Disaggregating Deliberation's Effects:

An Experiment within a Deliberative Poll*

Cynthia Farrar, Yale University

James S. Fishkin, Stanford University

Donald P. Green, Yale University

Christian List, London School of Economics

Robert C. Luskin, University of Texas at Austin

Elizabeth Levy Paluck, Yale University

Abstract

Using data from a randomized field experiment within a Deliberative Poll, we examine 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 issues were airport expansion and revenue-sharing in New Haven, Connecticut and its surrounding towns. Half the participants deliberated revenue-sharing, then the airport, the other half the reverse. This split-half design enables us to distinguish the effects of the formal on-site deliberations from those of other aspects of the Deliberative Polling treatment. We find that the formal on-site deliberations accounted for much of the Deliberative Polling effect on one issue, though not the other—thus both confirming deliberation's capacity to shape attitudes and preferences and raising the question of how its effects may depend on the kind of issue being deliberated. We suggest that deliberation's effects are larger for less salient issues.

Recent years have seen the emergence of the instrument of public consultation known as *Deliberative Polling*. Ordinary polls seek to gauge the opinions people actually hold, Deliberative Polls to gauge the opinions they *would* hold if they knew and thought more. The design provides random samples with information and gives them the opportunity of discussing the issues with one another and questioning policy experts about them. Beginning in 1994, there have been more than twenty Deliberative Polls, including eleven in the U.S. (two national and nine regional), five in Britain, two in Australia, and one each in Denmark, Hungary, Bulgaria, and China (all national, except for the last). Since 2003, there have also been four national online Deliberative Polls, all in the U.S. Fueled by normative and empirical concerns about the quality of popular decision-making, a growing body of research is using the data from Deliberative Polls to gauge how deliberation shapes people's views (as in Fishkin 1997; Fishkin and Luskin 1999; Luskin, Fishkin, and Jowell 2002; Luskin, Iyengar, and Fishkin 2003).[1]

The present study examines two important hypotheses about deliberation's effects. The first is that deliberation frequently alters policy attitudes, both at the individual level and in the aggregate. 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 (1785) and Arrow (1951).

Previous evidence from Deliberative Polling supports both hypotheses. The participants do frequently change their views, both at the individual level and in the aggregate, and their preferences do generally come closer to single-peakedness. Both effects, moreover, seem related to learning. The participants typically learn a great deal (Luskin, Fishkin, and Jowell 2002; Luskin, Fishkin, Jowell, and Park 1999; Luskin, Fishkin, and Plane 1999; Fishkin and Luskin 1999; Luskin et al. 2000), and those who emerge knowing the most tend both to change their

views the most (Luskin, Fishkin, and Jowell 2002, Luskin et al. 2000) and to account for most of the approach to single-peakedness (List, Luskin, Fishkin, and McLean 2006).[2]

But questions about the sources of these before-after changes remain. One question is the extent to which they result from the Deliberative Polling 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. Several previous Deliberative Polls have therefore compared the participants to "quasi control groups" consisting of either reinterviewed "nonparticipants" (members of the initial random sample who declined to participate) or an independent random sample interviewed at roughly the time the Deliberative Poll was ending.[3] These comparisons lend some considerable assurance that the before-after changes do indeed result from something in the Deliberative Polling experience (see, e.g., Luskin and Fishkin 1999, Luskin et al. 2000).[4]

A second, so far less explored question is the extent to which the before-after changes stem from deliberation, as distinct from other aspects of the Deliberative Polling experience. Previous Deliberative Polls have involved one, grand, undifferentiated treatment, consisting of everything bracketed by the initial interview and final questionnaire. That includes the invitation to participate; the briefing materials laying out competing arguments; a weekend's worth of formal, balanced deliberation; the casual, generally much less balanced anticipatory deliberation between the initial interview and the deliberative weekend; and the conversational spillover into corridors and dining rooms during the weekend. The treatment thus involves more than just discussion, and the discussion is not all of a kind, nor all confined to the weekend. Yet the heart of the intervention is the formal deliberation during the weekend, which constitutes the greatest

departure from the participants' everyday experience and the closest approximation to what theorists of "deliberative democracy" have in mind.

This study reports on the first Deliberative Poll designed to estimate the specific contribution of the formal on-site deliberation. In an enfolded randomized experiment, the participants are randomly assigned to deliberate one or the other of two distinct policy issues, then answer the same questions as when first interviewed and recruited, then deliberate the other issue, then answer the same questions again. The midterm measurement—at the point at which the participants have 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.

Our results show some interesting differences between the two issues. Both show the usual Deliberative Polling effects, but in one case the effects are stronger and seem to stem much more from the on-site deliberation than in the other. Thus we conclude by considering how deliberation's effects may vary with the issue. We suggest that deliberation's effects are apt to be stronger on less salient issues.

Deliberative Democracy and Deliberative Polling

Democracy, according to deliberative democrats, should not just aggregate preferences but help shape them. Votes and opinions should emerge from processes of discussion and reflection (Elster 1998). The attainability of this ideal is unclear. The public may never achieve much more than the modest and imperfect deliberation that already occurs. Competing demands for time and attention, coupled with typically minuscule probabilities of actually affecting the

outcome, may make ignorance too rational (Downs 1957). But while it may not be possible to get everyone to discuss and reflect seriously on policy issues, it *is* possible to get random samples of a few hundred to do so. The resulting distributions of policy and electoral preferences provide glimpses of what the whole public would think if it deliberated to the same degree. That is the strategy of Deliberative Polling.

The basics of the design are these: A random sample is drawn, interviewed, and invited to attend a weekend of deliberations at a common site. Those agreeing to attend are sent carefully balanced briefing materials laying out the major arguments for and against the major policy proposals. On-site, they discuss the issues in randomly assigned small groups led by trained moderators and question balanced panels of competing experts or policy-makers in plenary sessions. Then, at the end, they answer the same questions as at the beginning. They receive a financial inducement for participating, and the deliberations are generally televised.

The intent is to approach a counterfactual ideal in which deliberation is not only more pervasive but in several important senses "better"—more substantive, better informed, more balanced, more deeply reflective, and more inclusive in the sense of involving more socio-demographically and attitudinally diverse discussants. The briefing materials are provided to jump start the participants' learning and thinking about the issues. The briefing materials and expert panels are carefully balanced, and the small group moderators strive to ensure that all the arguments in the briefing materials get considered. The participants read, hear, and voice arguments and counter-arguments. The moderators keep the discussions on-topic and civil. The combination of random sampling and random assignment maximizes the heterogeneity of both the attitudes expressed and the people expressing them.

These features make the formal on-site deliberations very different from naturally occurring discussion in the real world—much closer to what those writing of "deliberative democracy" have in mind. The participants talk with people very unlike themselves, expressing views very unlike their own, and in circumstances in which it is difficult not to give them and their views serious attention. We suspect that these artificial—more ideal—features give the deliberation in Deliberative Polling much of its effect.

Deliberation, Policy Attitudes, and Proximity to Single-Peakedness

But let us say a bit more about the effects we are looking for, why we expect them, and why they matter. Our first hypothesis, again, is that deliberation frequently changes policy attitudes, both individually and in the aggregate, although the degree and direction of change will naturally depend on the nature of the issue and the circumstances of the day. (For some speculations, see Luskin 2003.) One possible mechanism is that the participants come to draw truer, tauter connections between their policy attitudes and their own more fundamental values and interests. Another is that they come to redefine their interests or re-weight their values. In particular, we suspect that they sometimes gravitate toward thinking in terms of a wider public interest.[5] In either case, we should expect change. Of course the individual-level changes may 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 absent any reason to expect such balancing to be the rule—and if anything there is reason to suspect the contrary, given some correlation between interests and initial thought and information—we may expect net change to be quite common (as the evidence from previous Deliberative Polls, reported, e.g. in Fishkin and Luskin 1999, Luskin, Fishkin, and Jowell 2002, confirms).

Our second hypothesis—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 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 distant in either direction from it. This is single-peakedness as originally defined by Black (1948) and Arrow (1951)—sometimes also called "ordinal," as distinct from "spatial" or "cardinal" single-peakedness in the spatial voting model.[6]

Single-peakedness matters because it affords an escape from the possibility of cyclical collective preferences in pairwise majority voting, as in Condorcet's paradox (1785). 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 generalizations (e.g., Arrow 1953, McKelvey 1979) undermine the meaningfulness of majority rule (Riker 1982). But single-peakedness precludes cycling (Black 1948), 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 isn't. In populations (or samples) of any size, it is exceedingly unlikely ever to obtain. Following List, Luskin, Fishkin, and McLean (2006), we therefore define *proximity to single-peakedness*, a non-binary property, as $S = m/n$, where m is the size of a largest subset of sample members whose combination of preferences is single-peaked and n is the overall sample size ($m \leq n$).[7] If there were a dimension on which everyone's preferences were single-peaked, m would equal n, and S would equal 1.[8] Proximity to single-peakedness bears a strong positive

relationship to the probability that a Condorcet winner exists and a strong negative relationship to the probability of cycles (Niemi 1969).[9]

Deliberation, we argue, should increase proximity to single-peakedness (see Miller 1992, Knight and Johnson 1994, Dryzek and List 2003, List 2002, List, Luskin, Fishkin, and McLean 2006).[10] As people talk, learn, and think about the relationships among the alternatives and the criteria for choosing among 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.[11]

Both these hypotheses reflect important effects. To the extent that deliberation changes the distribution of policy attitudes, majorities supporting given policies, parties, or candidates may appear or disappear. A more deliberative democracy might bring different governments, enacting different policies. It should also bring more meaningful majorities, strengthening the case for democracy, at least for deliberative democracy.[12]

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 Deliberative Poll. A random sample drawn from the fifteen towns surrounding New Haven, Connecticut, deliberated two issues: the level of service to be provided by the local airport and what if any sharing there should be of property-tax revenues from new commercial development. The on-site deliberations extended from Friday evening, March 1, through midday Sunday, March 3, 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 "nonparticipants" (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.

At the beginning of the weekend, the participants were randomly assigned to one of sixteen small groups, and the small groups in turn randomly assigned to one of the two possible orders in which the two issues could be deliberated. Eight groups (containing 64 participants) deliberated the airport Saturday morning and revenue-sharing Saturday afternoon, the other eight (containing 68 participants) the reverse. We denote these two treatment groups as "A-first" and "R-first," respectively.[13]

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.[14] 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, anticipatory learning and deliberation occurring 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 T2 measurement is particularly revealing. At that point, one treatment group had deliberated the airport but not revenue-sharing, while the other had deliberated revenue-sharing but not the airport. Both 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 randomization provides assurance that they differed minimally in other ways. On each issue, therefore, the contrast between the A-first and R-first T2 attitudes should reflect the effect of the on-site deliberations on that issue.

The T3 measurement 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 T3-T2 comparison should be expected to mirror the T2-T1 comparison is therefore unclear.

The questionnaire asked respondents to both rate and rank the main policy alternatives on each issue. For the airport, these were:

> A1. "Commercial passenger service to nearby cities should be maintained but not expanded to serve a larger market."
>
> A2. "Commercial passenger service should be expanded to provide more flights to more places."
>
> A3. "Commercial passenger service should be ended, leaving only service for private airplanes."

For short, these alternatives were to *maintain*, *expand*, or *end* the existing service.

For revenue-sharing, the main alternatives were:

> R1. "My town should maintain local control over all of its tax revenues from new businesses and industries"
>
> R2. "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. "The state should provide incentives for towns in the region to share some tax revenues from new businesses and industries."

R4. "The state should require towns in the region to share some tax revenues from new businesses and industries."

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

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 the respondent which option was his or her 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. We use the rating questions for the analysis of policy attitudes, the ranking questions for the analysis of proximity to single-peakedness.

Policy Attitudes

As we shall see, the Deliberative Poll moved our participants toward wanting to end rather than expand airport services and toward favoring non-mandatory revenue sharing as opposed to either mandatory sharing or local control. To summarize airport attitudes, we therefore subtracted the expanding from the ending 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 -1 for the reverse. To summarize revenue-sharing attitudes we created two companion indices, one pitting the two middle, non-mandatory sharing options against local control, the other pitting them against mandatory sharing. In each case, we averaged the ratings of the two non-mandatory

options and subtracted 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 -1 for the reverse.

Table 1 shows the results. From start to finish, the sample preferred expanding airport service to ending it, as indicated by mean differences well above zero. But that attitude faded significantly by the midterm (T2) measurement following the first round of on-site deliberation, resurging only insignificantly thereafter. Over the course of the experiment, the mean decreased from .540 to .434. On revenue sharing, the sample initially preferred local control to non-mandatory sharing and the latter to mandatory sharing, but reversed the first preference and

(Table 1 about here)

strengthened the second as the experiment proceeded. By the end, they distinctly preferred non-mandatory sharing or to either local control or mandatory sharing. There are significant changes in this direction from both T1 to T2 and T2 to T3.[15] These results are consistent with those of previous Deliberative Polls, which have shown statistically significant net attitude change more often than not.

The present question, however, is the extent to which this sort of change results from the formal on-site deliberations rather than other aspects of the larger treatment. Again, the contrast between the two treatment groups over the T1 to T2 interval is particularly revealing. Consider first revenue-sharing. The mean attitude shifts dramatically from local control toward non-mandatory sharing in the R-first group, discussing the issue during this interval, but scarcely budges in the A-first group, discussing the airport instead. The change is .313 in the A-first group, only -.012 in the R-first. The difference is highly significant ($p < .001$). When the question is non-mandatory versus mandatory sharing, the comparison is fainter but similar.

From T1 to T2, the R-first group moves twice as far toward non-mandatory sharing, although in this case the change is not quite significant ($p = .115$) in the R-first group and insignificantly greater in the R-first than the A-first group. In all, these results suggest that the on-site deliberations drove most of the attitude change on revenue sharing.

The T2-T3 comparison reinforces the inference. Here it is the A-first group, now discussing revenue sharing, that moves furthest toward voluntary sharing or incentives. 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 .318 in the A-first group, .090 in the R-first group; when it is non-mandatory versus mandatory sharing, the figures are .221 and .097. In both cases, the difference is highly significant ($p < .001$, $p = .007$).

The attitude changes on the airport tell a somewhat different story. In the first place, there is less overall net change, of only .106, compared to .335 and .220 on the two revenue-sharing indices. From T1 to T2, both treatment groups shift toward ending service. Here too it is the group discussing the issue that changes noticeably more (.168 versus .085), and the change is significant in the A-first group, discussing the airport ($p = .027$), but not quite in the R-first group, discussing revenue-sharing ($p = .175$). The difference, however, is not statistically significant ($p = .392$), and neither treatment group shows any real change from T2 to T3.

These results may suggest some slight effect of the on-site deliberation, although the effect is much smaller and the suggestion much more diffident than in case of revenue sharing. The change *is* larger and more significant from T1 to T2 in the A-first group. On the other hand, the difference between the groups is insignificant. More certainly, the results suggest some effect of the at-home deliberations before the deliberative weekend. That fits with the noticeable

and relatively similar changes in both treatment groups from T1 to T2, the interval containing the at-home deliberations, and the absence of change, in either group, from T2 to T3. That appears to be the bulk of the story on this issue. We speculate about the reason for the difference between the two issues below.

Proximity to Single-Peakedness

Our second hypothesis is that deliberation tends to increase proximity to single-peakedness, defined, as above, as $S = m/n$. Note that the identity of the dimension on which the numerator is premised 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 separately at all three measurements.[16] The results mostly echo those on attitudes. Overall—across the whole sample from T1 to T3—both issues show increased proximity to single-peakedness. The increase is modest (only from .77 to .81) on the airport but

(Table 2 about here)

dramatic on revenue-sharing (from .52 to .80). On the airport, the increase occurs entirely from T1 to T2 (when *S* increases from .77 to .84). There is actually a slight recession (from .84 back to .81) from T2 to T3.[17] On revenue-sharing, *S* increases both from T1 to T2 (when *S* goes from .52 to .70) and again from T2 to T3 (when *S* goes from .70 to .80).[18]

Comparing the treatment groups again suggests that the on-site deliberation had little if any effect on the airport. The changes in *S* in both treatment groups are minor. On revenue-sharing, however, the on-site deliberation appears to have had a profound effect. From T1 to T2, *S* increased by .21 in the R-first group, then deliberating the issue, but only by .07 in the A-first

group, then deliberating the airport. From T2 to T3, it increased by .21 in the A-first group but decreased by .02 in the R-first group.[19]

Absent *S*'s sampling distribution, we refrain from assertions of statistical "significance or "insignificance." The bootstrapped standard errors, moreover, are for *S*, not for 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. Still, these estimated standard errors make it hard to imagine that the increases in proximity to single-peakedness on revenue-sharing in the R-first group from T1 to T2 and in the A-first group from T2 to T3 and or that the differences between the two groups over each interval are not significant.

Discussion

These results demonstrate experimentally that the formal on-site deliberations at least sometimes account for a substantial portion of the attitude change and increased proximity to single-peakedness produced by Deliberative Polling. This in turn strengthens the inference, based heretofore only on statistical associations with information gains, that these before-after changes are substantially driven by the deliberative content of Deliberative Polling. It also strengthens the logically prior inference, based heretofore on contrasts with quasi control groups and the usual absence of anything noticeable that could account for parallel, contemporaneous changes in the wider public, that the before-after changes are real. At least on some issues, serious, balanced deliberation with diverse conversational partners does seem to produce both net attitude change and increased proximity to single-peakedness.

The results also suggest, however, that these effects may sometimes be rooted elsewhere. On revenue-sharing, most of the very large effect seems to stem from the on-site deliberation. But on the airport, most of the more limited effect appears to occur during the anticipatory period

between the invitation and the weekend. Some of this off-site effect may stem from isolated perusal of the briefing materials. But some, we suspect much, of it must also stem from the casual and generally imbalanced conversations with friends, family, and coworkers, stirred by the prospect of the deliberative weekend. This off-site effect may therefore also be deliberative.

What accounts for the difference between the airport and revenue-sharing issues? The most obvious and, we should guess, most important factor is their pre-deliberation salience. Hot-button issues have already received a great deal of attention and real-world deliberation. If the real-world deliberation is sufficiently good, most people may already be near their full-information positions; if it is sufficiently faulty, many people may be entrenched far from them. In either case, there is less room for a Deliberative Poll to have much effect (although, in the latter case, it might, if it lasted far longer than a weekend). Thus previous Deliberative Polls have tended to show smaller net attitude change and smaller increases in proximity to single-peakedness for more salient issues. The tendency is particularly strong for proximity to single-peakedness (see List, Luskin, Fishkin, and McLean 2006).

In New Haven, the airport was far more salient than revenue-sharing. In the year preceding the Deliberative Poll, the region's most widely circulated daily newspaper mentioned revenue sharing only seven times but ran 74 articles on airport expansion, 13 of them during the two months immediately preceding the Poll (excluding the coverage of the Poll itself).[20] This coverage, along with editorials and letters from citizens, spanned the full range of commonly held views about airport expansion and maintenance. At the time of the first interview, therefore, the public—our participants included—had presumably deliberated more about the airport than about revenue-sharing. Their attitudes were presumably more firmly rooted, their preferences closer to single-peaked. We note in this connection that *S* is only .52 for revenue-

sharing but .77 for the airport[21] and that the overall net attitude change was only .106 on the airport index versus .335 and .220 on the two revenue-sharing indices.

This explanation finds further support in the pattern of factual information gains. At all three measurements, we asked the participants to say (1) whether the region's population was closest to 250,000, 350,000, 550,000, or 750,000; (2) whether its rate of job growth during the 1990s was more than, about the same as, or less than in the rest of the United States; (3) whether 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; (4) whether New Haven's population increased, decreased, or did not change during the 1990's; (5) whether state law allows communities to share property tax revenues; (6) whether those communities with the most valuable property tend to have the lowest, average, or the highest property tax rates; (7) whether the Federal Aviation Authority classifies the regional airport as a major hub, a medium hub, a minor hub, or not a hub; and (8) whether 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) property taxes, (4) decreased, (5) yes, (6) the lowest, (7) a non-hub, and (8) yes. Items (7) and (8) are specifically relevant to airport expansion, items (5) and (6) specifically relevant to revenue sharing, and items (1) through (4) 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. As in previous Deliberative Polls, the participants

(Table 3 about here)

absorbed a great deal of factual information. For the whole sample across all eight items, the average percentage answering correctly increased by 22.3%, from 36.8% to 59.1%. The largest

gains on this overall measure occur from T1 to T2, doubtless because that is the much longer interval and because it brackets not only the first on-site deliberative session but the receipt of the briefing materials and the anticipatory, off-site deliberations with friends, family, and coworkers.

We present the issue-specific results with some diffidence, since the airport and revenue-sharing indices consist of only two items apiece, and any comparison must rest on a rather shaky assumption of equal average difficulty. Still, the differences are generally sizable, and the pattern extremely satisfying. For the whole sample, revenue-sharing information increased by 34.5%, general information by 19.5%, and airport information by only 15.5%. From T1 to T2, both treatment groups gained more information on the issue they deliberated then but also gained on the other issue, undoubtedly because the briefing materials and anticipatory, off-site deliberations covered both issues. The A-first group gained 15.6% on the airport and 6.3% on revenue sharing, while the R-first group gained 40.4% on revenue-sharing and 8.8% on the airport. Both groups also gained, and to similar degree (19.1% and 19.9%), on regional politics more generally. From T2 to T3, however, each treatment group showed significant information gains *only* on the issue it deliberated during that interval. The A-first group gained 18.8% on revenue sharing, and the R-first group 5.9% on the airport. Between T1 and T2, in short, the participants learn a good deal (presumably before arriving on-site) about all the topics of deliberation and a good deal more (presumably after arriving) about the topic they deliberate during the first deliberative session. Between T2 and T3 they learn still more but only about the topic they deliberate during the second deliberative session. The on-site learning does seem to be deliberation-based.

These information gains line up nicely with the net attitude change and increase in proximity to single-peakedness on revenue-sharing. From T1 to T2, the A-first group learned a good deal about revenue sharing and changed its views a good deal. It also showed a modest increase in proximity to single-peakedness. The R-first group learned 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, learned a good deal on top of what it had learned 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 learned only a little, changed its views only a little, and showed no increase in proximity to single-peakedness. (Compare Tables 3 with Tables 1 and 2.)

This explanation receives additional support from the issue-specific information measures in Table 3. Again acknowledging that each measure rests on only two items, we note that the participants seem to have been much better informed about the airport than about revenue-sharing at T1, before they had ever heard of the Deliberative Poll, answering 25.4% of the revenue-sharing information items but 38.6% of the airport information items correctly. The revenue sharing information items do not appear simply to have been intrinsically harder, given that by T3 the participants actually fared slightly better on them (59.9% correct, versus 54.2% on the airport items).[22] Rather, the participants would appear to have entered the Deliberative Poll with better informed—and hence less easily changed and more highly structured—preferences about the airport than about revenue sharing.

A broad residual question is how, in two distinct senses, to apportion the on-site deliberation's effect. The first sense is more operational, a matter of disaggregating the Deliberative Polling experience. Recall that each deliberative session consisted of both small-

group discussion and plenary questions and answers with panels of policy experts or policy makers. To what extent did the changes result from the one versus the other? Also recall that the T2-T3 interval encompassed two deliberative sessions: the Saturday afternoon session on the second issue (revenue sharing for the A-first group, the airport for the R-first group) and the synoptic Sunday morning session on both issues. To what extent did the T2-T3 changes result from the deliberations on the second issue versus the ensuing deliberations on both issues? Even seemingly small features of the design may have effects worth trying to isolate. The requirement that each small group agree on one or more questions to pose to the expert panels may increase single-peakedness, for example.

The second sense is more theoretical, a matter of distinguishing the social and psychological mechanisms by which deliberation may affect attitudes and proximity to single-peakedness. How far do the changes stem from the sheer quantity of information acquired, from the degree to which it is balanced, from the social properties or intellectual content of small group discussions, from the attitudinal or socio-demographic heterogeneity of one's fellow discussants? How far do they depend on reading versus listening? On solitary versus social processes more generally? On the adoption of "empirical premises," which are debatable, versus the absorption of "facts," which are not (Luskin, Fishkin, and Jowell 2002)? On learning versus thinking? On empathy or identification versus argument? It is already clear, from both earlier statistical analyses (Luskin, Fishkin, and Jowell 2002) and the experimental results above, that the quantity of factual information absorbed plays a major intervening role, but much more remains to be explored.[23]

The answers to these empirical questions bear on more normative ones. To what extent is deliberation subject to the "pathologies" cited by critics of deliberative democracy (Stokes 1998,

Mendelberg 2001)? To what extent do the substantive focus, balance, and heterogeneity built into the deliberations in Deliberative Polling protect against them?

All these questions need further exploration. We do not wish to anoint any single approach, but further randomized field experiments will help—albeit at some cost in public consultation. An experimental manipulation that gave different subsets of participants more substantively different experiences, as of discussing different issues or discussing given issues differently, would be exposing only some (if indeed any) of the sample to the optimal treatment. The *n* available for estimating deliberative public opinion, as distinct from the effects of variations in the deliberative experience would be radically diminished. The beauty of the split-half design is that it folds a randomized experiment into a Deliberative Poll without sacrificing the latter's public mission or recommending force. Everyone eventually deliberates both issues. The only variation is in the timing.

The present findings are an important beginning. The American Founders believed that only an elected elite would be capable of informed deliberation about how best to reconcile particular interests with the welfare of the nation as a whole. U.S. Senators and the Electors in the Electoral College were once mainly elected by state legislatures. But such Madisonian institutions of deliberative "filtration" have gradually given way to formal and informal inputs of "unfiltered" public opinion (Fishkin 1997). The vestigial state-by-state aggregation in the Electoral College aside, elections are now mainly direct. Ballot initiatives and referenda have grown increasingly common. Nominations are generally decided in direct primary elections or by open, primary-like caucuses. With the advent of modern polling, elected representatives are deliberating less and following the polls more. The same trends are evident around the democratic world.

But ordinary polls tend to register top-of-the-head, even phantom opinions, and the level of thought and information underlying many votes in elections and referenda is scarcely greater. Whether from Downsian "rational ignorance" (Downs 1957) or sheer uncalculated obliviousness, not many people know, think, or talk about politics very much (Delli Carpini and Keeter 1996, Kinder 1998, Price 1988, Luskin 2002). When they do talk about politics, people tend to choose conversational partners, topics, and specific sources of information to minimize disagreement (Mutz and Martin 2001). Mostly, therefore, they talk with others very like themselves, expressing opinions very like their own (Kinder 1998). And what they *hear*, given "confirmatory" and similar psychological biases (Higgins and Bargh 1987), is still more like what they already think.

Can the public do better? The results here confirm that opportunities for serious, balanced discussion with heterogeneous fellow citizens can permit people to learn about the issues, weigh competing arguments, reflecting on their values in relation to given policies, reconsider their views, and arrive at a shared understanding of the criteria for preferring one alternative to another. The resulting preferences, much more than those in conventional polls or elections, command respect. They are the opinions of people who, much more than usual, know what they are supporting or opposing and why. They are the unfiltered deliberative opinions of representative citizens.

These results should hearten reformers interested in designing fully democratic, participatory, and effective public consultation (Crosby 1995, Crosby, Kelly and Shaefer 1986, Fishkin 1997, Gastil 2000, and Leib 2002). The results demonstrate the power of even relatively small doses of serious, balanced deliberation with a wide assortment of fellow citizens. More sustained deliberation of this kind can be expected to do still more.

References

Arrow, Kenneth. 1953. *Social Choice and Individual Values*. New York: Wiley. 1953.

Black, Duncan. 1948. "On the Rationale of Group Decision-Making," *Journal of Political Economy* 56: 23-34.

Brams, Steven J., Michael A. Jones, D. Marc Kilgour. 2002. "Single-Peakedness and Disconnected Coalitions", *Journal of Theoretical Politics*, 14(3): 359-383.

Condorcet, Marie Jean Antoine Nicolas Caritat (the Marquis de). 1785. *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*, Paris.

Converse, Phillip. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. David E. Apter. New York: Free Press.

Crosby, Ned. 1995. "Citizen Juries: One Solution for Difficult Environmental Questions." *In Fairness and Competence In Citizen Participation: Evaluating Models for Environmental Discourse*, ed. Ortwin Renn, Thomas Webler, and Peter Wiedemann.

Crosby, Ned, Janey M. Kelly, and Paul Schaefer. 1986. "Citizen Panels: A New Approach to Citizen Participation." *Public Administration Review* (46): 170.

Delli Carpini, Michael X., and Keeter, Scott. 1996. *What Americans Know About Politics and Why it Matters*. New Haven: Yale University Press.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper.

Dryzek, John and Christian List. 2003. "Social Choice Theory and Deliberative Democracy: A Reconciliation", *British Journal of Political Science*, 33 (1): 1-28).

Elster, Jon. 1998. "Introduction", in *Deliberative Democracy,* ed Jon Elster. New York: Cambridge University Press, 1-18.

Fishkin, James S. 1997. *The Voice of the People: Public Opinion and Democracy.* New Haven

and London: Yale University Press, expanded paperback edition.

Fishkin, James S., and Robert C. Luskin. 1999. "Bringing Deliberation to the Democratic Dialogue: The NIC and Beyond" in *A Poll with a Human Face: The National Issues Convention Experiment in Political Communication*, ed Maxwell McCombs and Amy Reynolds. Mahwah, NJ: Lawrence Erlbaum.

Gaertner, Samuel L., John F. Dovidio, Mary C. Rust, Jason A. Nier, Brenda S. Banker, Christine M. Ward, Gary Mottola, and Missy Houlette. 1999. "Reducing Intergroup Bias: Elements of Intergroup Cooperation." *Journal of Personality and Social Psychology* 76(March): 388-402.

Gastil, John. 2000. *By Popular Demand: Revitalizing Representative Democracy Through Deliberative Elections.* Berkeley, Los Angeles, and London: University of California Press.

Gilens, Martin. 2001. "Political Ignorance and Collective Policy Preferences." *American Political Science Review* 95(June): 379-396.

Higgins, E. Tori, and John Bargh. 1987. "Social cognition and social perception." In *Annual review of psychology*, vol 38, ed. Mark Rosenzweig and Lyman W. Porter: 369-425.

Holmes, Susan and Christian List. 2002. "Partial Single-Peakedness and the Probability of Cycles: A Correlation." Working paper.

Kinder, Donald R. 1998. "Opinion and Action in the Realm of Politics." In *The Handbook of Social Psychology*, vol 4, ed. Daniel Gilbert, Susan T. Fiske, and Gardener Lindzey. New York: McGraw-Hill.

Knight, Jack and James Johnson. 1994. "Aggregation and Deliberation: On the Possibility of Democratic Legitimacy", *Political Theory* 22: 277-296.

Leib, Ethan J. 2002. "Towards a Practice of Deliberative Democracy: A Proposal for a Popular Branch," *Rutgers Law Journal* 33: 359-455.

List, Christian. 2002. "Two Concepts of Agreement." *PEGS: The Good Society* 11(1): 72-79.

List, Christian, Robert C. Luskin, James S. Fishkin, and Iain McLean. 2006. " Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls." Manuscript, London School of Economics.

Luskin, Robert C. 2003. "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.

Luskin, Robert C. 2002. "From Denial to Extenuation (and Finally Beyond): Political Sophistication and Citizen Performance." In James H. Kuklinski (ed.), *Thinking about Political Psychology*. New York: Cambridge University Press.

Luskin, Robert C. and James S. Fishkin. 1998. "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, St Louis.

Luskin, Robert C., James S. Fishkin, Roger Jowell, and Allison Park. 1999. "Learning and Voting in Britain: Insights from the Deliberative Poll" Paper presented at the annual meeting of the American Political Science Association, Atlanta.

Luskin, Robert C., James S. Fishkin, Dennis L. Plane. 1999. "Deliberative Polling and Policy Outcomes: Electric Utility Issues in Texas." Paper presented at the annual meeting of the Association for Public Policy Analysis and Management, Washington DC.

Luskin, Robert C., James S. Fishkin, Ian McAllister, John Higley, and Pamela Ryan. 2000. "Information Effects in Referendum Voting: Evidence from the Australian Deliberative

Poll." Paper presented at the annual meeting of the American Political Science Association, Washington DC.

Luskin, Robert C., James S. Fishkin, and Roger Jowell. 2002. "Considered Opinions: Deliberative Polling in Britain." British Journal of Political Science 32(July): 455-487.

Manin, Bernard. 1997. *The Principles of Representative Government*. Cambridge: Cambridge University Press.

McKelvey, Richard. 1979. "General Conditions for Global Intransitivities in Formal Voting Models." *Econometrica* 47: 1085-111.

Mendleberg, Tali. 2001. "The Deliberative Citizen: Theory and Evidence." Forthcoming in *Research in Micropolitics*, *Volume 6*, edited by Michael Delli Carpini, Leonie Huddy, and Robert Y. Shapiro, Elsevier Press.

Miller, David. 1992. "Deliberative Democracy and Social Choice", *Political Studies* 40 (special issue): 54-67.

Mutz, Diana C. and Paul S. Martin. 2001. Facilitating Communication across Lines of Political Difference: The Role of Mass Media. *American Political Science Review*, 95 (1): 97-114.

Niemi, Richard G. 1969. "Majority Decision-Making with Partial Unidimensionality," *American Political Science Review* 63: 488-497.

Orbell, John M., Alphons van de Kragt, and Robyn M. Dawes. 1988. "Explaining Discussion-Induced Cooperation. *Journal of Personality and Social Psychology* 54: 811-819.

Price, Vincent. 1988. "On the Public Aspects of Opinion - Linking Levels of Analysis in Public-Opinion Research." *Communication Research* 15(December):659-679.

Price, Vincent, and Neijens, Peter. 1997. "Opinion quality in public opinion research."

International Journal of Public Opinion Research 9(Winter):336-360.

Price, Vincent, and Neijens, Peter. 1998. "Deliberative Polls: Toward Improved Measures of "Informed" Public Opinion?" *International Journal of Public Opinion Research* 10(Summer):145-176.

Riker, William H. 1982. *Liberalism against Populism*. San Franscisco (W. H. Freeman).

Sally, D. 1995. "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992." *Rationality and Society,* 7(January): 58-92.

Schofield, Norman. 1976. "Instability of Simple Dynamic Games", *Review of Economic Studies* 45: 575-594.

Stokes, Susan. 1998. "Pathologies of Deliberation." In Jon Elster, ed. *Deliberative Democracy*, Cambridge UP: 123-139

Van Mill, David. 1996. 'The Possibility of Rational Outcomes from Democratic Discourse and Procedures', *Journal of Politics,* 58: 734-52.

Table 1

A. The Airport: Ending vs. Expanding Service

	T1	T2	T3	T2-T1	Sig.	T3-T2	Sig.	T3-T1	Sig.
A-First (*n* = 64)	.500	.332	.336	.168 (.074)	.027	-.004 (.072)	.957	.164 (.078)	.040
R-First (*n* = 68)	.577	.493	.526	.085 (.062)	.175	-.033 (.074)	.655	.051 (.072)	.477
Whole Sample (*n* = 132)	.540	.415	.434	.125 (.048)	.010	-.019 (.052)	.714	.106 (.053)	.048

B. Revenue Sharing: Voluntary Sharing or Incentives vs. Local Control

	T1	T2	T3	T2-T1	Sig.	T3-T2	Sig.	T3-T1	Sig.
A-First (*n* = 64)	-.074	-.086	.232	-.012 (.049)	.812	.318 (.049)	.000	.307 (.059)	.000
R-First (*n* = 68)	-.153	.160	.250	.313 (.060)	.000	.090 (.043)	.042	.403 (.068)	.000
Whole Sample (*n* =132)	-.115	.041	.241	.155 (.040)	.000	.201 (.034)	.000	.356 (.045)	.000

C. Revenue Sharing: Voluntary Sharing or Incentives vs. Mandatory Sharing

	T1	T2	T3	T2-T1	Sig.	T3-T2	Sig.	T3-T1	Sig.
A-First (*n* = 64)	.172	.211	.432	.039 (.050)	.434	.221 (.059)	.000	.260 (.044)	.000
R-First (*n* = 68)	.160	.244	.342	.085 (.053)	.115	.097 (.032)	.004	.182 (.048)	.000
Whole Sample (*n* = 132)	.166	.228	.385	.063 (.036)	.087	.157 (.033)	.000	.220 (.033)	.000

NOTE: Standard errors in parentheses; *p*-values are two-tailed.
DKs at midpoint

Table 2

Changes in Single-Peakedness*

A. Whole Sample

	n	T1	T2	T3
Airport	132	0.77 (.033)	0.84 (.031)	0.81 (.032)
Revenue-Sharing	132	0.52 (.042)	0.70 (.037)	0.80 (.037)

B. Airport, by Treatment Group

	n	T1	T2	T3
A-First	64	0.80 (.043)	0.81 (.049)	0.86 (.049)
R-First	68	0.82 (.045)	0.88 (.049)	0.84 (.036)

C. Revenue-Sharing, by Treatment Group

	n	T1	T2	T3
A-First	64	0.56 (.060)	0.63 (.062)	0.84 (.044)
R-First	68	0.47 (.057)	0.78 (.052)	0.76 (.051)

*Standard errors (in parentheses) bootstrapped, as described in n. 17.

Table 3
Information Gains

A. Whole Sample, by Topic (*n* = 132)

	T1	T2	T3	T2-T1	T3-T2	T2-T1
Airport	.386	.508	.542	.121***	.034	.155***
Revenue-Sharing	.254	.492	.599	.239***	.106***	.345***
General	.417	.606	.612	.189***	.006	.195***
Overall	.368	.553	.591	.185***	.038**	.223***

B. Airport, by Treatment Group

	T1	T2	T3	T2-T1	T3-T2	T3-T1
A-First (*n* = 64)	.398	.555	.563	.156***	.008	.164***
R-First (*n* = 68)	.375	.463	.522	.088**	.059*	.147***

C. Revenue-Sharing, by Treatment Group

	T1	T2	T3	T2-T1	T3-T2	T3-T1
A-First (*n* = 64)	.273	.336	.523	.063*	.188***	.250***
R-First (*n* = 68)	.235	.640	.669	.404***	.029	.434***

D. General, by Treatment Group

	T1	T2	T3	T2-T1	T3-T2	T3-T1
A-First (*n* = 64)	.398	.601	.590	.203***	-.012	.191***
R-First (*n* = 68)	.434	.610	.632	.176***	.022	.199***

Note: *P*-values based on one-tailed tests.

* $p < .10$ ** $p < .05$ *** $p < .01$

Appendix A
Demographic Comparisons of Participants, Nonparticipants, and Voting Population

	Participants (n = 132)	Nonparticipants (n = 1024)	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%
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%[b]
High school graduate	9.1%	20.2%	30.8%
Some college	25%	21.1%	18.2%
College graduate	28%	29.1%	21.7%
Trade/Technical	3%	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%[c]
Caucasian	72.7%	75.3%	78.1%
Hispanic or Latino/a	3%	4.4%	9.6%
Other	9.1%	9.1%	4%
Gender			
Male	50.8%	47.2%	46.8%[c]
Female	49.2%	52.8%	53.2%

NOTE: percentages may not add up to 100 in nonparticipant blocks due to "refusal" category.
[a] For population15 and over.
[b] For population 25 and over.
[c]For population 18 and over.

NOTES

*Earlier versions of this paper were presented at the annual meeting of the International Society of Political Psychology, Boston, MA, July 6-9, 2003 and at the ECPR General Conference, Marburg, Germany, September, 18-21, 2003. We 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 Carnegie Corporation, the Ford Foundation, the Renee B. Fisher Foundation, Fannie Mae, William C. Graustein, New Haven Savings Bank, the South Central Connecticut Regional Water Authority, the United Way of Greater New Haven, and 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 much of the 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.

[1]Topics have included the choices in a parliamentary election (in Britain), in referenda (in Australia and Denmark), and on such policy issues as the future of the American family, how best to meet regional electricity needs, how to deal with crime, the future of Britain's National Health Service, and Britain's role in Europe.

[2]These findings also support the widely accepted propositions that many of the opinions in conventional polls are "top-of-the-head" (Converse 1964) and that political knowledge affects

policy and electoral preferences (Bartels 1996, Delli Carpini and Keeter 1996, Althaus 1997, Luskin and Globetti 1997, Gilens 2001).

[3]Admittedly, these comparisons still lack the full authority they would have with true random assignment.

[4]Barabas's (2005) analysis of much the same question involves broadly similar quasi control groups analysis (created in his case by matching rather than approximate randomization) but with a highly nonrandom participant sample, hence lesser external validity.

[5]To the extent that the first mechanism predominates, we should expect to see attitudes change so as to increase their predictability from sociodemographic variables, proxying interests; to the extent that the second predominates, they should change so as to decrease it (Luskin, Fishkin, and Jowell 2002, Luskin 2003).

[6]Spatial single-peakedness is a well-defined notion if and only if the alternatives are identified with points in a Euclidean space. In that case (which is not ours), an individual's preference ordering is *spatially single-peaked* if it is induced by the alternatives' Euclidean distance from that individual's most preferred point in the space, and a combination of preference orderings is *spatially single-peaked* if every individual preference ordering is. Spatial single-peakedness is sufficient (but not necessary) for single-peakedness in Black's and Arrow's sense if the space is one-dimensional but not even sufficient for it if the space is multi-dimensional. (On the distinction, see, e.g., Brams, Jones, and Kilgour 2002.)

[7]See also Niemi (1969). More precisely, let N be the set of n individuals, and X the set of k alternatives. Each individual $i \in N$ holds a preference ordering R_i over the alternatives in X which is reflexive, transitive and connected (i.e., it allows strict preferences as well as ties). We write xP_iy as an abbreviation for [xR_iy and not yR_ix]. A combination (n-tuple) of preference orderings

across the individuals in N is abbreviated $(R_i)_{i \in N}$. $(R_i)_{i \in N}$ is single-peaked if there exists a one-to-one function $\Omega : X \rightarrow \{1, 2, ..., k\}$ (representing a dimension) such that, for every triple of alternatives $x,y,z \in X$ and every individual $i \in N$, if $\Omega(x) < \Omega(y) < \Omega(z)$ or $\Omega(z) < \Omega(y) < \Omega(x)$, then xR_iy implies xP_iz. To define the proximity to single-peakedness of $(R_i)_{i \in N}$, let M be a maximal subset of N such that $(R_i)_{i \in M}$ (i.e., the combination of preference orderings across the individuals in M) is single-peaked, and let $S = m/n$. For details, see List, Luskin, Fishkin, and McLean (2006).

[8]Note that S expresses the cohesion or 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 (2006).

[9]This is also supported by unpublished computer simulations by Christian List and Susan Holmes.

[10]Cf. Van Mill (1996), countered by Dryzek and List (2003).

[11]This does *not* necessarily mean that deliberation will lead people to converge on some particular ranking, what List (2002) calls "agreement at a substantive level." In fact, the results in List, Luskin, Fishkin, and McLean (2006) suggest that it tends to *decrease* agreement of this sort. Our only claim here is that deliberation increases proximity to single-peakedness, an instance of what List (2002) calls "agreement at a meta-level."

[12]Ironically, it was Riker (1982, p. 128) who first noted that discussion might create "a common view of the political dimension" at issue, thereby preventing cycles. "If by reason of discussion, debate, civic education and political socialization, voters have a common view of the political dimension (as evidenced by single-peakedness)," he remarked, "then a transitive outcome is guaranteed."

[13]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.

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

[15]Counting one p value of .087 as "significant."

[16]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.

[17]The dimension along which the largest subsample is single-peaked remains the same throughout, ordering the alternatives above as [A2 A1 A3], and the Condorcet winner, throughout, is A2 (expanding service).

[18]The dimension along which the largest subsample is single-peaked remains the same throughout, ordering the alternatives above as [R1 R2 R3 R4]. 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.

[19]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.

[20]Property taxes for homeowners were at issue in the region, and a great deal of discussion was devoted to property tax relief for the elderly, disabled, and low income families. But regional revenue-sharing was only briefly mentioned in one editorial on the future of the region and as a low-profile issue in the platform of an unsuccessful local mayoral candidate.

[21]Although *S* is also to some degree a decreasing function of the number of alternatives, which is 4 for revenue-sharing but only 3 for the airport. See List, Luskin, Fishkin, and McLean (2006).

[22]This does not necessarily mean that there was more learning during the Deliberative Poll about revenue sharing than about airport expansion. Information indices like these are subject to ceiling effects—the participants answering both airport items correctly at T1 could not *show* any learning, but since great deal of literature in both psychology and communications research argues and finds that the information-rich tend to get information-richer, they were very likely learning a lot, unobservably. Accordingly, there even may have been *more* learning about the airport than about revenue sharing during the Deliberative Poll. We simply cannot see as much of it. For more extended discussion of this general issue, see Luskin, Fishkin, and Jowell (2002) and Luskin (2002).

[23]See Dryzek and List (2003) and Luskin (2003) for further discussion of possible mechanisms.