $g$ th group be  $s_{gjt}$ ; and, assuming some mutually exclusive, exhaustive division into advantaged and disadvantaged, its time- $t$  mean for the  $g$ th group’s advantaged and disadvantaged members be  $\bar{A}_{gjt}^a$  and  $\bar{A}_{gjt}^d$ . In these terms:

The *homogenization* of the  $g$ th group’s attitudes on the  $j$ th issue is:<sup>5</sup>

$$H_{gj} = s_{gj1} - s_{gj2}, \tag{1}$$

which  $> 0$  for homogenization,  $< 0$  for variegation, and  $= 0$  for neither. Figure 1 illustrates, representing the within-group variation by more or less elongated ellipses. Regardless of what happens to the mean (compare Panels A1a and A2a with Panels A1b and A2b), a decreasing variance (as in Panels A1a and A2a) is homogenization, an increasing one (as in Panels A1b and A2b) variegation.  $H_{gj}$  is at its most positive (0.5) when the participants are evenly split between the polar attitudes (half at 0, half at 1) before deliberating but all have exactly the same attitude (whatever it may be) after doing so – changing, that is, from perfect dissensus ( $s_{gj1} = 0.5$ ) to perfect consensus ( $s_{gj2} = 0$ ). It is at its most negative (–0.5) for the opposite change, from perfect consensus to perfect dissensus. The binary version is  $H_{gj}^b = 1$  if  $H_{gj} > 0$  and  $= 0$  if  $H_{gj} \leq 0$ . Redundantly, though perhaps usefully for later exposition, the complementary binary variable for variegation can be defined as  $V_{gj}^b = 1$  if  $H_{gj} < 0$  and  $= 0$  if  $H_{gj} \geq 0$ .<sup>6</sup>

The *polarization* of the  $g$ th group’s attitudes on the  $j$ th issue is:

$$P_{gj} = (\bar{A}_{gj2} - \bar{A}_{gj1})S_{gj1}, \tag{2}$$

where  $S_{gj1}$  indicates the  $g$ th group’s time-1 side on the  $j$ th issue:  $S_{gj1} = 1$  for  $\bar{A}_{gj1} > 0.5$  and  $= -1$  for  $\bar{A}_{gj1} < 0.5$ . The multiplication by  $S_{gj1}$  ensures that  $P_{gj} > 0$  for polarization,  $< 0$  for moderation, and  $= 0$  for neither (no mean attitude change).<sup>7</sup> Panels B3a and B4a show  $\bar{A}_{gj1}$  moving toward the nearer extreme (polarization); Panels B3b, B4b, and B5b show it moving in the opposite direction, toward or beyond the midpoint (moderation).<sup>8</sup>  $P_{gj}$  is at its most positive (just barely

<sup>5</sup>This preserves the original unit of measurement. The difference or ratio of the variances  $s_{gj1}^2$  and  $s_{gj2}^2$  would yield similar results.

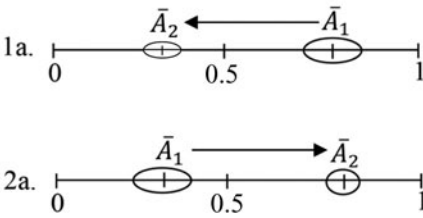
<sup>6</sup>The redundancy is that  $V_{gj}^b = 1 - H_{gj}^b$ , except in the rare cases in which  $H_{gj} = 0$ . The same, in terms defined below, can be said of  $M_{gj}^b$  and  $O_{gj}^b$  vis-à-vis  $P_{gj}^b$  and  $D_{gj}^b$ .

<sup>7</sup> $P_{gj}$  is undefined for  $\bar{A}_{gj1} = \text{exactly } 0.5$ , in which case the group’s mean attitude cannot move toward or away from the nearer extreme, since neither extreme is nearer than the other.

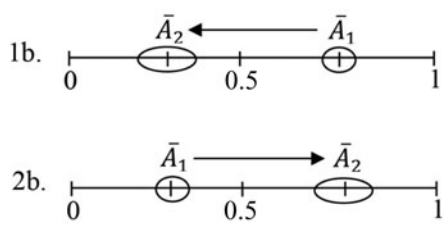
<sup>8</sup>Moderation, therefore, need not mean decreasing extremity. A group that polarizes becomes more extreme, but so may one that moderates (as when  $\bar{A}_{gj1} = 0.6$ , and  $\bar{A}_{gj2} = 0.1$ ).

**A. Homogenization/variegation**

$H > 0$  (Homogenization)

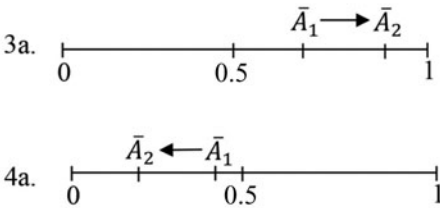


$H < 0$  (Variegation)

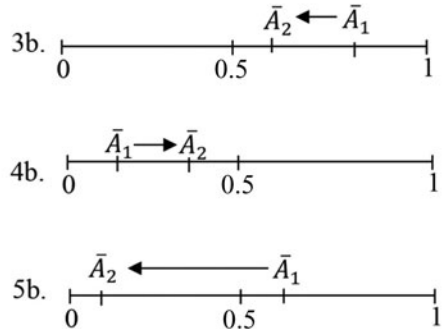


**B. Polarization/moderation**

$P > 0$  (Polarization)

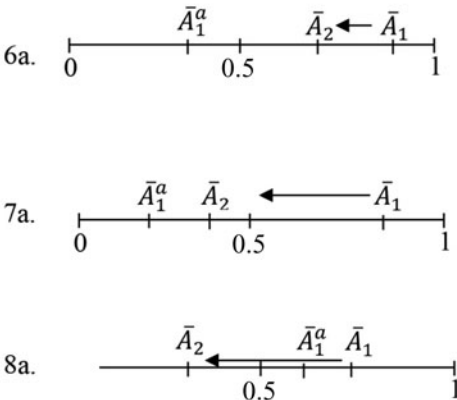


$P < 0$  (Moderation)



**C. Domination/opposition**

$D > 0$  (Domination)



$D < 0$  (Opposition)

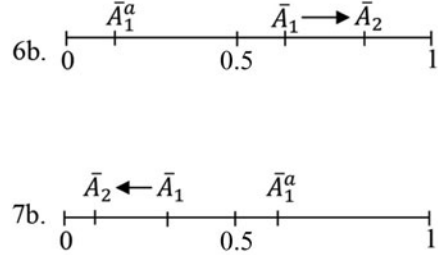


Figure 1. Illustrating the definitions.

under 0.5) when the mean is either just barely above 0.5 before deliberation and exactly 1 after or just barely below 0.5 before deliberation and exactly 0 after. It is at its most negative (just barely above  $-1$ ) when the mean is just fractionally toward the midpoint from the nearer pole (just

barely below 1 or above 0) before deliberating and at the opposite pole (0 or 1) after.<sup>9</sup> The binary version is  $P_{gj}^b = 1$  if  $P_{gj} > 0$  and  $= 0$  if  $P_{gj} \leq 0$ . Its complement, for moderation, is  $M_{gj}^b = 1$  if  $P_{gj} < 0$  and  $= 0$  if  $P_{gj} \geq 0$ .

The *domination* of the  $g$ th group's attitudes on the  $j$ th issue (with respect to a given dimension of advantage) is:

$$D_{gj} = (\bar{A}_{gj2} - \bar{A}_{gj1})R_{gj1}, \tag{3}$$

where  $R_{gj1}$  indicates the ordinal relation between  $\bar{A}_{gj1}$  and  $\bar{A}_{gj1}^a$ :  $R_{gj1} = 1$  for  $\bar{A}_{gj1}^a > \bar{A}_{gj1}$  and  $= -1$  for  $\bar{A}_{gj1}^a < \bar{A}_{gj1}$ . Thus  $D_{gj} > 0$  for domination,  $< 0$  for opposition and  $= 0$  for neither (no mean attitude change).<sup>10</sup> In Figure 1, C6a, C7a, and C8a show  $\bar{A}_{gj1}$  moving toward or beyond  $\bar{A}_{gj1}^a$  (domination), while C6b and C7b show it moving in the opposite direction, away from  $\bar{A}_{gj1}^a$  (opposition).  $D_{gj}$  is at its most positive (just barely  $< 1$ ) when the disadvantaged start at 1 or 0, the advantaged start just barely toward the midpoint from that (as, therefore, does the whole group), and everyone, whether advantaged or disadvantaged, moves all the way to the opposite pole (0 or 1), not only toward but as far as possible beyond  $\bar{A}_{gj1}^a$ . It is at its most negative (just barely  $> -1$ ) when the *advantaged* start at 1 or 0, the *disadvantaged* start just barely toward the midpoint from that, as therefore does the whole group), and everyone moves all the way to the opposite pole (0 or 1), as far as possible away from  $\bar{A}_{gj1}^a$ .<sup>11</sup> The binary version is  $D_{gj}^b = 1$  if  $D_{gj} > 0$  and  $= 0$  if  $D_{gj} \leq 0$ . Its complement, for opposition, is  $O_{gj}^b = 1$  if  $D_{gj} < 0$  and  $= 0$  if  $D_{gj} \geq 0$ .<sup>12</sup>

We shall examine three dimensions of advantage – gender, education, and income – both individually and all three combined.<sup>13</sup> For gender, a matter simply of sociodemographic group membership, the threshold of advantage (maleness) is relatively clear. For education and income, matters of having more or less of a numerical or ordinal property, it is less clear. But division at each DP's sample median makes sense for several reasons. First, the sample median varies from sample to sample, tacitly recognizing that what is highly educated or high income varies by time and place. Social advantage is relative. Second, the sample median, unlike the small-group median, lets the proportions of advantaged versus disadvantaged vary from group to group. Third, the median, compared to other sample-dependent cut-points, minimizes the proportion of small groups for which the number of either disadvantaged or advantaged members scrapes zero. Fourth, the median is a good guess when we do not know where to draw the line. If the actual proportion of the sample that is disadvantaged has a symmetric (Bayesian) probability distribution centered at 0.5 (the uniform distribution being a special case), the minimum mean-squared-error guess is 0.5, corresponding to division at the median.

<sup>9</sup> $P_{gj}$  can be larger in magnitude at its most negative than at its most positive because its positive values gauge movement on the 'short side of the field'.  $\bar{A}_{gj1}$  cannot move as far toward the nearer pole. This asymmetry does not tilt the results toward moderation, however. It is no easier to get a  $P_{gj}$  of  $-0.3$  than one of  $+0.3$ , and we virtually never actually see  $P_{gj} < -0.5$ .

<sup>10</sup> $D_{gj}$  is undefined for  $\bar{A}_{gj1}$  exactly =  $\bar{A}_{gj1}^a$  (implying that both also exactly =  $\bar{A}_{gj1}^a$ ), in which case  $\bar{A}_{gj1}$  cannot move toward  $\bar{A}_{gj1}^a$ . It is already there.

<sup>11</sup>Domination, therefore, need not mean decreasing distance from  $\bar{A}_{gj1}^a$ . A group exhibiting opposition increases its distance from  $\bar{A}_{gj1}^a$ , but so may one exhibiting domination (as when  $\bar{A}_{gj1}^a = 0.6$ ,  $\bar{A}_{gj1} = 0.7$ , and  $\bar{A}_{gj2} = 0.3$ ).

<sup>12</sup>As these formalizations make clear,  $H_{gj}$ ,  $P_{gj}$ , and  $D_{gj}$  (and  $H_{gj}^b$ ,  $P_{gj}^b$ , and  $D_{gj}^b$ ) are conceptually distinct.  $P_{gj}$  and  $D_{gj}$  are both mean attitude change, though with different points of reference (the nearer extreme versus  $\bar{A}_{gj1}^a$ ). In magnitude, both are  $|\bar{A}_{gj2} - \bar{A}_{gj1}|$ , but their signs are as apt to be opposite as the same.  $H_{gj}$  is a change not in the mean attitude but in the within-group standard deviation. A positive correlation between  $H_{gj}$  and  $P_{gj}$  does seem likely (and can be found in Sunstein's (2002) and Sunstein and Hastie's (2014) analyses, of data from other deliberative designs), but not on account of any built-in overlap. Rather,  $H_{gj}$  may causally affect  $P_{gj}$ , or the two may be 'spuriously' correlated, each resting on the same or correlated other variables. In our DP data (described below), the  $H_{gj}$ - $P_{gj}$  correlation is only a modest 0.358, and the  $H_{gj}$ - $D_{gj}$  and  $D_{gj}$ - $P_{gj}$  correlations range from  $-0.062$  to  $0.105$ , averaging only 0.012.

<sup>13</sup>In this last case,  $\bar{A}_{gj1}^a$  is the initial mean attitude of those who are male, better-educated, and higher-income.



## Theory, Expectations and Inferences

In broad strokes, our central proposition is that homogenization, polarization and domination rest (and therefore shed light) on the deliberative quality of the discussion. It will help in developing the why's and how's to note that the population means of  $H_{gj}$ ,  $P_{gj}$ , and  $D_{gj}$  (averaging across all possible group-issue pairs, a sense of 'population' about which we say a bit more below) are  $E(H_{gj})$ ,  $E(P_{gj})$ , and  $E(D_{gj})$ , where  $E(\cdot)$  denotes mathematical expectation. Positive values indicate the extent to which, on average, homogenization exceeds variegation, polarization exceeds moderation, and domination exceeds opposition; negative values, the reverse. Similarly, the relative frequencies of homogenization, polarization, and domination are  $E(H_{gj}^b)$ ,  $E(P_{gj}^b)$ , and  $E(D_{gj}^b)$ , and those of variegation, moderation, and opposition  $E(V_{gj}^b) = 1 - E(H_{gj}^b)$ ,  $E(M_{gj}^b) = 1 - E(P_{gj}^b)$ , and  $E(O_{gj}^b) = 1 - E(D_{gj}^b)$ . It will also help, at points, to take  $|\bar{A}_{gjt} - \bar{A}_{gj}^*|$  and  $|s_{gjt} - s_{gj}^*|$  – the distances between a group's sample-mean attitude ( $\bar{A}_{gjt}$ ) and its sample-mean full-consideration attitude ( $\bar{A}_{gj}^*$ ) and between the sample standard deviations of its members'  $j$ th-issue attitudes ( $s_{gjt}$ ) and of their full-consideration attitudes ( $s_{gj}^*$ ) – as simple, tractable reflections of group-level 'appropriateness'. The smaller these distances, the more appropriate the group's attitudes.

## Social Dynamics versus Weighing the Merits

We see two broad mechanisms by which a discussion may change policy attitudes.

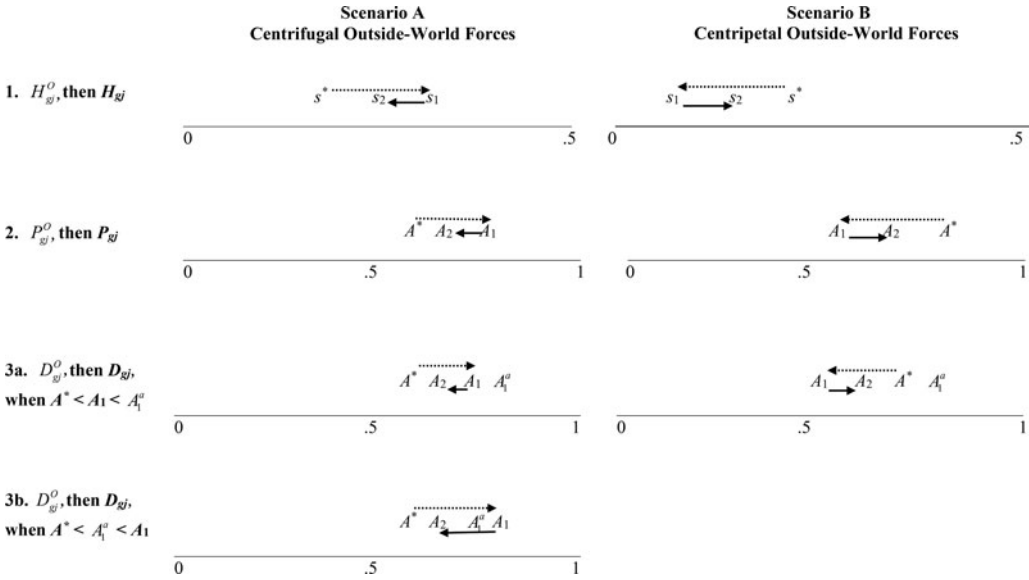
### Social dynamics (SD)

The first lies in the discussion's social interactions, the relevant features of which we shall call its social dynamics (SD). People commonly seek approval and sidestep disagreement. That should shrink the initial within-group standard deviation  $s_{gj1}$  (Cialdini and Goldstein 2004; Gerber et al. 2012; Huckfeldt and Sprague 1995; Huckfeldt, Johnson and Sprague 2004; Mutz 2006; Suhay 2015; Sunstein 2002; Sunstein 2009; Sunstein and Hastie 2014) and pull the initial group mean attitude  $\bar{A}_{gj1}$  toward the nearer extreme (Suhay 2015; Sunstein 2002; Sunstein 2009; Sunstein and Hastie 2014; Wojcieszak 2011; Zuber, Crott and Werner 1992). In addition, some participants, often concentrated among the socially disadvantaged, will normally be less articulate, less assertive, or less heeded than others. That should move  $\bar{A}_{gj1}$  toward  $\bar{A}_{gj1}^a$  (Fraser 1993; Karpowitz and Mendelberg 2014; Karpowitz, Mendelberg and Shaker 2012; Sanders 1997; Young 2000). Hence SD should produce homogenization, polarization, and domination – not always strongly, nor in every instance, but on average and more often than not. More formally, we should expect  $E(H_{gj})$ ,  $E(P_{gj})$ ,  $E(D_{gj})$  all  $\gg 0$ , and  $E(H_{gj}^b)$ ,  $E(P_{gj}^b)$ , and  $E(D_{gj}^b)$  all  $\gg 0.5$  and, *ipso facto*,  $\gg E(V_{gj}^b)$ ,  $E(M_{gj}^b)$ , and  $E(O_{gj}^b)$ .

### Weighing the merits (WM)

The second mechanism is the participants' open-minded, even-handed, and earnest weighing of the merits (the arguments and evidence), as they see them – the deliberation in the discussion, call it WM. This where Habermas's (1990, 1996) 'unforced force of the better argument' resides. In WM, participants can be expected to absorb nontrivial quantities of new information, higher-than-everyday proportions of which are accurate and counterattitudinal, thus increasing  $C_{ij1}$ ,  $F_{ij1}$ , and  $B_{ij1}$  (as the results in Barabas 2004; Hansen 2004; Luskin, Fishkin, and Jowell 2002; and Farrar et al. 2010 suggest). That, in turn, should allow them to see more clearly how given policies may serve or thwart their values and interests (which they may also come to see more clearly), thus moving  $A_{ij1}$  closer to  $A_{ij}^{*14}$  and, at the group level, reducing the distances

<sup>14</sup>With some exceptions. When  $A_{ij1}$  is already close to  $A_{ij}^*$  ( $\alpha_{ij1}$  close to 1), WM may leave the attitude unchanged ( $A_{ij2} = A_{ij1}$ ,  $\alpha_{ij2} = \alpha_{ij1}$ ), just better grounded.



**Figure 2.** Homogenization, Polarization, and Domination to Be Expected in the Outside World, then from Weighing the Merits. Note:  $A_1, A_2, A^*, A_1^a, s_1, s_2,$  and  $s^*$  are short for the text's  $\bar{A}_{gj1}, \bar{A}_{gj2}, \bar{A}_{gj}^*, s_{gj1}, s_{gj2},$  and  $s_{gj}^*$ . We assume, without loss of generality, that  $\bar{A}_{gj1} > 0.5$ .

$|\bar{A}_{gjt} - \bar{A}_{gj}^*|$  and  $|s_{gjt} - s_{gj}^*|$ . There is no obvious reason to expect these changes to constitute homogenization, polarization, or domination (or their opposites), although ...

**A Closer Look at WM's Effects**

We can actually reason out some rough expectations about WM-induced homogenization/variegation, polarization/moderation, and domination/opposition by taking account of the pre-existing homogenization/variegation, polarization/moderation, and domination/opposition (call them  $H_{gj}^o, P_{gj}^o,$  and  $D_{gj}^o$ ) of the outside-world attitudes with which the discussion begins. These latter can most sensibly be defined as differences between the initial  $s_{gj1}$  and  $\bar{A}_{gj1}$  and the full-consideration  $s_{gj}^*$  and  $\bar{A}_{gj}^*$  (in contrast to  $H_{gj}, P_{gj},$  and  $D_{gj}$ , which are differences between the initial  $s_{gj1}$  and  $\bar{A}_{gj1}$  and the post-discussion  $s_{gj2}$  and  $\bar{A}_{gj2}$ ).<sup>15</sup>

Figure 2 illustrates the pre-existing  $H_{gj}^o, P_{gj}^o,$  and  $D_{gj}^o$  alongside the corresponding WM-induced  $H_{gj}, P_{gj},$  and  $D_{gj}$ . To avoid redundant mirror-image cases, we assume, without loss of generality, that  $\bar{A}_{gj1} > 0.5$ , making 1 the nearer extreme. The lower, solid arrows depict the WM-induced attitude changes, moving  $s_{gjt}$  and  $\bar{A}_{gjt}$  toward  $s_{gj}^*$  and  $\bar{A}_{gj}^*$ , and the upper, dashed ones the prior effects of outside-world forces, pulling the initial  $s_{gj1}$  and  $\bar{A}_{gj1}$  above the full-consideration  $s_{gj}^*$  and  $\bar{A}_{gj}^*$  in Scenario A and below them in Scenario B. Scenario A consists of pre-existing variegation ( $s_{gj1} > s_{gj}^*$ ), polarization ( $\bar{A}_{gj}^* < \bar{A}_{gj1}$ ), and domination ( $\bar{A}_{gj}^* < \bar{A}_{gj1} < \bar{A}_{gj1}^a$  or  $\bar{A}_{gj}^* < \bar{A}_{gj1}^a < \bar{A}_{gj1}$ , given  $\bar{A}_{gj}^* < \bar{A}_{gj1}$ ), Scenario B of pre-existing homogenization ( $s_{gj1} < s_{gj}^*$ ), moderation ( $\bar{A}_{gj1} < \bar{A}_{gj}^*$ ), and opposition ( $\bar{A}_{gj1} < \bar{A}_{gj}^* < \bar{A}_{gj1}^a$ , given  $\bar{A}_{gj1} < \bar{A}_{gj}^*$ ).<sup>16</sup> We shall return to the two possibilities for pre-existing domination in Figure 2's rows 3a and 3b.

<sup>15</sup>More precisely,  $H_{gj}^o = s_{gj}^* - s_{gj1}, P_{gj}^o = (\bar{A}_{gj1} - \bar{A}_{gj}^*)S_{gj}^*$ , and  $D_{gj}^o = (\bar{A}_{gj1} - \bar{A}_{gj}^*)R_{gj}^*$ , where  $S_{gj}^* = 1$  for  $\bar{A}_{gj}^* > 0.5$  and  $= -1$  for  $\bar{A}_{gj}^* < 0.5$ , and  $R_{gj}^* = 1$  for  $\bar{A}_{gj1} > \bar{A}_{gj}^*$  and  $= -1$  for  $\bar{A}_{gj1} < \bar{A}_{gj}^*$ . Thus  $H_{gj}^o, P_{gj}^o,$  and  $D_{gj}^o$  are pre-existing homogenization, polarization, and domination when  $> 0$  and pre-existing variegation, moderation, and opposition when  $< 0$ .

<sup>16</sup>The inequalities are strict because we exclude  $s_{gj1} = s_{gj}^*, \bar{A}_{gj1} = 0.5,$  and  $\bar{A}_{gj1} = \bar{A}_{gj1}^a$  as both vanishingly rare and leaving the movements of  $s_{gj1}$  toward  $s_{gj}^*$  and  $\bar{A}_{gj1}$  toward 0.5 and  $\bar{A}_{gj1}^a$  undefined (in these cases,  $s_{gj1}$  cannot get any closer to  $s_{gj}^*$ , nor  $\bar{A}_{gj1}$  any closer to 0.5 or  $\bar{A}_{gj1}^a$ ).



Scenarios A and B are alternative legacies of the forces shaping outside-world attitudes. Some of those forces – notably, communications siloing (residential and other sorting, homophily, selective media consumption) and social inequalities – are *centrifugal*, pulling the initial attitudes away from 0.5, toward both the nearer extreme and the mean attitude of the advantaged ( $\bar{A}_{gj1}^a$ ), and thus also (since the nearer extreme is 0 for some group members but 1 for others) increasing their variance. Other forces – principally, inattention, ignorance, and irreflection – are *centripetal*, keeping the initial attitudes closer to 0.5 and thus also restraining their variance. The centrifugal forces make for Scenario A, the centripetal ones for Scenario B.<sup>17</sup> The world, of course, involves a mix of both kinds of forces. The centrifugal ones are presently much noted (and understandably bemoaned). But, behind the scenes, the centripetal ones are at the same time etiologating the attitudes of the less politically engaged – and, on low-salience issues, many of those of the more politically engaged as well.

What Figure 2 makes clear is that WM can be expected to ‘correct’ what the outside world has done – producing homogenization, moderation, and opposition that reduce or reverse the pre-existing variegation, polarization, and domination in Scenario A and variegation, polarization, and domination that reduce or reverse the pre-existing homogenization, moderation, and opposition in Scenario B. As drawn, the arrows (shorter from  $s_{gj1}$  to  $s_{gj2}$  and from  $\bar{A}_{gj1}$  to  $\bar{A}_{gj2}$  than from  $s_{gj}^*$  to  $s_{gj1}$  and from  $\bar{A}_{gj}^*$  to  $\bar{A}_{gj1}$ ) depict reductions, which seem likelier than reversals (its being hard for a few days of discussion to completely negate the accumulated effects of a lifetime of prior circumstances and experiences). The reduction or reversal may be slightly smaller for Scenario A’s pre-existing domination, which can yield WM-induced opposition in 3a but WM-induced domination in 3b, than for Scenario A’s pre-existing variegation or polarization (or anything in Scenario B). But the centrifugal forces pulling  $\bar{A}_{gj1}$  above  $\bar{A}_{gj}^*$  should also pull the better-educated and better-off, who tend to be more exposed to the information and messaging involved, still further above it ( $\bar{A}_{gj1} < \bar{A}_{gj1}^a$ ), making 3a much more common than 3b.

The lesson for the frequencies of WM-induced homogenization/variegation, polarization/moderation, and domination/opposition is that they ultimately depend on the balance between centrifugal and centripetal forces. Absent much reason to see either as greatly stronger than the other, the safest guess, and our expectation, is that they are about equally strong. In this case, WM should produce homogenization, polarization, and domination about half the time and variegation, moderation, and opposition about half the time:  $E(H_{gj}^b) \cong E(V_{gj}^b) \cong E(P_{gj}^b) \cong E(M_{gj}^b) \cong E(D_{gj}^b) \cong E(O_{gj}^b) \cong 0.5$ . To the extent that these proportions depart from 0.5, however, we might expect to see slightly more homogenization than variegation but slightly less polarization and domination than moderation and opposition. These are increasingly tribal days (Achen and Bartels 2016), in which the balance of outside-world forces may be tipping toward the centrifugal. Probably not too grossly, to be sure. Inattention, ignorance, and irreflection remain forever widespread and potent, and should keep  $E(H_{gj}^b)$ ,  $E(P_{gj}^b)$ , and  $E(D_{gj}^b)$  (and  $E(V_{gj}^b)$ ,  $E(M_{gj}^b)$ , and  $E(O_{gj}^b)$ ) from departing too much from 0.5.

### Theoretical Asides

Two side notes, important in different ways to our post-analysis reflections below, need sounding.

#### Motivated reasoning

The first is that both WM and SD entail varieties of ‘motivated reasoning’, a term often simplistically reduced to ways of ignoring, discounting, or reasoning around counterattitudinal information. This too-narrow sense is apt enough when the motivations are strictly or mainly ‘defensive’. In that case, discussion should tend to produce little attitude change. But defensive motivations

<sup>17</sup>The centrifugal-centripetal distinction resembles Leeper’s (2014) between ‘slanted information environments’ and ‘low-importance attitudes’.

may be less pervasive than previously thought (Arceneaux and Vander Wielen 2017; Druckman 2012; Druckman and McGrath 2019; Hahn and Harris 2014; Hart *et al.* 2009; Leeper and Slothuus 2014; Már and Gastil 2020), and are hardly the only motivations in play (Kunda 1990), not even the only ‘directional’ ones (Hart *et al.* 2009). In discussions, social approval motivations (to please or favorably impress others) are important to *SD*, and accuracy motivations (to ‘process information in an objective, open-minded fashion ...’, Hart *et al.* 2009, 558) important to *WM*. Both can be expected to change attitudes – the former in directions tending to yield homogenization, polarization, and domination, the latter in the direction of full-consideration attitudes.

### Contextual factors

The second note concerns the conditions under which the discussion occurs. A discussion involving a serious weighing of the merits is deliberative, but the deliberation may be more or less ‘effective’ – more or less helpful to its participants in considering what their attitudes should be – depending on who is ‘in the room’ and (not unrelatedly) the information readily available to them. The more demographically and attitudinally heterogeneous the discussants, and the more plentiful, balanced, and accurate the information, the harder it is for the discussion and its participants to misconstrue, slight, or ignore relevant information and arguments (see Mercier and Landemore 2012; Tuller *et al.* 2015). These variables, too, are part of what separates successful deliberative designs from everyday discussions, in which homophily and sorting make the discussants relatively homogeneous, and the information is generally confined to whatever the discussants bring with them, which is, for most of them, in most discussions, sparse, imbalanced and/or inaccurate.<sup>18</sup>

### In Sum

A discussion’s effects on homogenization, polarization, and domination should depend on how deliberative it is – on how much it revolves around *WM*. Everyday discussions, involving much *SD* and little *WM*, can be expected to yield decidedly more homogenization than variegation, polarization than moderation, and domination than opposition. The deliberative discussions spawned by successful deliberative designs should not. If anything, they may produce some slight homogenization, but also, if so, some slight *moderation* and *opposition*.

These widely different expectations allow the observed distributions of  $H_{gj}$ ,  $P_{gj}$ , and  $D_{gj}$  and  $H_{gj}^b$ ,  $P_{gj}^b$ , and  $D_{gj}^b$  to form a rough diagnostic. If the sample means  $\bar{H}$ ,  $\bar{P}$ ,  $\bar{D}$  (estimating  $E(H_{gj})$ ,  $E(P_{gj})$ , and  $E(D_{gj})$ ) are well above 0, or the sample frequencies  $\bar{H}^b$ ,  $\bar{P}^b$ , and  $\bar{D}^b$  (estimating  $E(H_{gj}^b)$ ,  $E(P_{gj}^b)$ , and  $E(D_{gj}^b)$ ) well above 0.5, the discussion is probably not very deliberative, involving little beyond *SD*. If instead  $\bar{H}$ ,  $\bar{P}$ , and  $\bar{D}$  are all close to 0, and  $\bar{H}^b$ ,  $\bar{P}^b$ , and  $\bar{D}^b$  all close to 0.5, the discussion is probably quite deliberative, involving a healthy dose of *WM*.

### Data

We take our data from Deliberative Polling, a well-known deliberative design (described, for example, in Luskin, Fishkin and Jowell 2002). Its signal features include randomly sampled participants, randomly assigned to small groups; honoraria to help recruit hard-to-get participants, including those unenticed by the prospect of discussing the policy issue; balanced, factually accurate briefing materials provided in advance; lightly moderated small-group discussions alternating with plenary question-and-answer sessions with panels of policy experts; and anonymous questionnaires to gauge policy attitudes and other relevant variables both before and after deliberation. The twenty-one DPs supplying our data are summarized in Table 1. Five were in Britain, eleven in the United States, two in the EU, and one each in China, Australia, and Bulgaria. Sixteen were

<sup>18</sup>Opposing imbalances in individual-level information could cancel out, but since the discussants tend to be homogenous, the group-level information tends to be imbalanced as well.

Table 1. DPs analyzed

Broad Topic(s)	<i>n</i>	Country	Region/Town	Year	Mode	Policy Indices	Groups
Britain's Role in the EU	238	UK		1995	F2F	4	16
National Health System	230	UK		1998	F2F	11	15
British Monarchy	258	UK		1996	F2F	4	15
UK General Election	275	UK		1997	F2F	4	15
Making Australia a Republic	347	Australia		1999	F2F	5	24
Crime	299	UK		1994	F2F	5	20
EU Expansion, Pension Reform, Foreign Policy	344	EU <sup>a</sup>		2007	F2F	7	18
Meeting future electricity needs	216	US	CP&L Service Area	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	EU <sup>a</sup>		2009	F2F	2	25
Future of Airport, Revenue-Sharing	132	US	New Haven, CT	2004	F2F	3	16
Foreign policy	340	US		2003	F2F	9	24
Foreign Policy	245	US		2003	OL	9	15
US General Election	246	US		2004	OL	6	15
US Presidential Primaries	434	US		2004	OL	3	16
Housing Policy	239	US	San Mateo, CA	2008	F2F	7	26
Health Care, Education	454	US		2005	OL	11	30
Meeting future electricity needs	230	US	WTU Service Area	1996	F2F	7	14
Crime, the family, foreign policy	466	US		1996	OL	9	30
Meeting future electricity needs	232	US	SEP Service area	1996	F2F	7	14
Total	6,084					134	397

Note: WTU, West Texas Utilities; CP&L, Central Power & Light; SEP, Southwestern Electric Power.

<sup>a</sup>All (then twenty-seven) Member-States.

face to face, five online. The topics spanned policy issues from foreign policy to health care. In all, the data encompass 372 small groups (containing, all told, 5,736 participants), 139 policy issues (counting each policy attitude index as tapping an at least somewhat different issue), and 2,601 group-issue pairs.<sup>19</sup> Appendix A describes the indices and their ingredients in greater detail.

In the fullest accounting, we are sampling time-indexed person-populations (for example, Great Britain in 1994, which is in our sample, or Paraguay in 2011, which is not), then both the individuals within those time-indexed person-populations and the policy issues facing them (for example, crime in Bulgaria in 2007, which is in our sample, and climate change in the United States in 2009, which is not). The samples of individuals are almost always random draws. The samples of time-indexed person-populations and policy issues are not. Yet we hope they are large and varied enough to afford 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 and conducted face-to-face,  $\bar{H}$ ,  $\bar{P}$ , and  $\bar{D}$  are only trivially different for the group-issue pairs from other countries and in online mode, as can be seen in Appendix B.

## Results

So how much homogenization, polarization, and domination does there appear to be? The short answer is, very little. Figure 3 shows the distributions of the group-issue-level  $H_{gj}$ ,  $P_{gj}$ , and  $D_{gj}$  across group-issue pairs. All are packed tightly and symmetrically around near-zero means. Some group-issue pairs exhibit homogenization, some variegation; some exhibit polarization,

<sup>19</sup>Not  $372 \times 139 = 51,708$ , since each index is confined to just one DP and its small groups.

some moderation; some exhibit domination, some opposition. But often and on average very little of any.

**Table 2** homes in on the means ( $\bar{H}$ ,  $\bar{P}$ ,  $\bar{D}$ ) and relative frequencies ( $\bar{H}^b$ ,  $\bar{P}^b$ , and  $\bar{D}^b$ ). The top row shows  $\bar{H}$ ,  $\bar{P}$ ,  $\bar{D}$ ,  $\bar{H}^b$ ,  $\bar{P}^b$ , and  $\bar{D}^b$ , the lower rows the Huber-White estimates of the standard errors (White 1980),<sup>20</sup> and two-tailed  $p$ -values for the null hypotheses that  $E(H_{gj}) = E(P_{gj}) = E(D_{gj}) = 0$  and  $E(H_{gj}^b) = E(V_{gj}^b) = E(P_{gj}^b) = E(M_{gj}^b) = E(D_{gj}^b) = E(O_{gj}^b) = 0.5$  – that in the population of all possible group-issue pairs the mean levels of homogenization, polarization, and domination are 0 and that each occurs only as often as its opposite (as would be expected from *WM*, assuming centrifugal and centripetal forces to be equally strong).<sup>21</sup> The alternative hypotheses are that  $E(H_{gj})$ ,  $E(P_{gj})$ ,  $E(D_{gj})$  all  $> 0$  and that  $E(H_{gj}^b) > 0.5 > E(V_{gj}^b)$ ,  $E(P_{gj}^b) > 0.5 > E(M_{gj}^b)$ , and  $E(D_{gj}^b) > 0.5 > E(O_{gj}^b)$  – that the mean levels of homogenization, polarization, and domination are all positive and that each occurs more often than its opposite.

In their stronger versions ( $\bar{H}$ ,  $\bar{P}$ , and  $\bar{D} \gg 0$  and  $\bar{H}^b$ ,  $\bar{P}^b$ , and  $\bar{D}^b \gg 0.5$ ), these alternatives are what would be expected from a discussion involving mostly *SD*. But **Table 2** shows nothing of the sort. True, six of the table's ten estimates are statistically significant ( $p < 0.05$ ). In these cases, we can be quite sure that, in the population from which we are sampling, there is *some* homogenization or variegation (depending on the signs of  $\bar{H}$  and  $\bar{H}^b - 0.5$ ), *some* polarization or moderation (depending on the signs of  $\bar{P}$  and  $\bar{P}^b - 0.5$ ), and *some* domination or opposition (depending on the signs of  $\bar{D}$  and  $\bar{D}^b - 0.5$ ). But how much? At a glance,  $\bar{H}^b$ ,  $\bar{P}^b$ , and  $\bar{D}^b$  are close to 0.5,  $\bar{H}$ ,  $\bar{P}$ , and  $\bar{D}$  close to 0.

A closer look reinforces that impression. Take homogenization.  $\bar{H}^b = 0.595$ , distinctly above but still quite close to 0.5 (less than 20 percent of the way to 1). This is far from 'routine'. On average, moreover, it is weak. To see this, imagine a group of twenty participants whose initial attitudes are as follows: four at 0.6 and two each at every other integer multiple of 0.1 from 0.2 to 1. This initial distribution has a mean of 0.6 and a middling standard deviation of 0.245, close to halfway between  $s_{git}$ 's maximum of 0.5 and minimum of 0. Now let one of the participants initially at 0.2 move to 0.3 and one of those initially at 1 move to 0.9 – in each case, 0.1 closer to the mean. This is not much homogenization: the distribution is almost completely unaltered. The mean remains 0.6, while the standard deviation shrinks from 0.245 to 0.230. This unimposing scenario thus yields  $H_{gj} = 0.015$ . The observed  $\bar{H} = 0.013$  is still lower.

Next take polarization.  $\bar{P}^b = 0.454$ , meaning that slightly *fewer* than half the group-issue pairs polarize (more than half moderate), and  $\bar{P} = -0.022$ , meaning that, on average, their mean attitudes move slightly *toward*, not away from, the midpoint, likewise representing moderation. To contextualize this  $\bar{P}$ , take again a group of twenty with an initial mean attitude  $\bar{A}_{gj1}$  of 0.6. If just five of twenty members decrease their scores, from whatever starting points, by just 0.1 apiece (a scenario involving a bit more attitude change than the one sketched just above, but still not much),  $\bar{P}$  is only  $-0.025$ . The observed  $\bar{P}$  is still smaller ( $-0.022$ ).

Finally, domination. Across our four dimensions of advantage,  $\bar{D}^b$  runs only from 0.447 to 0.485. No matter what the dimension, *fewer* than 50 percent of the group-issue pairs move toward the initial mean attitude of the advantaged, meaning that more than 50 percent move *away* from it. This is (weak) opposition, not domination. The  $\bar{D}$ 's tell much the same tale:  $\bar{D} = 0.008$  for gender,  $= -0.013$  for education,  $< 0.0005$  for income, and  $= -0.015$  for threefold advantage. Again take a group of twenty members. Let  $\bar{A}_{gj1} = 0.6$  and  $\bar{A}_{gj1}^a = 0.8$ . If just three of the twenty participants decrease their scores by just 0.1 apiece,  $D_{gj} = -0.015$ . If just two do so,  $D_{gj} = -0.10$ . These scenarios involve scant attitude change, but their  $D_{gj}$ 's bracket the negative  $\bar{D}$ 's for education and threefold

<sup>20</sup>A given group may 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>21</sup>The  $p$ -value is a stretch for  $H_{gp}$ , a difference of standard deviations, rather than of means.

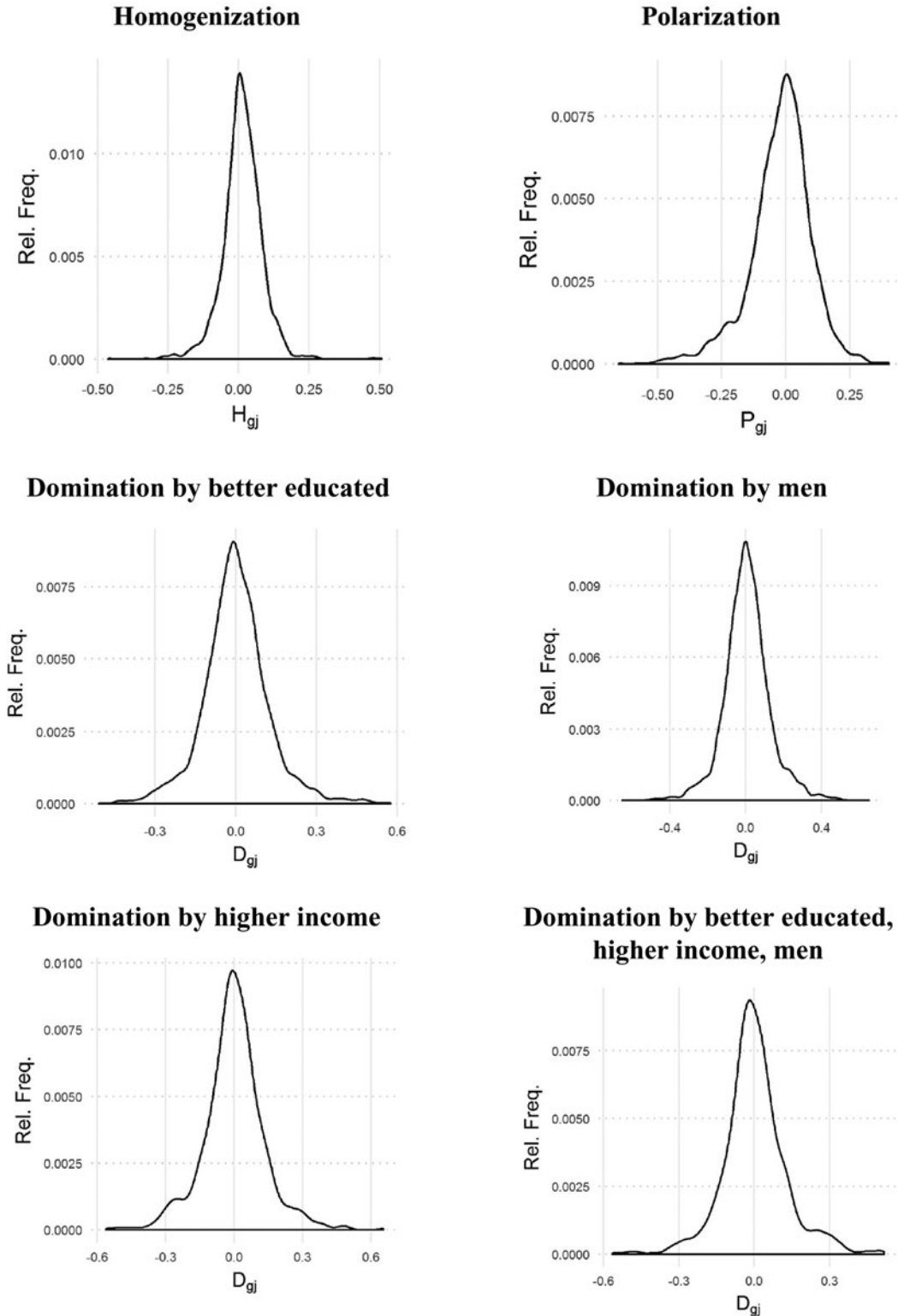


Figure 3. Distributions of group-issue pairs on  $H_{gj}$ ,  $P_{gj}$  and  $D_{gj}$ .

**Table 2.** Homogenization, polarization and domination: occurrence and (signed) magnitude

	Homogenization		Polarization		Domination							
					By Gender		By Education		By Income		By All 3	
	$H_{gj}$	$H_{gj}^b$	$P_{gj}$	$P_{gj}^b$	$D_{gj}$	$D_{gj}^b$	$D_{gj}$	$D_{gj}^b$	$D_{gj}$	$D_{gj}^b$	$D_{gj}$	$D_{gj}^b$
Mean	0.013	0.595	-0.022	0.454	0.008	0.464	-0.013	0.447	0.000	0.485	-0.015	0.466
s.e	0.003	0.017	0.010	0.027	0.004	0.011	0.005	0.012	0.005	0.017	0.007	0.020
P	0.000	0.000	0.032	0.088	0.031	0.002	0.008	0.000	0.951	0.425	0.029	0.102

Note: in the ‘Mean’ row, the entries for  $H_{gj}$ ,  $P_{gj}$ , and  $D_{gj}$  are the means of those variables ( $\bar{H}$ ,  $\bar{P}$ , and  $\bar{D}$ ). Those for  $H_{gj}^b$ ,  $P_{gj}^b$ , and  $D_{gj}^b$  ( $\bar{H}^b$ ,  $\bar{P}^b$ , and  $\bar{D}^b$ ) are the relative frequencies with which  $H_{gj} > 0$ ,  $P_{gj} > 0$ , and  $D_{gj} > 0$ .

advantage. The positive  $\bar{D}$ 's for gender and income are still smaller. There is more opposition than domination, but not much of either.

The principal lesson is clear. The homogenization, polarization, and domination here are much too modest to suggest attitude changes stemming largely from *SD* but are consistent with attitude changes stemming largely from *WM*. A further suggestion, too faint and uncertain to be a ‘lesson’, lies in the signs of  $\bar{H}$ ,  $\bar{P}$ , and  $\bar{D}$  and of  $\bar{H}^b$ 's,  $\bar{P}^b$ 's, and  $\bar{D}^b$ 's departures from 0.5. These show some slight homogenization, moderation, and opposition, a combination suggesting that the deliberation may be redressing outside-world variegation, polarization, and domination.

**Elaborations and Reflections**

By way of follow-up, it may help to say a bit more about what ‘domination’, the most polyglot of our terms, does and does not mean; to elaborate on and probe our findings regarding it; to consider motivated reasoning’s implications for our rough diagnostic; and to suggest some of the likeliest reasons for which our results diverge from some other studies’.

***D<sub>gj</sub>* and the Meaning of Domination**

‘Domination’, here, is simply attitude change – specifically, the group’s mean attitude change toward the initial mean attitudes of its advantaged members, distilled in  $D_{gj}$ . It is not itself a matter of dialogic inequalities or other asymmetries in the discussion’s social interactions, although they presumably affect it. Equivalently,  $D_{gj}$  is also the weighted mean of  $\bar{A}_{gj1}^d$ ’s movement toward  $\bar{A}_{gj1}^a$  and  $\bar{A}_{gj1}^a$ ’s movement toward  $\bar{A}_{gj1}^d$ :

$$D_{gj} = d_g M_{gj}^d + (1 - d_g) M_{gj}^a, \tag{4}$$

where  $M_{gj}^d = (\bar{A}_{gj2}^d - \bar{A}_{gj1}^d)$ ,  $M_{gj}^a = (\bar{A}_{gj2}^a - \bar{A}_{gj1}^a)$ , and  $d_g \neq 0$  is the proportion of the *g*th group who are disadvantaged. Note that both  $M_{gj}^d > 0$  and  $M_{gj}^a > 0$  represent movement toward the advantaged or, equivalently, away from the advantaged (assuming, without loss of generality, that  $\bar{A}_{gj1} > 0.5$  and  $\bar{A}_{gj1} < \bar{A}_{gj1}^a$ , implying  $\bar{A}_{gj1}^d < \bar{A}_{gj1}$ ).

But this is not the only possible way of looking at domination qua attitude change. It may thus be clarifying to contrast  $D_{gj}$  with a couple of alternatives. One is the unweighted mean:

$$D'_{gj} = \frac{1}{2} (M_{gj}^d + M_{gj}^a),$$

the special case of  $D_{gj}$  in which  $d_g = 1/2$  for all *g*. Removing  $D_{gj}$ ’s dependence on  $d_g$  does more to contrast  $M_{gj}^d$  and  $M_{gj}^a$ .  $D'_{gj}$  shares  $D_{gj}$ ’s sign when  $\bar{A}_{gj1}^d$  and  $\bar{A}_{gj1}^a$  move in the same direction but can



have the opposite sign when they move in opposite directions ( $M_{gj}^d > 0$  and  $M_{gj}^a < 0$  or vice versa). If, for example,  $M_{gj}^d = 0.4$  and  $M_{gj}^a = -0.2$ ,  $D_{gj}$  can be negative when the disadvantaged are sufficiently few relative to the advantaged ( $D_{gj} = -0.08$  for  $d_g = 0.2$ ), but  $D'_{gj}$  is always positive (in this case,  $D'_{gj} = 0.1$ ), because the disadvantaged are moving further toward the advantaged than the advantaged toward the disadvantaged.

A second alternative, doing still more to contrast  $M_{gj}^d$  and  $M_{gj}^a$ , is

$$D''_{gj} = \frac{1}{2} (|M_{gj}^d| - |M_{gj}^a|) Q_{gj},$$

where  $Q_{gj} = 1$  for  $M_{gj}^d > 0$  and  $-1$  for  $M_{gj}^d < 0$ .  $D''_{gj} > 0$  when the disadvantaged move further toward the advantaged than the advantaged move in that same direction, and  $< 0$  when the disadvantaged move further away from the advantaged than the advantaged move in that same direction. For example, if  $M_{gj}^d = 0.2$ , and  $M_{gj}^a = 0.4$ ,  $D_{gj}$  and  $D'_{gj}$  ( $= 0.3$ ) both  $> 0$ , because the disadvantaged, the advantaged, and ergo the whole group are moving toward the advantaged, whereas  $D''_{gj} = -0.1$ , because the disadvantaged are moving less in that direction than the advantaged, whereas if  $M_{gj}^d = -0.2$ , and  $M_{gj}^a = -0.4$ , both  $D_{gj}$  and  $D'_{gj}$  ( $= -0.3$ )  $< 0$ , because the disadvantaged, the advantaged, and ergo the whole group are moving away from the advantaged, whereas  $D''_{gj} = 0.1$ , because the disadvantaged are moving less in that direction than the advantaged. Appendix C supplies a fuller account.

These alternatives would make sense for more sociometric notions of ‘domination’, comparing subgroup A’s influence on subgroup B with B’s influence on A. But what we are studying here – what is most relevant to deliberative democracy – is the bottom-line effects on the whole group’s attitudes. And for that,  $D_{gj}$  (like  $H_{gj}$  and  $P_{gj}$ ) is the best fit – and would be, whether we call it ‘domination’ or something else. (Juliet was right about names.)

### Parsing $D_{gj}$

It is nevertheless interesting to separate  $M_{gj}^d$ ’s and  $M_{gj}^a$ ’s contributions to  $D_{gj}$ . Given Equation (4),  $D_{gj} > 0$  may stem from  $M_{gj}^d > 0$ ,  $M_{gj}^a > 0$ , or both;  $D_{gj} < 0$  from  $M_{gj}^d < 0$ ,  $M_{gj}^a < 0$ , or both. The separate means and relative frequencies, in Table 3, evince two interesting patterns. First, the disadvantaged and advantaged move toward each other, each drawing the other’s attitudes in their direction ( $\bar{M}^d > 0$ ,  $\bar{M}^a < 0$ ). Second, the advantaged move slightly further toward the disadvantaged than vice versa (on all three dimensions, though not quite as far on the three combined), consistent with the slightly negative  $\bar{D}$ ’s.

### $D_{gj}$ ’s Dependence on $d_g$

A more extrapolatory question is the extent to which our results might differ for other dimensions of advantage. Let the whole-sample proportion who are disadvantaged be  $d$  (which, given random assignment, should be close to the unweighted mean of  $d_g$ ). For the individual advantages we examine here,  $d \cong 0.5$  – inherently for gender and by virtue of division at the whole-sample median for education and income. As we have argued, these operational thresholds make sense for these advantages. Our results for them are what they are. But what of other advantages, for which  $d$  might be much higher or lower? For home ownership in the United States,  $d < 0.5$ ; for having attended private school in the United States,  $d > 0.5$ . Let us therefore consider what  $\bar{D}$  might have been if  $d$  had been markedly higher or lower.

A simple approach is to estimate a bivariate, linear equation for  $D_{gj}$  as a function of  $d_g$ . The OLS-estimated slope is small and insignificant, and the  $R^2 < 0.001$ , which is already telling. The estimates imply, moreover, that, for  $d_g = 0.2$ ,  $\bar{D} = 0.012$  for gender,  $-0.013$  for education,  $-0.007$  for income, and  $-0.014$  for all three, while for  $d_g = 0.8$ ,  $\bar{D} = 0.002$  for gender,  $-0.005$

**Table 3.** Parsing domination

	$\bar{D}$	$\bar{M}^d$	$\bar{M}^a$	$\bar{D}^b$	$\bar{M}^{db}$	$\bar{M}^{ab}$
(A) By gender						
Mean	0.008	0.027	-0.033	0.464	0.495	0.440
s.e.	0.004	0.003	0.006	0.011	0.009	0.015
<i>p</i>	0.031	0.000	0.000	0.002	0.560	0.000
(B) By education						
Mean	-0.013	0.023	-0.057	0.447	0.523	0.412
s.e.	0.005	0.005	0.006	0.012	0.015	0.012
<i>p</i>	0.008	0.000	0.000	0.000	0.018	0.000
(C) By income						
Mean	0.000	0.036	-0.043	0.485	0.521	0.465
s.e.	0.005	0.007	0.006	0.017	0.017	0.016
<i>p</i>	0.951	0.000	0.000	0.425	0.203	0.033
(D) By all three						
Mean	-0.015	0.066	-0.031	0.466	0.495	0.455
s.e.	0.007	0.011	0.007	0.020	0.022	0.019
<i>p</i>	0.029	0.000	0.000	0.102	0.824	0.019

Note:  $\bar{D}$ ,  $\bar{M}^d$ , and  $\bar{M}^a$  denote the whole group’s mean movement toward the initial mean of disadvantaged, its disadvantaged members’ mean movement toward the initial mean attitude of the advantaged, and its disadvantaged members’ mean movement in that same direction.  $\bar{D}^b$ ,  $\bar{M}^{db}$ , and  $\bar{M}^{ab}$  are the relative frequencies with which those means are greater than 0.  $\bar{D}$  does not necessarily =  $\frac{1}{2}(\bar{M}^d + \bar{M}^a)$ , because  $d_g$  varies across groups and is likely correlated with  $M_{gj}^d$  and  $M_{gj}^a$ .

for education, 0.013 for income, and 0.003 for all three.<sup>22</sup> That is,  $\bar{D}$  would still show a bit more opposition than domination but not much of either if the disadvantaged were only 20 percent of each group and slightly more domination than opposition, but next-to-none of either if the disadvantaged were 80 percent of each group. In fine,  $\bar{D}$  does not appear to depend much on  $d_g$ .

### Motivated Reasoning Redux

Could the near-zero homogenization, polarization, and domination in our *DPs* be a mere artefact of motivated reasoning? In everyday discussions, defensive and social approval motivations may limit attitude change, thus reducing  $H_{gj}$ ,  $P_{gj}$ , and  $D_{gj}$  and (assuming no correlation between signs and reduction in magnitude)  $\bar{H}$ ,  $\bar{P}$ , and  $\bar{D}$ . But this is hardly a convincing explanation for our near-zero  $\bar{H}$ ,  $\bar{P}$ , and  $\bar{D}$ . *DPs* are not everyday discussions. Their briefing materials and expert panels afford more information and make uncongenial information harder to ignore. They explicitly cultivate *WM*, promoting even-handed evaluations, a sense of accountability for one’s views, and civic-mindedness, all of which should strengthen accuracy motivations (Bolsen, Cook, and Druckman 2014; Kam 2007; Lerner and Tetlock 1999). They also involve direct interactions with more heterogeneous others, strengthening *WM*’s ability to change attitudes (Mercier and Landemore 2012; Tuller et al. 2015), specifically toward their full-consideration counterparts.

In fact, *DPs* do generally produce considerable attitude change, as our present data make clear (see also, for example, Luskin, Fishkin, and Jowell 2002).<sup>23</sup> Across our 21 *DPs*, the mean absolute net change,  $Mean(|Mean(A_{ij2} - A_{ij1})|)$ , is 0.092, and the mean gross change,  $Mean(Mean(|A_{ij2} - A_{ij1}|))$ , 0.203.<sup>24</sup> Since these numbers may not speak for themselves, consider a simple, artificial scenario that would closely approximate them. Take a familiar five-point, Likert-type scale, with response categories running from ‘strongly disagree’ to ‘strongly agree’, linearly projected onto the [0, 1] interval as 0, 0.25, 0.5,

<sup>22</sup>For the three combined, for which the actual  $d$  (the whole-sample proportion who are either female or of below-median education or of below-median income) is already a shade over 0.8, we may also try  $d_g = 0.95$ . But that still yields  $\bar{D}$  = only 0.008.

<sup>23</sup>So, in some cases, do other deliberative designs (as in, e.g., Barabas 2004; Gastil, Black, and Moscovitz 2008; Himmelroos and Christensen 2014).

<sup>24</sup>In both cases, the inner mean is over  $i$ , and the outer one over  $j$ , then  $g$ .

0.75, and 1. Let 20 percent of the sample change from neutrality to agreement, another 20 percent from neutrality to strong agreement, and only 24 percent in the opposite direction, from disagreement to strong disagreement. In sum, nearly two-thirds (64 percent) of the sample change their response, nearly a third of them (20 percent of the sample) by two response categories. And two-thirds again as many (40 percent versus 24 percent) change toward 1 as toward 0, doing so, on average, by one-and-a-half times as much (1.5 categories versus 1.0). Here the mean absolute net change is 0.090, and the mean gross change 0.210—numbers within hairs of the 0.092 and 0.203 in our results. So the latter represent quite a lot of attitude change, both gross and net. The reason that  $\bar{H}$ ,  $\bar{P}$ , and  $\bar{D}$  hug 0 is *not* that the participants are simply clinging to their time-1 attitudes.

### *Deliberative Polling versus other Deliberative Designs*

Other deliberative designs do sometimes yield routine and strong homogenization, polarization, and domination. Although there is not yet much systematic evidence on specific design features' effects (notably excepting Karpowitz, Mendelberg, and Shaker 2012), several features characteristic of *DPs* but rare among other designs do figure to promote *WM*, inhibit *SD*, and thus limit homogenization, polarization, and domination (and their opposites). Three, in particular, leap out:

- (1) The task being set. Are the participants asked to reach a conscious, collective decision? To reach a consensus? Or simply to talk, listen, learn, and think about the issues? When the goal is consensus, homogenization is a demand characteristic. It is hardly surprising or informative when a design seeking consensus approaches it (consistent with research on compliance and conformity, as in, for example, Cialdini and Goldstein 2004; Carlson and Settle 2016). Striving to reach a conscious, collective decision, too, may create incentives to indulge emerging pluralities. Voilà, homogenization. More subtly, the pressure to agree may also hinder *WM* and allow *SD* freer rein, thus facilitating polarization and domination as well.<sup>25</sup> Designs asking the participants only to decide what they individually think entail no such task-based impetus toward homogenization, polarization, or domination.
- (2) The encouragement versus discouragement of interim, public expressions of bottom-line attitudes ('I prefer Policy X') as opposed to tributary beliefs about likely consequences or valuations thereof ('Policy X would produce more/less of Y, which would be a good/bad thing because ...'). For example, many designs require or encourage publicly tallied votes or shows-of-hands.<sup>26</sup> This can be regarded as a subtler version of (1), and it, too, may constitute a shove toward homogenization, polarization, and domination (consistent with Brauer et al. 2004; Levy and Sakaiya 2020).
- (3) The methods by which the participants are sampled, then assigned to groups. The ideal is random sampling, followed by random assignment, making every group a small random sample. Many designs attempt neither. Random sampling, even when attempted, is never fully realized in practice. Not everyone who is randomly selected can be reached or interviewed, and not everyone interviewed attends the event. Men, the young, the less well educated, and the socially marginal are particularly under-represented. So, still more relevantly, are the least interested in and knowledgeable about the topic. The magnitudes of these biases depend on details like the number of callbacks, the insistence with which anyone besides the designated respondent is excluded, the existence and size of an

<sup>25</sup>Collective decision making may also promote polarization by increasing risk-acceptance (Sunstein and Hastie 2014). Or by increasing homogenization, if the movement toward the group mean is concentrated among group members with weaker, less firmly anchored initial attitudes, presumably including those between 0.5 and  $\bar{A}_{g1}$  and those held by the disadvantaged.

<sup>26</sup>Some *DPs* have had midway measurements, but only via confidential questionnaires. The participants have no way of knowing the results.

honorarium, the venue's being away from home, etc. Thus some designs claiming to practice random sampling come reasonably close, while others do not. Even random assignment can vary in attainment. Participants sometimes have their own ideas about what group to join. All this matters because random samples should, on average, across hypothetically iterated sampling, be just as demographically and attitudinally heterogeneous as the population from which they are drawn. And the more heterogeneous the groups, the wider-ranging and more balanced the information their members exchange should be – which, as previously argued, should limit homogenization, polarization, and domination (see Levendusky, Druckman, and McLain 2016 and Strandberg, Himmelroos, and Gronlund 2019).

In all these respects, DPs stand apart. They do not task their participants with reaching any conscious, collective decision,<sup>27</sup> nor urge them toward (or away from) consensus; they sternly discourage interim public expressions of bottom-line opinion, including votes and shows-of-hands; and they employ high-quality random sampling,<sup>28</sup> followed by thoroughly random assignment, or the closest possible approximations thereof. The recruitment is well-organized and persistent, the participants are offered honoraria, and their travel and lodging are paid for. Small wonder, in this light, that DPs tend to produce much less homogenization, polarization, and domination than many other deliberative designs.

### Closing Remarks

Part of this study's value lies in its data. Scattered analyses of individual DPs and other deliberative events have reported broadly similar results regarding homogenization and polarization (Luskin, Fishkin, and Jowell 2002; Fishkin *et al.* 2010; Fishkin *et al.* 2011; Grönlund, Herne, and Setälä 2015). In finer grain, Siu (2009) finds that the disadvantaged and advantaged speak about equal numbers of words and for about equal lengths of time, consistent with little domination. But a more convincing test requires a larger number of groups, deliberating on a larger number and wider variety of policy issues, in a larger number and wider variety of contexts. Here we have examined twenty-one DPs, in multiple countries and at different times, encompassing 372 small groups, 139 policy issues, and 2,601 group-issue pairs.

The results show only irregular and feeble homogenization, polarization, and domination. The means are close to 0, the relative frequencies close to 0.5. This is *not* simply because the participants' attitudes do not change very much, as might be expected from heavily defensive motivated reasoning. They do change, appreciably, just not in ways regularly constituting homogenization, polarization, or domination. This faintness of pattern suggests a relatively deliberative discussion, involving considerable weighing of the merits, rather than just the social dynamics that would yield routine and strong homogenization, polarization, and domination.

The deliberative quality of the discussion – what makes a discussion a deliberation – matters because of its *effects* (cf. Lindell *et al.* 2017). Among other things, it should increase the participants' understanding of the issues, respect for others holding different views, and feelings of political efficacy. Still more critically for democracy, and more centrally for this article, it should move the participants' policy attitudes toward their full-consideration attitudes, thus refining – not distorting – the public will.

Our results do show slightly more homogenization than variegation, but also slightly more moderation than polarization and opposition than domination: a dash of one of our three worried-about attitude changes but dashes of the opposites of the other two. These inequalities

<sup>27</sup>A number of DPs have had policy consequences, affecting subsequent decision-making by other bodies. But the DPs themselves entailed no conscious, collective decision-making.

<sup>28</sup>The survey houses have included the National Opinion Research Center at the University of Chicago; the Survey Research Center at UC-Berkeley; Social and Community Planning Research, now the National Centre for Social Research in the UK; Polimetrix/YouGov; and Knowledge Networks.

are faint (hence not to be taken too seriously) but contingently plausible (hence not to be swept completely out of mind). They are what we should expect when the centrifugal forces in outside-world communications siloing and social inequalities slightly outweigh the centripetal ones in outside-world inattention, ignorance, and irreflection. As such, they represent deliberative *corrections to outside-world* distortions. We should expect neither the distortions nor therefore the corrections to be overwhelming. It is hard to imagine that the outside world's centrifugal and centripetal forces, both perennially strong, can ever be too out of balance, although the faint homogenization, moderation, and opposition we currently see could grow somewhat stronger, should our outside-world politics continue to wax more centrifugal (perish the thought).

This line of reasoning recalls the chasm between deliberation and everyday discussion. Many of the claims that deliberation inevitably produces homogenization, polarization, and domination rest on observations of everyday discussions or results from deliberative designs insufficiently different from them. Indeed, the prevalence and magnitude of homogenization, polarization, and domination suggests itself as a criterion for evaluating deliberative designs. A discussion in which they preponderate and are strong may not be much of a deliberation.

Among the further, beckoning questions are these: What accounts for the variation across groups and issues? Fuller-fledged explanatory models may provide some answers. What about combinations of these phenomena? For example, are polarization and domination (both changes in the mean) more problematic when accompanied by homogenization (shrinking variance)? What about the effects of specific elements of deliberative design? To what extent do the differences between DPs and other designs stem from the former's being consensus-neutral versus consensus-seeking, on their involving more rigorous random sampling and random assignment, etc.? Experimentally varying such features may permit some estimation. So may analyses of still more-meta meta-datasets, encompassing results from multiple designs (not just multiple DPs), varying in such features. But these present results should at least allay the concerns that deliberation intrinsically yields homogenization, polarization, and domination. In deliberative designs sufficiently capturing the deliberative ideal – sufficiently unlike everyday discussions – it does not.

**Supplementary material.** Online appendices are available at <https://doi.org/10.1017/S0007123421000168>.

**Data availability statement.** Data replication sets are available in Harvard Dataverse at: <https://doi.org/10.7910/DVN/D7GILO>

**Acknowledgements.** We are grateful to Jane Mansbridge, Mark Warren, Bruno Cautrès, Pierre Bréchon, Bernard Reber, Martial Foucault, Frédéric Gonthier, Mathieu Brugidou, Kalli Giannelos, Vincent Tournier, Peter Loewen, and Alice Siu, among others, along with several anonymous reviewers and the editor, for comments, and to Pete Mohanty for research assistance.

## References

- Achen CH and Bartels LM. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton University Press: Princeton, NJ.
- Albarracín D and Shavitt S (2018) Attitudes and attitude change. *Annual Review of Psychology* **69**, 299–327.
- Althaus S (2003) *Collective Preferences in Democratic Politics: Opinion Surveys and the Will of the People*. Cambridge: Cambridge University Press.
- Arceneaux K and Vander Wielen RJ (2017) *Taming Intuition: How Reflection Minimizes Partisan Reasoning and Promotes Democratic Accountability*. New York: Cambridge University Press.
- Barabas J (2004) How deliberation affects policy opinions. *American Political Science Review* **98**(4), 687–701.
- Bartels LM (1996) Uninformed voters: information effects in presidential elections. *American Journal of Political Science* **40**, 194–230.
- Beck PA et al. (2002) The social calculus of voting: interpersonal, media, and organizational influences on presidential choices. *American Political Science Review* **96**(1), 57–73.
- Bennett SE, Flickinger RS, and Rhine SL (2000) Political talk over here, over there, over time. *British Journal of Political Science* **30**, 99–119.
- Bohman J and Rehg W (eds.). (1997) *Deliberative Democracy: Essays on Reason and Politics*. Cambridge, MA: MIT Press.
- Bohner G and Dickel N (2011) Attitudes and attitude change. *Annual Review of Psychology* **62**, 391–417.

- Bolsen T, Cook F and Druckman JN** (2014) The influence of partisan motivated reasoning on public opinion. *Political Behavior* **36**(2), 235–262.
- Brauer M et al.** (2004) The relationship between expertise and evaluative extremity: the moderating role of experts' task characteristics. *Journal of Personality and Social Psychology* **86**(1), 5–18.
- Butters R and Hare C** (2020) Polarized networks? New evidence on American voters' political discussion networks. *Political Behavior*. Doi: 10.1007/s11109-020-09647-w.
- Carlson TN and Settle JE** (2016) Political chameleons: an exploration of conformity in political discussions. *Political Behavior* **38**(4), 817–859.
- Cialdini RB and Goldstein NJ** (2004) Social influence: compliance and conformity. *Annual Review of Psychology* **55**, 591–621.
- Cohen J** (1989) Deliberation and democratic legitimacy. In Hamlin A and Pettit P (eds), *The Good Polity*. Oxford: Blackwell.
- Converse PE** (1970) Attitudes and non-attitudes: continuation of a dialogue. In Tufte ER (ed.), *The Quantitative Analysis of Social Problems*. Reading, PA: Addison- Wesley.
- Cowan SK and Baldassarri D** (2018) 'It could turn ugly': selective disclosure of attitudes in political discussion networks. *Social Networks* **52**, 1–17.
- Delli Carpini MX and Keeter S** (1996) *What Americans Know About Politics and Why It Matters*. New Haven, CT: Yale University Press.
- Druckman JN** (2012) The politics of motivation. *Critical Review* **24**(2), 199–216.
- Druckman JN and McGrath MC** (2019) The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change* **9**, 111–119.
- Dryzek J** (2002) *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. Oxford: Oxford University Press.
- Elster J** (ed.). (1998) *Deliberative Democracy*. Cambridge: Cambridge University Press.
- Farrar C et al.** (2010) Disaggregating deliberation's effects: an experiment within a deliberative poll. *British Journal of Political Science* **40**, 333–347.
- Fishkin JS** (1991) *Democracy and Deliberation: New Directions for Democratic Reform*. New Haven, CT: Yale University Press.
- Fishkin JS et al.** (2010) Deliberative democracy in an unlikely place: deliberative polling in China. *British Journal of Political Science* **40**, 435–448.
- Fishkin JS et al.** (2011) Deliberative agenda setting: piloting reform of direct democracy in California. *Perspectives on Politics* **13**(4), 1030–1042.
- Fraser N** (1993) Rethinking the public sphere: a contribution to the critique of actually existing democracy. In Robbin B (ed.), *The Phantom Public Sphere*. Minneapolis: University of Minnesota Press.
- Gastil J, Black L, and Moscovitz K** (2008) Ideology, attitude change, and deliberation in small face-to-face groups. *Political Communication* **25**, 23–46.
- Gerber AS et al.** (2012) Disagreement and the avoidance of political discussion: aggregate relationships and differences across personality traits. *American journal of Political Science* **56**(4), 849–874.
- Grönlund K, Herne K, and Setälä M** (2015) Does enclave deliberation polarize opinions? *Political Behavior* **37**(4), 995–1020.
- Gutmann A and Thompson D** (1996) *Democracy and Disagreement*. Cambridge, MA: Harvard University Press.
- Gutmann A and Thompson D** (2004) *Why Deliberative Democracy?* Princeton, NJ: Princeton University Press.
- Habermas J** (1990) Discourse ethics: notes on a program of philosophical justification. In Benhabib S and Dallmayer F (eds), *The Communicative Ethics Controversy*. Cambridge, MA: MIT Press, pp. 60–110.
- Habermas J** (1996) *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge, MA: MIT Press.
- Hahn U and Harris AJL** (2014) What does it mean to be biased: motivated reasoning and rationality. *Psychology of Learning and Motivation* **61**, 41–102.
- Hansen KM** (2004) *Deliberative Democracy and Opinion Formation*. Odense, Denmark: University Press of Southern Denmark.
- Hart W et al.** (2009) Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological Bulletin* **135**(4), 55–588.
- Himmelroos S and Christensen HS** (2014) Deliberation and opinion change: evidence from a deliberative mini-public in Finland. *Scandinavian Political Studies* **37**(1), 41–60.
- Huckfeldt R, Johnson PE, and Sprague J** (2004) *Political Disagreement: The Survival of Diverse Opinions Within Communication Networks*. New York: Cambridge University Press.
- Huckfeldt R and Sprague J** (1995) *Citizens, Politics, and Social Communication*. Cambridge: Cambridge University Press.
- Kam CD** (2007) When duty calls, do citizens answer? *Journal of Politics* **69**(1), 17–29.
- Karpowitz CF and Mendelberg T** (2014) *The Silent Sex: Gender, Deliberation, and Institutions*. Princeton, NJ: Princeton University Press.
- Karpowitz CF, Mendelberg T and Shaker L** (2012) Gender inequality in deliberative participation. *American Political Science Review* **106**(3), 533–547.
- Kinder DR and Kalmoe NP** (2017) *Neither Liberal nor Conservative: Ideological Innocence in the American Public*. Chicago, IL: University of Chicago Press.
- Kunda Z** (1990) The case for motivated reasoning. *Psychological Bulletin* **108**(3), 480–498.
- Lau RR and Redlawsk DP** (1997) Voting correctly. *American Journal of Political Science* **91**(3), 585–598.



- Leeper TJ (2014) The informational basis for mass polarization. *Public Opinion Quarterly* 78(1), 27–46.
- Leeper TJ and Slothuus R (2014) Political parties, motivated reasoning, and public opinion formation. *Political Psychology* 35, 129–156.
- Lerner JS and Tetlock PE (1999) Accounting for the effects of accountability. *Psychological Bulletin* 125(2), 255–275.
- Levendusky MS, Druckman JN, and McLain A (2016) How group discussions create strong attitudes and strong partisans. *Research & Politics* 3(2), 1–6.
- Levy HI and Sakaiya S (2020) Effect of deliberation style on the gender gap in deliberative participation. *Japanese Journal of Political Science* 21(3), 158–175.
- Lindell M et al. (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.
- Lodge M and Taber CS (2013) *The Rationalizing Voter*. New York: Cambridge University Press.
- Luskin RC (1987) Measuring political sophistication. *American Journal of Political Science* 31, 856–899.
- Luskin RC (2003) The heavenly public: what would the ideal democratic citizenry be like? In MacKuen MB and Rabinowitz G (eds), *Electoral Democracy*. Ann Arbor: University of Michigan Press.
- Luskin RC, Fishkin JS and Jowell R (2002) Considered opinions: deliberative polling in Britain. *British Journal of Political Science* 32, 455–487.
- Luskin RC, Sood G, Fishkin JS, Hahn KS (2021), Replication data for: Deliberative distortions? Homogenization, polarization, and domination in small group deliberations, <https://doi.org/10.7910/DVN/D7G1LO>, Harvard Dataverse, V1, UNF:6:TXrLWTP8cXLVzPOWQytJ/A== [fileUNF]
- Mansbridge J (1983) *Beyond Adversary Democracy*. Chicago, IL: University of Chicago Press.
- Mansbridge J (1999a) On the idea that participation makes better citizens. In Elkin S and Soltan KE (eds.), *Citizen Competence and Democratic Institutions*. State College, PA: Penn State University Press.
- Mansbridge J (1999b) Everyday talk in the deliberative system. In Macedo S (ed.), *Deliberative Politics: Essays on Democracy and Disagreement*. New York: Oxford University Press.
- Már K and Gastil J (2020) Tracing the boundaries of motivated reasoning: how deliberative minipublics can improve voter knowledge. *Political Psychology* 41(1), 107–127.
- McPherson M, Smith-Lovin L, and Cook JM (2001) Birds of a feather: homophily in social networks. *Annual Review of Sociology* 27, 415–444.
- Mercier H and Landemore H (2012) Reasoning is for arguing: understanding the successes and failures of deliberation. *Political Psychology* 33(2), 243–258.
- Mutz D (2006) *Hearing the Other Side: Deliberative Versus Participatory Democracy*. Cambridge: Cambridge University Press.
- Ozanne JL, Corus C, and Saaticioglu B (2009) The philosophy and methods of deliberative democracy: implications for public policy and marketing. *Journal Of Public Policy & Marketing* 28(1), 29–40.
- Sanders LM (1997) Against deliberation. *Political Theory* 25(3), 347–376.
- Searing DD et al. (2007) Public discussion in the deliberative system: does it make better citizens? *British Journal of Political Science* 37, 587–618.
- Siu A (2009) *Look Who's Talking: Examining Social Influence, Opinion Change, and Argument Quality in Deliberation*. PhD dissertation, Department of Communication, Stanford University.
- Squires J (2008) Deliberation, domination and decision-making. *Theoria* 55(117), 104–133.
- Strandberg K, Himmelroos S, and Gronlund K (2019) Do discussions in like-minded groups necessarily lead to more extreme opinions? Deliberative democracy and group polarization. *International Political Science Review* 40(1), 41–57.
- Suhay E (2015) Explaining group influence: the role of identity and emotion in political conformity and polarization. *Political Behavior* 37, 221–251.
- Sunstein CR (2002) The law of group polarization. *The Journal of Political Philosophy* 10(2), 175–195.
- Sunstein CR (2009) *Going to Extremes: How Like Minds Unite and Divide*. Oxford: Oxford University Press.
- Sunstein CR and Hastie R (2014) *Wiser: Getting Beyond Groupthink to Make Groups Smarter*. Cambridge, MA: Harvard Business Review Press.
- Tuller HM et al. (2015) Seeing the other side: perspective taking and the moderation of extremity. *Journal of Experimental Social Psychology* 59, 18–23.
- White HA (1980) Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48: 817–838.
- Wojcieszak M (2011) Deliberation and attitude polarization. *Journal of Communication* 61, 596–617.
- Young IM (2000) *Inclusion and Democracy*. Oxford: Oxford University Press.
- Zuber JA, Crott HW and Werner J (1992) Choice shift and group polarization: an analysis of the status of arguments and social decision schemes. *Journal of Personality and Social Psychology* 62(1), 50–61.

**Cite this article:** Luskin RC, Sood G, Fishkin JS, Hahn KS (2022). Deliberative Distortions? Homogenization, Polarization, and Domination in Small Group Discussions. *British Journal of Political Science* 52, 1205–1225. <https://doi.org/10.1017/S0007123421000168>