

This is a preprint of an article published in [Farrar, Cynthia, James S. Fishkin, Donald P. Green, Christian List, Robert C. Luskin, and Elizabeth Levy Paluck. 2010. Disaggregating Deliberations Effects: An Experiment within a Deliberative Poll. *British Journal of Political Science* 40(2): 333-347. Reprinted (2010) in Jostein Gripsrud, Hallvard Moe, Anders Molander, and Graham Murdock (eds.), *The Public Sphere*. Thousand Oaks, CA: SAGE Publications]. Pagination of this preprint may differ slightly from the published version.

Disaggregating Deliberation's Effects: An Experiment within a Deliberative Poll

CYNTHIA FARRAR, JAMES S. FISHKIN, DONALD P. GREEN, CHRISTIAN LIST, ROBERT C. LUSKIN AND ELIZABETH LEVY PALUCK*

Using data from a randomized field experiment within a Deliberative Poll, this paper examines deliberation's effects on both policy attitudes and the extent to which ordinal rankings of policy options approach single-peakedness (a help in avoiding cyclical majorities). The setting was New Haven, Connecticut, and its surrounding towns; the issues were airport expansion and revenue sharing - the former highly salient, the latter not at all. Half the participants deliberated revenue sharing, then the airport; the other half the reverse. This split-half design helps distinguish the effects of the formal on-site deliberations from those of other aspects of the treatment. As expected, the highly salient airport issue saw only a slight effect, while much less salient revenue-sharing issue saw a much larger one.

Fed by normative concerns and empirical questions about the quality of public opinion and popular decision making, recent years have seen the emergence of a growing body of research on deliberation's ability to improve matters.¹ Some of the most apt data, in our view, come from *Deliberative Polling*, which seeks to gauge the opinions people would

* Farrar and Green: both at Yale University, Institution for Social and Policy Studies, and Department of Political Science, respectively; Fishkin: Department of Communication, Stanford University; List, Departments of Philosophy and Political Science, London School of Economics and Political Science; Luskin, Department of Government, University of Texas at Austin (email: rcluskin@mail. [texas.edu](mailto:rcluskin@mail.texas.edu)); Paluck: Department of Psychology and Woodrow Wilson School of Public and International Affairs, Princeton University. Earlier versions of this article were presented at the annual meeting of the International Society of Political Psychology, Boston, Mass., 2003, and at the ECPR General Conference, Marburg, Germany, 2003. The authors are grateful to Ethan Leib, the League of Women Voters of Connecticut Education Fund Inc.; The Guild Group and Regional Plan Association for assistance in organizing and carrying out the split-half Deliberative Poll; and to the Carnegie Corporation, the Ford Foundation, the Renee B. Fisher Foundation, Fannie Mae, William C. Graustein, the New Haven Savings Bank, the South Central Connecticut Regional Water Authority, the United Way of Greater New Haven and the United Illuminating Company for making the project possible. The Center for Deliberative Polling and the Public Policy Clinic at the University of Texas at Austin, the Community Foundation for Greater New Haven and Yale University's Institution for Social and Policy Studies also provided financial and in-kind support. The research and some of the analysis and writing were conducted while Fishkin and Luskin were Fellows at the Center for Advanced Study in the Behavioral Sciences, supported by grants from the William and Flora Hewlett Foundation (Grant #2000-5633), the Center General Fund and the University Research Institute of the University of Texas. Alice Siu contributed invaluable research assistance. *Deliberative Polling*® is a trade mark of James S. Fishkin. Any fees from the trade mark are used to support research at the Center for Deliberative Democracy at Stanford University.

¹ See, e.g., Robert C. Luskin, James S. Fishkin and Roger Jowell, 'Considered Opinions: Deliberative Polling in Britain', *British Journal of Political Science*, 32 (2002), 455-87; Jason Barabas, 'How Deliberation Affects Policy Opinions', *American Political Science Review*, 98 (2004), 687-701; James N. Druckman,

hold if they knew and thought more about the issues. The design gives random samples both balanced information and the opportunity of talking with one another and questioning policy experts. Beginning in 1994, there have been more than thirty national, regional/statewide/provincial and local Deliberative Polls (DPs) in the United States, Britain, Australia, Denmark, Hungary, Bulgaria, Italy, Northern Ireland, China and Brazil. More recently, there have also been two transnational (pan-European) and four on-line DPs (the latter all national and all in the United States).²

The present study examines three important hypotheses about deliberation's effects. The first is that deliberation frequently alters policy *attitudes* (continuous dispositions towards policy alternatives) – not only at the individual level ('gross change') but in the aggregate ('net change'). The second is that deliberation tends to bring policy *preferences* (ordinal rankings of policy alternatives) closer to *single-peakedness*, a help in avoiding cyclical majorities of the sort identified by Condorcet and Arrow.³ The third is that both these effects tend to be stronger for less salient issues, on which less deliberation has already occurred. Since very few issues are highly salient, these hypotheses together imply an important role for deliberation in shaping majorities and making them meaningful.

The before-and-after contrasts in Deliberative Polling are consistent with all three hypotheses. The experience does frequently change policy attitudes and bring policy preferences closer to single-peakedness.⁴ The less salient the issue, moreover, the greater these changes.⁵ But this evidence still leaves some important questions.

One preliminary question is the extent to which the before-and-after changes result from the DP experience at all. The whole public, after all, could be changing at the same time, in the same ways, and to the same degree. That may be generally unlikely but is at least possible for issues sufficiently in the headlines and on people's lips.⁶ Several previous DPs have therefore compared the participants to 'quasi-control groups' consisting of either reinterviewed 'non-participants' (members of the initial random sample who declined to participate) or an independent random sample interviewed at roughly the time

(F'note continued)

'Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects', *American Political Science Review*, 98 (2004), 671–86.

² See James S. Fishkin, *The Voice of the People: Public Opinion and Democracy* (New Haven, Conn.: Yale University Press, expanded paperback edition, 1997); Robert C. Luskin, James S. Fishkin and Roger Jowell, 'Considered Opinions: Deliberative Polling in Britain', *British Journal of Political Science*, 32 (2002), 455–87; James S. Fishkin and Robert C. Luskin, 'Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion', *Acta Politica*, 40 (2005), 284–98.

³ Marie Jean Antoine Nicolas Caritat (the Marquis de) Condorcet, *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix* (Paris, 1785); Kenneth Arrow, *Social Choice and Individual Values* (New York: Wiley, 1953).

⁴ See Robert C. Luskin, James S. Fishkin and Kyu S. Hahn, 'Deliberation and Net Attitude Change' (paper presented at the biennial General Conference of the European Consortium for Political Research, Pisa, Italy, 2007); Christian List, Robert C. Luskin, James S. Fishkin and Iain McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls' (unpublished, London School of Economics, 2007).

⁵ See Luskin, Fishkin and Hahn, 'Deliberation and Net Attitude Change'; List, Luskin, Fishkin and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy'.

⁶ This is a plausible worry for DPs conducted during and about referendum or election campaigns, as well as in scattered other cases, including DPs about educational policy in Northern Ireland, just as the current accord was being reached, and about foreign policy in the United States on the eve of the Iraq war.

the DP was ending.⁷ These comparisons lend some reasonable assurance that the before-and-after changes do indeed result from 'something' in the DP experience.⁸

That something, moreover, would seem to be a function of deliberation.⁹ The participants typically learn a great deal,¹⁰ and those who emerge knowing the most tend both to change their views the most¹¹ and to account for most of the approach to single-peakedness.¹²

But that still leaves the question of exactly what aspect(s) of the treatment is (are) responsible. The DP treatment includes everything from the initial interview to the final questionnaire: among other things, the invitation to participate; briefing materials laying out competing arguments; a weekend's worth of formal, balanced deliberation in both small group discussions and plenary questioning of panels of policy experts and policy makers; the casual, generally much less balanced anticipatory deliberation between the initial interview and the weekend; and the conversational spillover into corridors and dining rooms during the weekend. Of particular interest is the formal, on-site deliberation. The whole intervention may be considered broadly deliberative, but this is the heart of it, as well as the greatest departure from the participants' everyday experience and the closest approximation to what theorists of 'deliberative democracy' have in mind.

This study uses a randomized experiment within a DP to estimate the effect of the formal, on-site deliberation. The experiment randomly assigns the participants to deliberate one or the other of two distinct policy issues, one much more salient than the other,

⁷ Of course, both samples are corroded – unequally – by non-response/non-participation, depriving these comparisons of the full authority of true random assignment. Barabas's analysis of much the same question involves broadly similar quasi-control groups (created in his case by matching), though with a highly non-random participant sample and hence lesser external validity.

⁸ See also James S. Fishkin and Robert C. Luskin, 'Bringing Deliberation to the Democratic Dialogue: The NIC and Beyond', in Maxwell McCombs and Amy Reynolds, eds, *A Poll with a Human Face: The National Issues Convention Experiment in Political Communication*, (Mahwah, NJ: Lawrence Erlbaum, 1999); Robert C. Luskin, James S. Fishkin, Ian McAllister, John Higley and Pamela Ryan, 'Information Effects in Referendum Voting: Evidence from the Australian Deliberative Poll' (paper presented at the annual meeting of the American Political Science Association, Boston, Mass.; 2002); Barabas, 'How Deliberation Affects Policy Outcomes'.

⁹ The changes could conceivably if implausibly stem simply from the stimulation of the initial interview, which even without the invitation to and anticipation of the DP should produce some heightened attention and thought. Luskin, Fishkin, McAllister, Higley, and Ryan, 'Information Effects in Referendum Voting: Evidence from the Australian Deliberative Poll'; Robert C. Luskin and James S. Fishkin, 'Deliberative Polling, Public Opinion and Democracy: the Case of the National Issues Convention' (paper presented at the annual meeting of the American Association for Public Opinion Research, 1998); and Robert C. Luskin, Shanto Iyengar and James S. Fishkin, 'Considered Opinions on U.S. Foreign Policy: Face-to-Face versus Online Deliberative Polling' (Center for Deliberative Democracy, Stanford University, 2006) address this concern by comparing to quasi control groups likewise interviewed before as well as after the deliberations.

¹⁰ See, e.g., Luskin, Fishkin and Jowell, 'Considered Opinions'; Robert C. Luskin, James S. Fishkin, Roger Jowell and Allison Park, 'Learning and Voting in Britain: Insights from the Deliberative Poll' (paper presented at the annual meeting of the American Political Science Association, Atlanta, Ga. 1999); Robert C. Luskin, James S. Fishkin and Dennis L. Plane, 'Deliberative Polling and Policy Outcomes: Electric Utility Issues in Texas' (paper presented at the annual meeting of the Association for Public Policy Analysis, 1999); Luskin, Fishkin, Higley and Ryan, 'Information Effects in Referendum Voting'.

¹¹ See, e.g., Luskin, Fishkin and Jowell, 'Considered Opinions'; and Luskin, Fishkin, Higley and Ryan, 'Information Effects in Referendum Voting'.

¹² See List, Luskin, Fishkin and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy'.

then has them answer the same questions as when first interviewed, then has them deliberate the other issue, and finally has them answer the same questions again. The midterm measurement – at the point at which the participants have all had the same treatment, except for deliberating one issue versus the other – is particularly revealing. To the extent that it is the on-site deliberation that is producing the overall change, the attitude change and approach to single-peakedness should be greater, on each issue, among those who have just finished deliberating that issue than among those who have just finished deliberating the other issue. We also expect the deliberation to have a greater effect on the less salient issue.

DELIBERATION AND SALIENCE, POLICY ATTITUDES AND PROXIMITY TO SINGLE-PEAKEDNESS

Our first hypothesis, again, is that deliberation frequently produces aggregate change in policy *attitudes*, regarded as continuous dispositions towards policy alternatives. Individual-level change should be common. Even in ordinary surveys, respondents frequently give different answers when reinterviewed, partly or mostly as a result of non-attitudinal bouncing-around.¹³ The deliberation in Deliberative Polling should make for greater and more purposive change. Some participants can be expected to draw truer, tauter connections between their policy attitudes and their own more fundamental values and interests. Others may redefine their interests or re-weight their values (perhaps sometimes gravitating towards thinking in terms of a wider public interest).¹⁴ Of course, these individual-level changes could largely cancel out, with some participants moving one way, and others moving equally the other way: there could be much gross but little net change. But there is no reason to expect that, and considerable reason to suspect the contrary, given some correlation between interests and initial thought and information. Which way any net change should run is another question, beyond our scope here. The answer undoubtedly depends on the issue and the setting, but net change should in any event be common – as the evidence from previous DPs confirms.¹⁵

Our second hypothesis concerns policy *preferences*, regarded as ordinal rankings of policy alternatives. Here, our contention – that deliberation tends to increase what we shall call ‘proximity to single-peakedness’ – needs more explanation. A combination of preferences is *single-peaked* across individuals if the alternatives can be aligned on some ‘structuring’ dimension, say from left to right, such that every individual has a most preferred alternative and a decreasing preference for other alternatives as they get more

¹³ Philip E. Converse, ‘The Nature of Belief Systems in Mass Publics’, in David E. Apter, ed., *Ideology and Discontent* (New York: Free Press, 1964). Cf. Jennifer Hill and Hanspeter Kriesi, ‘An Extension and Test of Converse’s “Black-and-White” Model of Response Stability’, *American Political Science Review*, 95 (2001), 397–414.

¹⁴ To the extent that the first mechanism predominates, we should expect to see attitudes change so as to increase their predictability from socio-demographic variables, proxying interests; to the extent that the second predominates, they may change so as to decrease it. See Luskin, Fishkin and Jowell, ‘Considered Opinions’; Robert C. Luskin, ‘The Heavenly Public: What Would the Ideal Citizenry Be Like?’, in Michael B. MacKuen and George Rabinowitz, eds, *Electoral Democracy* (Ann Arbor: University of Michigan Press, 2003).

¹⁵ Across fifty policy attitude indices in seven DPs, the mean absolute net change of opinion ranges from 0.068 to 0.117, averaging 0.092, impressively large for net attitude change (for reasons sketched in Luskin 2002). Thirty-four (68 per cent) of the fifty indices show statistically significant net change. See Luskin, Fishkin and Hahn, ‘Deliberation and Net Attitude Change’.

distant in either direction from it.¹⁶ This is single-peakedness as originally defined by Black and Arrow,¹⁷ sometimes called 'ordinal' single-peakedness, as distinct from the 'cardinal' or 'spatial' single-peakedness in the spatial voting model.¹⁸

Single-peakedness matters because it affords an escape from the possibility of cyclical collective preferences in pairwise majority voting, as in Condorcet's paradox.¹⁹ If one third of an electorate prefer x to y to z , another third prefer y to z to x , and the remaining third prefer z to x to y , two-to-one majorities prefer x to y , y to z , and z to x . The winning alternative depends on the pair of alternatives put forward. Such 'majority cycles' and their numerous generalizations²⁰ undermine the meaningfulness of majority rule.²¹ But single-peakedness precludes cycling,²² ensuring a *Condorcet winner* (an alternative that beats, or is tied with, all others in pairwise majority voting).

Note that single-peakedness is a binary property – a combination of preferences is either single-peaked, or it is not – that in populations (or samples) of any size is exceedingly unlikely ever to obtain. Following List, Luskin, Fishkin and McLean,²³ we therefore define *proximity to single-peakedness* as $S = m/n$, where m is the size of the largest subset of sample members whose combination of preferences is single-peaked and n is the overall sample size ($m \leq n$).²⁴ If there were a dimension on which everyone's preferences were single-peaked, m would equal n , and S would equal 1.²⁵ Proximity to

¹⁶ For example, consider a population/sample of four individuals, with (declining) preference orderings (x, y, z) , (z, y, x) , (y, x, z) , and (y, z, x) . These preference orderings are single-peaked, with structuring dimension $[x, y, z]$. Each individual has a first preference, a second preference on one side of it, and a third preference no 'closer' (in terms of ordinal slots) to it. By contrast, if a fifth individual with the preference ordering (x, z, y) is added to this population/sample, the preferences are no longer single-peaked. They are not single-peaked with respect to the structuring dimension $[x, y, z]$, because the fifth individual prefers x , then z , which is 'further' than y from x on this dimension. And there is no other possible structuring dimension (left-right ordering of x, y and z) with respect to which all five preference orderings are single-peaked. For a more precise definition, see List, Luskin, Fishkin and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy'.

¹⁷ Duncan Black, 'On the Rationale of Group Decision-Making', *Journal of Political Economy*, 56 (1948), 23–34 and Kenneth Arrow, *Social Choice and Individual Values*.

¹⁸ See Steven J. Brams, Michael A. Jones and D. Marc Kilgour, 'Single-Peakedness and Disconnected Coalitions', *Journal of Theoretical Politics*, 14 (2002), 359–83. Cardinal single-peakedness is sufficient but unnecessary for ordinal single-peakedness if the space is one-dimensional, but neither sufficient nor necessary if the space is multi-dimensional.

¹⁹ Condorcet, *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*.

²⁰ For example, Arrow, *Social Choice and Individual Values*, and Richard McKelvey, 'General Conditions for Global Intransitivities in Formal Voting Models', *Econometrica*, 47 (1979), 1085–111.

²¹ William H Riker, *Liberalism against Populism*, (San Francisco: W. H. Freeman, 1982).

²² Black, 'On the Rationale of Group Decision-Making'.

²³ List, Luskin, Fishkin and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy'.

²⁴ In the second example of fn. 16, the largest subset of the population/sample of five whose preferences are single-peaked contains four members (the four of the first example, with respect to the structuring dimension $[x, y, z]$). Thus $S = 4/5 = 0.8$. For a more precise definition and discussion, see List, Luskin, Fishkin and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy'. See also Richard G. Niemi, 'Majority Decision-Making with Partial Unidimensionality', *American Political Science Review*, 63 (1969), 488–97.

²⁵ Note that S concerns the aggregate patterning of preferences *across* individuals, not necessarily the cognitive organization of individuals' preferences (although it may partly reflect that). See the further discussion below and in List, Luskin, Fishkin and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy'.

single-peakedness bears a strong positive relationship to the probability of the existence of a Condorcet winner and a strong negative relationship to the probability of cycles.²⁶

Deliberation should increase proximity to single-peakedness.²⁷ As people talk, learn and think about the relationships between the alternatives and the criteria for choosing between them, they may simply adopt an ordering they come to recognize as conventional among political elites. Or they may influence each other's thinking, acquiring more of a shared understanding of what the relevant issue-space is and how the alternatives are positioned within it. Or they may independently excogitate a natural ordering urged if not quite compelled by logic. By whatever mix of such mechanisms, deliberation should tend to make preferences more single-peaked,²⁸ and the evidence from previous DPs suggests that it does.²⁹

Our third hypothesis is that both these effects should be weaker for more salient issues, which have already received more real-world deliberation. If the issue is salient enough, many people may already be near their full-information positions, if the real-world deliberation is sufficiently good, or entrenched far from them, if not. In either case, there is less opportunity for Deliberative Polling to have much effect (although, if the real-world deliberation is sufficiently flawed, it might, if the DP lasted far longer than a weekend). Thus the before-and-after contrasts in previous DPs have tended to show larger net attitude changes and smaller increases in proximity to single-peakedness for less salient issues.³⁰

THE SPLIT-HALF DELIBERATIVE POLL: DESIGN AND MEASUREMENT

To isolate the effects of the formal on-site deliberations, we have built a fully randomized field experiment into a DP. A random sample drawn from the fifteen towns surrounding New Haven, Connecticut, deliberated two issues: the level of service to be provided by the

²⁶ Niemi, 'Majority Decision-Making with Partial Unidimensionality'. This is also supported by computer simulations reported in List, Luskin, Fishkin, and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy'.

²⁷ See David Miller, 'Deliberative Democracy and Social Choice', *Political Studies*, 40 (1992: special issue): 54–67; Jack Knight and James Johnson, 'Aggregation and Deliberation: On the Possibility of Democratic Legitimacy', *Political Theory*, 22 (1994), 277–96; Christian List, 'Two Concepts of Agreement', *PEGS: The Good Society*, 11 (2003), 72–9; List, Luskin, Fishkin and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy'. Cf. David Van Mill, 'The Possibility of Rational Outcomes from Democratic Discourse and Procedures', *Journal of Politics*, 58 (1996), 734–52, countered by John Dryzek and Christian List, 'Social Choice Theory and Deliberative Democracy: A Reconciliation', *British Journal of Political Science*, 33 (2003), 1–28.

²⁸ This does *not* necessarily mean that deliberation will lead people to converge on some particular ranking, what List calls 'agreement at a substantive level'. In fact, the results in List, Luskin, Fishkin and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy', suggest that it may *decrease* agreement of this sort. Our only claim here is that deliberation increases proximity to single-peakedness, an instance of what List calls 'agreement at a meta-level'. See List, 'Two Concepts of Agreement'.

²⁹ Across thirteen issues, the index of 'proximity to single-peakedness', defined below, increases by an average of 0.101, on a scale that runs from a variable minimum always > 0 to 1.0. Across the ten low to moderate salience issues, the average was 0.134 (computed from Table 2 in List, Luskin, Fishkin and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy').

³⁰ Across seven DPs, the correlation between pre-deliberation knowledge, proxying salience, and the mean absolute net change for all the policy attitude indices in that DP is 2.61. See Luskin, Fishkin, and Hahn, 'Deliberation and Net Attitude Change'. Across thirteen DPs, the correlation between pre-deliberation knowledge and the increase in "proximity to single-peakedness" (defined below) is 2.59 (computed from Tables 1 and 3 in List, Luskin, Fishkin, and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy').

local airport and what, if any, sharing there should be of property-tax revenues from new commercial development.

These issues differed markedly in salience. The airport was centre stage, revenue sharing barely in the theatre. Over the preceding year, the region's most widely circulated daily newspaper mentioned revenue sharing only seven times³¹ but ran seventy-four articles, plus editorials and letters from citizens, on airport expansion. The airport coverage spanned the full range of commonly held views. Thus the participants would already have thought, learnt and talked far more about the airport than about revenue sharing. Their airport attitudes would have been more firmly rooted, their airport preferences already closer to single-peaked.

The on-site deliberations extended from the evening of Friday, 1 March, through midday on Sunday, 3 March 2002. Of an initial interview sample of 1,032, a total of 133 showed up. The Friday evening session, at which participants dined with members of their randomly assigned small groups, was designed to orient the participants and acquaint them with one another. The actual deliberations began Saturday morning. Those interviewees who said they would attend were sent the briefing materials, and those who did attend were paid \$200 on completing the final questionnaire.

The participants were generally representative. Compared with the 'non-participants' (the initial interviewees who did not attend), they were somewhat more highly educated and more likely to be from New Haven itself but comparable in income, gender, race and voter registration (see Appendix A for details). The geographic bias, probably attributable to the longer commute from suburbs and outlying towns, does not seem to affect the results. The views of those residing in New Haven moved in the same direction and to the same extent as the views of those residing in the surrounding towns.

The sixteen small groups were randomly assigned to one of the two possible orders in which the two issues could be deliberated. Eight groups (containing sixty-four participants) deliberated the airport Saturday morning and revenue sharing Saturday afternoon, and the other eight (containing sixty-eight participants) the reverse. We denote these two treatment groups as 'A-first' and 'R-first', respectively.³²

The formal on-site deliberations consisted of three 'deliberative sessions', each involving both small-group discussions and plenary questions-and-answers with panels of policy experts and advocates. The first two sessions, occupying the whole of Saturday, concentrated on one issue apiece, with the first confined to the airport for the A-first treatment group and to revenue sharing for the R-first treatment group, and the second to revenue sharing for the A-first group and to the airport for the R-first group.³³ The third, on Sunday morning, was more synoptic, with all the participants revisiting both issues in their small groups and then questioning a panel of local and state officials about both.

There were three waves of measurement: the initial telephone interview (T1), a written version of the same questionnaire after the first deliberative session (T2), and the same written version (plus a few additional questions) again at the end of the weekend (T3). The T1-T2 interval thus spans both the first deliberative session on-site and the casual,

³¹ There was much discussion of residential property taxes and property tax relief for the elderly, disabled and low-income, but regional revenue sharing was only mentioned briefly in one editorial and as a low-profile issue in the platform of one unsuccessful mayoral candidate.

³² We shall thus be referring to two sorts of 'groups' – the small groups, within which the issues are discussed, and the treatment groups, each consisting of eight small groups, which tackle the issues in different sequences.

³³ The A-first and R-first groups thus had different plenary sessions on Saturday but shared a plenary session on Sunday.

anticipatory learning and deliberation between the first interview at T1 and the beginning of the weekend. The T2–T3 interval spans the second and third deliberative sessions on-site.

From the standpoint of the randomized experiment, the T1–T2 changes are particularly revealing. At T2, both treatment groups had had the experience of deliberating in a casual, less balanced way, with relatively homogeneous interlocutors, in the period between the initial interview and their arrival on site. Both had also had the experience of the more formal and balanced deliberation with more heterogeneous interlocutors on site, but on different issues. The A-first group had deliberated the airport but not the revenue sharing, and the R-first group the reverse. Thus the contrast between their T1–T2 changes should reflect the effects of the A-first group's airport deliberations and of the R-first group's revenue-sharing deliberations.

The T2–T3 changes can be used similarly, but less certainly, to assess the effects of the second deliberative session, in which the R-first small groups switched to the airport, and the A-first small groups to revenue sharing. This comparison is harder to interpret, since by T2 each group had already deliberated the other issue, and the T2–T3 interval bracketed not only the second deliberative session but also the third, during which both issues were deliberated. Perhaps discussing revenue sharing is better preparation for discussing the airport than the reverse. Perhaps considering them simultaneously at the end alters the effect of the earlier sequencing. Or perhaps discussing an issue from T1 to T2 continues to have an effect from T2 to T3, even while the other issue is discussed. How far the T2–T3 comparison should be expected to mirror the T1–T2 comparison is, therefore, unclear.

The questionnaire asked about the main policy alternatives on each issue. For the airport, these were:

- A1. *Maintaining*: 'Commercial passenger service to nearby cities should be maintained but not expanded to serve a larger market.'
- A2. *Expanding*: 'Commercial passenger service should be expanded to provide more flights to more places.'
- A3. *Ending*: 'Commercial passenger service should be ended, leaving only service for private airplanes.'

For revenue sharing, the main alternatives were:

- R1. *Local Control*: 'My town should maintain local control over all of its tax revenues from new businesses and industries.'
- R2. *Voluntary Sharing*: 'My town should try for a voluntary agreement with other towns in the region to share some of the tax revenues from new businesses and industries.'
- R3. *State-encouraged Sharing*: 'The state should provide incentives for towns in the region to share some tax revenues from new businesses and industries.'
- R4. *Mandatory Sharing*: 'The state should require towns in the region to share some tax revenues from new businesses and industries.'

With regard to the expected level of sharing, the two non-mandatory sharing options (voluntary and state-encouraged) lie between local control, on the one side, and mandatory sharing, on the other.

We asked policy questions of two distinct sorts: *ratings*, to gauge attitudes, and *rankings*, to gauge preferences. The rating questions asked whether the respondent agreed strongly, agreed somewhat, neither agreed nor disagreed, disagreed somewhat, or disagreed strongly with each policy option. The ranking questions asked which option was the respondent's first choice, then which was his or her second choice, and then, in the case of revenue-sharing, which was his or her third choice. The lowest-ranked choice can be inferred from the others.

POLICY ATTITUDES

As we shall see, the DP moved the participants towards wanting to end rather than expand airport services and towards favouring non-mandatory revenue sharing over both mandatory sharing and local control. To summarize airport attitudes, therefore, we subtract the ending from the expanding service rating. Scoring both items from 0, for strong disagreement, to 1, for strong agreement, yields a difference that runs from 1 for strong agreement with expanding service and strong disagreement with ending it to 21 for the reverse. To summarize revenue-sharing attitudes, we create two companion indices, one pitting the two middle, non-mandatory sharing options against local control, and the other pitting them against mandatory sharing. In each case, we average the ratings of the two non-mandatory options and subtract the rating of the alternative. The indices run from 1 for strong agreement with non-mandatory sharing and strong disagreement with mandatory sharing/local control and 21 for the reverse.³⁴

Table 1 shows the results. On the airport, although the sample, from start to finish, was more inclined to expand than to end airport services, that inclination faded significantly ($p < 0.05$) over the course of the experiment. The mean airport attitude decreased from 0.540 to 0.434. On revenue sharing, the sample initially favoured local control over non-mandatory sharing and the latter over mandatory sharing but came to favour non-mandatory sharing over both other options. The mean of the index opposing non-mandatory sharing to local control increased from 20.115 to 0.241 ($p < 0.001$), and that of the index opposing non-mandatory to mandatory sharing from 0.166 to 0.385 ($p < 0.001$). Note that the more salient airport issue shows less net change (only 0.106, compared to 0.335 and 0.220 on the two revenue-sharing indices). Again these results echo those of previous DPs, which have frequently shown statistically significant net attitude changes and have done so more frequently on less salient issues.

The present question, however, is the extent to which the net change results from the formal on-site deliberations. Thus consider the contrast between the two treatment groups over the T1-T2 interval. Take first the less salient revenue-sharing issue, where the net change is much greater. The mean attitude shifts dramatically from local control towards non-mandatory sharing in the R-first group, discussing the issue during this interval (0.313), but scarcely budges in the A-first group discussing the airport instead (20.012). The difference is highly significant ($p < 0.001$). The contrast is fainter but similar for non-mandatory versus mandatory sharing. From T1 to T2, the R-first group moves twice as far towards non-mandatory sharing (0.085 versus 0.039), although in this case the 0.085 is not quite significant ($p = 0.115$), and the difference between 0.085 and 0.039 is insignificant. In all, these results suggest that the on-site deliberations drove most of the net attitude change on revenue sharing.

³⁴One might expect preferences to reflect attitudes – more after deliberation than before and more, at least initially, on the airport than on revenue sharing – and the results bear this out. Maximum likelihood estimation of binomial logit models expressing preferences for expanding airport service over ending it, for non-mandatory revenue sharing over local control, and for non-mandatory over mandatory revenue sharing as functions of the respective attitude indices produce pre- and post-deliberation (McFadden's) pseudo- R^2 s of 0.479 and 0.680, 0.377 and 0.345, and 0.099 and 0.371, respectively. Preferences are indeed relatively well predicted by attitudes by the end of the process, are indeed initially better predicted on the airport than on revenue sharing, and are indeed better predicted after than before deliberation in two of the three cases. More detailed results are available on request.

TABLE 1 *Policy Attitude Changes*

	T1	T2	T3	T2-T1	Sig.	T3-T2	Sig.	T3-T1	Sig.
<i>A. The Airport: Ending v. Expanding Service</i>									
A-First (<i>n</i> = 64)	0.500	0.332	0.336	0.168 (0.074)	0.027	-0.004 (0.072)	0.957	0.164 (0.078)	0.040
R-First (<i>n</i> = 68)	0.577	0.493	0.526	0.085 (0.062)	0.175	-0.033 (0.074)	0.655	0.051 (0.072)	0.477
Whole sample (<i>n</i> = 132)	0.540	0.415	0.434	0.125 (0.048)	0.010	-0.019 (0.052)	0.714	0.106 (0.053)	0.048
(A-First) - (R-First)	-0.077	-0.161	-0.190	0.083 (0.096)	0.386	0.029 (0.104)	0.779	0.113 (0.106)	0.291
<i>B. Revenue Sharing: Voluntary Sharing or Incentives v. Local Control</i>									
A-First (<i>n</i> = 64)	-0.074	-0.086	0.232	-0.012 (0.049)	0.812	0.318 (0.049)	0.000	0.307 (0.059)	0.000
R-First (<i>n</i> = 68)	-0.153	0.160	0.250	0.313 (0.060)	0.000	0.090 (0.043)	0.042	0.403 (0.068)	0.000
Whole sample (<i>n</i> = 132)	-0.115	0.041	0.241	0.155 (0.040)	0.000	0.201 (0.034)	0.000	0.356 (0.045)	0.000
(R-First) - (A-First)	-0.078	0.246	0.018	0.324 (0.078)	0.000	-0.228 (0.065)	0.001	0.096 (0.091)	0.293
<i>C. Revenue Sharing: Voluntary Sharing or Incentives v. Mandatory Sharing</i>									
A-First (<i>n</i> = 64)	0.172	0.211	0.432	0.039 (0.050)	0.434	0.221 (0.059)	0.000	0.260 (0.044)	0.000
R-First (<i>n</i> = 68)	0.160	0.244	0.342	0.085 (0.053)	0.115	0.097 (0.032)	0.004	0.182 (0.048)	0.000
Whole sample (<i>n</i> = 132)	0.166	0.228	0.385	0.063 (0.036)	0.087	0.157 (0.033)	0.000	0.220 (0.033)	0.000
(R-First) - (A-First)	-0.012	0.034	-0.090	0.045 (0.073)	0.533	-0.123 (0.066)	0.064	-0.078 (0.066)	0.240

Note: Standard errors in parentheses; *p*-values are two-tailed.

The parallel comparison over the T2-T3 interval reinforces the inference. Here it is the A-first group, now discussing revenue sharing, that moves furthest towards non-mandatory sharing. The R-first group continues to move in the same direction, perhaps as a delayed effect of their earlier deliberation, but less so. When the question is non-mandatory sharing versus local control, the change is 0.318 in the A-first group versus 0.090 in the R-first group; when it is non-mandatory versus mandatory sharing, it is 0.221 versus 0.097. In both cases, the difference is highly significant ($p < 0.001$, $p = 0.064$).

The results on the airport, a more salient issue, where there is less net attitude change to apportion, are weaker and more mixed. From T1 to T2, both treatment groups shift towards ending the service. Here too it is the group discussing the issue that changes noticeably more (0.168 versus 0.085), and the 0.168 is statistically significant ($p = 0.027$), while the 0.085 is not ($p = 0.175$). The difference between the 0.168 and the 0.085, however, is statistically insignificant ($p = 0.386$), and neither treatment group shows any real change from T2 to T3. These results suggest some slight effect of the on-site deliberation but also suggest some effect of the casual at-home deliberations preceding (and stirred by the prospect of) the deliberative weekend: thus the noticeable and not-too-different changes in both treatment groups from T1 to T2, the interval containing the at-home deliberations, and the stability of both from T2 to T3.

PROXIMITY TO SINGLE-PEAKEDNESS

Our second hypothesis is that deliberation tends to increase proximity to single-peakedness, defined, as above, as $S = m/n$. Two further points about S are worth noting. First, S necessarily ≤ 1 , and if all individual preference orderings are strict (no one is indifferent between any pair of alternatives), S also $\geq S_m = 2^{(k-1)}/k! > 0$, where k is the number of alternatives.³⁵ When $k = 53$ (as on the airport), $S_m = 2/3$; when $k = 4$ (as on revenue-sharing), $S_m = 1/3$. Since our data inevitably contain incomplete rankings, implying non-strict orderings, however, S_m is only an approximate lower bound. Secondly, the dimension on which m is premised (the one on which the largest subset of participants has single-peaked preferences) may vary with the treatment group, over time, or both. This leaves S 's sampling distribution unknown, although its standard error may be bootstrapped.

Table 2 shows S and its bootstrapped standard error for both the whole sample and the two treatment groups at all three measurements.³⁶ Absent S 's sampling distribution, we refrain from assertions of statistical 'significance' or 'insignificance',³⁷ but the estimated standard errors make it hard to imagine that the larger changes and differences would be insignificant.

The results mostly parallel those for attitudes. Overall – across the whole sample from T1 to T3 – both issues show increased proximity to single-peakedness. Again, and again in keeping with other DP results,³⁸ the change is minor (0.77 to 0.81) on the high salience

³⁵ Niemi, 'Majority Decision-Making with Partial Unidimensionality'.

³⁶ A thousand random samples of the relevant subjects were drawn with replacement. The standard deviation of the resampled S provides the estimate of the standard error.

³⁷ The bootstrapped standard errors, moreover, are for S , not the difference between the values of S at different times, which may be somewhat larger or (less likely) smaller, depending on the sign and magnitude of the covariance.

³⁸ List, Luskin, Fishkin and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy'.

TABLE 2 *Changes in Proximity to Single-Peakedness (S)**

	<i>n</i>	T1	T2	T3
<i>A. Whole Sample</i>				
Airport	132	0.77 (0.033)	0.84 (0.031)	0.81 (0.032)
Revenue Sharing	132	0.52 (0.042)	0.70 (0.037)	0.80 (0.037)
<i>B. Airport, by Treatment Group</i>				
A-First	64	0.80 (0.043)	0.81 (0.049)	0.86 (0.049)
R-First	68	0.82 (0.045)	0.88 (0.049)	0.84 (0.036)
<i>C. Revenue Sharing, by Treatment Group</i>				
A-First	64	0.56 (0.060)	0.63 (0.062)	0.84 (0.044)
R-First	68	0.47 (0.057)	0.78 (0.052)	0.76 (0.051)

*Standard errors (in parentheses) bootstrapped, as described in fn. 36.

airport issue but major (0.52 to 0.80) on the low salience revenue-sharing one.³⁹ On the airport, the contrast between treatment groups suggests little evidence of the on-site deliberation's having had much effect. *S* hardly budges, for either treatment group. On revenue sharing, however, each treatment group shows a large increase, and almost entirely during the interval during which it deliberates the issue on site. From T1 to T2, *S* increases by 0.31 in the R-first group but only by 0.07 in the A-first group. From T2 to T3, it increases by 0.21 in the A-first group but decreases by 0.02 in the R-first group.⁴⁰ As in the case of net attitude change, the DP's effect on proximity to single-peakedness appears to be mostly an effect of the on-site deliberation.

DISCUSSION

These results comport with the notion that deliberation tends to produce net attitude change and bring preferences closer to single-peakedness, at least on issues of low to moderate salience, if not necessarily on the most salient. We know from previous DP results that there is a strong negative correlation between the approach to single-peakedness and the magnitude of the net change, on the one hand, and the salience of the issue, on the other. On sufficiently salient issues, both the mean attitude and the proximity to single-peakedness may sit relatively still. But on less salient issues, where the DP produces more movement, it appears from these results, admittedly based on only two issues, to be the on-site deliberation that is responsible for the lion's share of it. The highly salient airport issue sees only minor change, while the not at all salient revenue sharing issue sees very large changes, mostly during the interval during which the issue is deliberated on site.

³⁹ On the airport, the dimension along which the largest sub-sample is single-peaked remains the same throughout, ordering the alternatives as [A2 A1 A3]. The Condorcet winner, throughout, is A2 (expanding service). On revenue sharing, the dimension along which the largest sub-sample is single-peaked also remains the same throughout, ordering the alternatives as [R1 R2 R3 R4], although the Condorcet winner changes from one non-mandatory sharing option to the other – from R3 (state-encouraged sharing) at T1 to R2 (voluntary sharing) at T3.

⁴⁰ In each treatment group, the Condorcet winner changes from R3 to R2 over the interval during which the group deliberates the issue – from T1 to T2 in the R-first group and from T2 to T3 in the A-first group.

We previously based this assessment of relative salience on local newspaper coverage, but our data provide additional evidence. Consider in this light the T1 proximity to single-peakedness: at T1, S is only 0.52 for revenue sharing but 0.77 for the airport. Of course, that gap might be thought to exist simply because S 's approximate lower bound S_m is 1/3 for the revenue-sharing issue, with its four alternatives, but 2/3 for the airport, with its mere three. But the adjusted measure $S' = (S - S_m)/(1 - S_m)$, taking account of the number of alternatives, is still noticeably higher for the airport (0.32) than for revenue sharing (0.27) at T1.⁴¹ So the initial difference in proximity to single-peakedness is not simply due to the difference in the number of alternatives. The participants would appear to have entered the process already having thought more about and thus acquired more single-peaked preferences about the airport issue.

The levels of relevant T1 knowledge lend further support. At all three waves, we asked the participants whether (1) the region's population was closest to 250,000, 350,000, 550,000 or 750,000; (2) its rate of growth in employment during the 1990s was more than, about the same as, or less than in the rest of the United States; (3) New Haven's population increased, decreased or did not change during the 1990s; (4) the major source of revenue for most of the region's town governments is sales taxes, property taxes, direct state subsidies or direct federal subsidies; (5) state law allows communities to share property tax revenues; (6) those communities with the most valuable property tend to have the lowest, average or the highest property tax rates; (7) the Federal Aviation Authority classifies the regional airport as a major hub, a medium hub, a minor hub or not a hub; and (8) maintaining the regional airport at its current level of service would require any significant investment. The correct answers are (1) 555,000, (2) less, (3) decreased, (4) property taxes, (5) yes, (6) the lowest, (7) a non-hub, and (8) yes. Items (7) and (8) are specifically relevant to the airport, items (4) – (6) specifically relevant to revenue sharing, and items (1) – (3) generally relevant to the politics of the region.

Table 3 shows the percentages of the whole sample and of the two treatment groups answering the two airport items, the three revenue-sharing items, the three general items and all eight items correctly at T1, T2 and T3. At T1, the participants averaged answering 38.6 per cent of the airport questions but only 25.4 per cent of the revenue-sharing questions correctly. Of course, the revenue-sharing knowledge items could simply be intrinsically harder. But the fact that by T3 the participants actually fared slightly *better* on the revenue-sharing items (59.9 per cent versus 54.2 per cent) suggests otherwise. The participants would appear to have entered the process knowing substantially more about the airport because it was the more salient issue.

Table 3 also strengthens the claim that it was in fact deliberation producing the net attitude change and approach to single-peakedness. Deliberation entails learning, and, as is usual in Deliberative Polling, the participants learnt a great deal. For the whole sample across all eight items, the mean percentage correct increased by 22.3 per cent.⁴² The gain

⁴¹ In words, S' is the extent to which S exceeds its minimum, expressed as a fraction of the extent to which it could do so. Note that S' necessarily $\leq S$ (indeed $< S$ for $S < 1$) and that because S_m is only approximate, S' should be taken as heuristic. For more on S and S' , see List, Luskin, Fishkin and McLean, 'Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy'.

⁴² The T1-T2 gains tend to exceed their T2-T3 counterparts, no doubt because T1-T2 is a much longer interval, bracketing the receipt of the briefing materials and the anticipatory, off-site deliberations, as well as the first on-site deliberative session.

TABLE 3 *Knowledge Gains*

	T1	T2	T3	T2-T1	T3-T2	T3-T1
<i>A. Whole Sample, by Topic (n = 132)</i>						
Airport	0.386	0.508	0.542	0.121***	0.034	0.155***
Revenue sharing	0.254	0.492	0.599	0.239***	0.106***	0.345***
General	0.417	0.606	0.612	0.189***	0.006	0.195***
Overall	0.368	0.553	0.591	0.185***	0.038**	0.223***
<i>B. Airport, by Treatment Group</i>						
A-First (n = 64)	0.398	0.555	0.563	0.156***	0.008	0.164***
R-First (n = 68)	0.375	0.463	0.522	0.088**	0.059*	0.147***
(A-First)-(R-First)	0.023	0.091*	0.040	0.068	-0.051	0.017
<i>C. Revenue Sharing, by Treatment Group</i>						
A-First (n = 64)	0.273	0.336	0.523	0.063*	0.188***	0.250***
R-First (n = 68)	0.235	0.640	0.669	0.404***	0.029	0.434***
(R-First)-(A-First)	-0.038	0.304***	0.146**	0.342***	-0.158**	0.184**

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (based on one-tailed tests).

was greater on revenue sharing, the less salient issue. The percentage correct increased by 34.5 per cent on revenue sharing versus 15.5 per cent on the airport.

The on-site learning, moreover, appears to be deliberation-based, as the lower tiers of Table 3 suggest. From T1 to T2, the participants learnt something, presumably before arrival, about the issue they were not initially deliberating on site (since the briefing materials and anticipatory, off-site deliberations covered both) but a good deal more, presumably after arrival, about the topic they were deliberating. The A-first group gained 15.6 per cent on the airport but only 6.3 per cent on revenue sharing, while the R-first group gained 40.4 per cent on revenue sharing but only 8.8 per cent on the airport. From T2 to T3, there was some additional learning but only about the topic being deliberated then. The A-first group gained 18.8 per cent on revenue sharing but less than 1 per cent on the airport, while the R-first group gained 5.9 per cent on the airport but only 2.9 per cent on revenue sharing.⁴³

The net attitude change and increase in proximity to single-peakedness on revenue sharing line up nicely with these knowledge gains. From T1 to T2, the A-first group learnt a good deal about revenue sharing, changed its views a good deal, and showed a modest increase in proximity to single-peakedness. But the R-first group learnt still more, changed its views still more, and showed a much greater increase in proximity to single-peakedness. From T2 to T3, the A-first group, then deliberating revenue sharing, learnt a good deal on top of what it had learnt from T1 to T2, changed its views a good deal further, and added greatly to its T1-T2 increase in proximity to single-peakedness. The R-first group learnt only a little, changed its views only a little, and showed no increase in proximity to single-peakedness (compare Table 3 with Tables 1 and 2). Thus the net attitude change and approach to single-peakedness not only occur *when* the participants

⁴³ By contrast, the general knowledge items show significant but roughly equal gains in both the A-first and R-first groups from T1 to T2 and thus from T1 to T3 (results not shown). On these more general matters, there is no reason to expect the members of either treatment group to learn more during either interval, and they do not.

deliberate the issue, but, as these coincident knowledge gains suggest, because they deliberate it.

All told, these results buttress the case for deliberative democracy. They support the proposition that at least on policy issues that are not too salient – the great majority of policy issues – deliberation frequently changes attitudes and makes preferences more single-peaked. A more deliberative democracy should, therefore, sometimes produce majorities favouring different policies, parties or candidates and in turn, at least sometimes, bring different governments, enacting different policies. It should also produce more meaningful majorities, based on preferences that are closer to single-peaked.

APPENDIX A : DEMOGRAPHIC COMPARISONS OF PARTICIPANTS, NON - PARTICIPANTS AND VOTING POPULATION

	Participants (n = 132)	Non-participants (n = 1,024)	Voting population
Age (in years)	50.1	50.1	47.0
Registered to vote	90.2%	88.6%	78.7%
Marital status			
Single	30.3%	29.3%	29.0%
Married	49.2%	52.4%	52.5%
Divorced/Separated/Widowed	19.7%	17.3%	18.5%
Education			
Less than/some high school	2.3%	4.2%	17.0%
High school graduate	9.1%	20.2%	30.8%
Some college	25.0%	21.1%	18.2%
College graduate	28.0%	29.1%	21.7%
Trade/Technical	3.0%	2.5%	-
Graduate school	32.6%	22.2%	12.4%
Income	\$61–70,000	\$61–70,000	\$64,018
Race			
African-American	12.1%	7.2%	8.3% [§]
Caucasian	72.7%	75.3%	78.1%
Hispanic or Latino*	3.0%	4.4%	9.6%
Other	9.1%	9.1%	4.0%
Gender			
Male	50.8%	47.2%	46.8% [§]
Female	49.2%	52.8%	53.2%

Note: Percentages may not add up to 100 in non-participant blocks due to 'refusal' category.

*For population 15 and over.

†For population 25 and over.

§For population 18 and over.