Considered Opinions on U.S. Foreign Policy:

Face-to-Face versus Online Deliberative Polling*

Robert C. Luskin, University of Texas at Austin

James S. Fishkin, Stanford University

Shanto Iyengar, Stanford University

May, 2004

Actual citizenries bear little resemblance to the democratic ideal. Some variation by individual, by issue, and by circumstances notwithstanding, not many people know or think much about politics (Converse 1964, Luskin 1987, Delli Carpini and Keeter 1996, Kinder 1998, Price 1999). As a rule, therefore, the public opinion revealed by everyday polling is neither particularly informed nor particularly thoughtful. But suppose people did know and think more about politics. What would public opinion be like then?

Deliberative Polling addresses this question by drawing and interviewing a random sample, providing its members with briefing documents laying out the arguments for and against policy alternatives, getting them to discuss the issues in small groups, giving them opportunities to question competing experts and policy-makers, and then gauging their opinions again (Fishkin xxxx; Luskin, Fishkin, and Jowell 2002). Typically, a separate random sample, answering the same questions at the end of the process, provides a control group.[1] The enterprise, therefore, is not only a form of public consultation providing a view of more considered public opinion but also and a form of social-scientific quasi-experimentation shedding light on deliberation's effects on information, attitudes, and behavior.[2]

Until very recently, the deliberation in Deliberative Polling has been entirely face-to-face. The small group discussions and questioning of policy experts and policy makers have taken place at a common site, typically over a weekend. Advances in information technology, however, now make it possible to implement Deliberative Polling online, provided a representative sample of the whole public, not just the online population, can be obtained.

Here we describe the first online Deliberative Poll, conducted in parallel with a traditional face-to-face Deliberative Poll, both about U.S. foreign policy on the eve of our war with Iraq. The online sample, from Knowledge Networks, is a true random sample (subject only

to the qualifications attaching to any "random sample" of human beings). So of course was the face-to-face sample. The discussions addressed the same issues and spanned approximately the same period.[3] The briefing documents were identical, the questionnaires almost so. Thus we can both describe and compare the results.

Our broad questions are two: How did deliberation affect public opinion on these issues, and how (if at all) do the online and face-to-face effects differ? How far, to begin with, are the changes of preference similar in thrust? How far do the two sorts of deliberation move their participants in the same direction and to the same degree? To what extent, moreover, are effects similar in "kind." The results from previous face-to-face Deliberative Polls suggest that deliberation often changes preferences (both individually and in the net), that it increases knowledge, and that it does the former at least partly on account of the latter (see, for example, Luskin, Fishkin, and Jowell 2002). They also suggest that opinions within discussion groups do not routinely become more polarized or more homogeneous. To what extent do these same patterns obtain when the deliberation is online? The answers begin to illuminate the difference between online and face-to-face deliberation, as well as the potential of online Deliberative Polling.

Online versus Face-to-Face Deliberative Polling

Online Deliberative Polling is an important innovation, with some significant practical advantages over the traditional face-to-face design. These include:

- *Cost*. Physically assembling a random sample for a weekend at a single site is both cost and labor intensive. The expenses, mounting into six, sometimes seven figures for national samples, include transportation, hotel accommodations, meals, and honoraria for participating. Online deliberations, by contrast, do not require

participants to leave their homes. Those initially lacking online access must be given computers, but a representative sample can still be recruited for a tiny fraction of the cost of transporting participants to a single location and lodging and feeding them there. This advantage, moreover, is likely to increase, as the proportion of the population already online and thus not needing to be given computers increases.

- *Duration*. Work, family, and other obligations may keep some prospective participants from being able to spend even a weekend away from home, and most would find more than a weekend impossible. For face-to-face deliberation, therefore, the late-Thursday-through-midday-Sunday span of the 1996 "National Issues Convention" is probably the limit.[4] Online deliberations can extend much longer, indeed without any obvious limit.
- *Lead Time*. Face-to-face Deliberative Polls require extensive logistics and preparation. Online versions can be organized much more quickly. This creates the possibility of deliberating about real-world events and decisions in something much closer to real time.[5]
- *Instrumentation*. Much more of what the participants are doing is visible—and readily recorded—online than face to face. Their resort to the briefing documents, reading of the answers from expert panelists, and the like can be recorded automatically. This opens up new possibilities for unpacking the deliberative treatment's effects.

Not every consideration necessarily points the same way, of course. One obvious but conditionally needless worry is representativeness. Access to technology remains closely tied to socio-economic standing, as the phrase "digital divide" suggests. As already noted, however, the digital divide can be overcome by providing free access, including free equipment, a strategy

pioneered by Knowledge Networks Inc., which offers random samples free online access in exchange for regular participation in market research and opinion surveys. This combination of offline random sampling and providing online access is a *sine qua non* for decent representation.

There are also differences in the social content of the deliberation. Online deliberations may use audio, even visual channels, but cannot achieve the same immediacy as face-to-face interactions. They are also confined to the formal sessions, whereas on-site deliberations spill over into causal conversations over meals and drinks. It is possible that the firsthand human contact in face-to-face events makes them more involving and that they convey certain kinds of information—about what other people quite different from oneself are like—more fully. On the other hand, such differences may not militate entirely in favor of face-to-face designs. Perhaps the more mediated interactions in online deliberation are emotionally cooler, more cognitive, and perhaps that is to the good.

The question, in the end, is empirical. Online designs have some obvious practical advantages, but what are we sacrificing in using them? Our data speak to this question.

Research Design

The subject of both Deliberative Polls was "America in the World." More concretely, the briefing materials and discussions focused on the broad issues of military intervention, promoting democracy, trade and economic relations with other countries, helping other countries with problems like poverty and AIDS, and preserving the global environment. The briefing materials were prepared by the National Issues Forums and the Kettering Foundation.

The face-to-face Deliberative Poll was conducted in collaboration with MacNeil/Lehrer Productions, which produced the event, and the Survey Research Center of the University of California at Berkeley, which drew the sample and conducted the initial interviews. The on-site

deliberations, termed the "National Issues Convention," took place in Philadelphia from Friday, January 10 through Sunday, January 12, 2003. The participants alternated between discussions in randomly assigned small groups led by trained moderators and putting questions to panels of policy experts and policy makers in plenary sessions. They then answered the same questions as when initially interviewed. At roughly the same time (the end of that weekend), the Program on International Policy Attitudes at the University of Maryland asked most of the same questions of a separate random sample, which thus provided a post-test-only control group.

The online Deliberative Poll was administered by the Political Communication Lab at Stanford University. The samples of 280 for the treatment group and 219 for the control group were supplied by Knowledge Networks. They were therefore already committed to participating in a series of online surveys, and those without computers had already received web TVs. We took Knowledge Networks' largess a step further, providing previously offline respondents with personal computers in return for their participation. Computer owners were given a cash incentive of $300 instead. There was also a control group, also from Knowledge Networks, which completed the same questionnaire both before and after the deliberations but did not deliberate.

After completing an initial online questionnaire, the participants deliberated twice a week, an hour at a time, for four weeks in randomly assigned small groups led by trained moderators. The deliberations began the week of December 9-13, 2002, and ended the week of Jan 7 - 16, 2003. The discussion, via Lotus Sametime software, was voice- rather than text-based, for two reasons. We feared the less literate would be daunted by the necessity of reading other participants' contributions and typing their own, and we believed voice would allow more of the affective bonding and mutual understanding that characteristically emerges in face-to-face

deliberations. The software permitted the participants to request and release the microphone and identified the speaker and the list of those wishing to speak. The *Online Newshour*, also a partner in the online experiment, relayed questions formed in the small groups to panels of competing experts and posted their answers on the *Online Newshour* web site.

In the main, then, the online design closely resembled the face-to-face design. The principal differences, apart from the online versus face-to-face nature of the discussions, were in the expert panelists, the methods of recruitment, the length and timing of the small group discussions, and the nature of the control groups. The panelists for the face-to-face event were gathered over a single weekend. For the online event, they were scattered over five weeks. In both cases, they were selected for balance and relevance by the *Newshour with Jim Lehrer*. The face-to-face participants were all sampled from scratch, whereas the online participants were drawn from Knowledge Networks' existing panels. The "elapsed time," from start to finish, of the small group discussion was longer online, while the unbroken stretches of "processing time" were much longer face-to-face. The face-to-face control group was only surveyed at the posttest, while the online control group was survey both before and after the experiment.

Data

The dependent variables for this study are policy attitudes, and the key explanatory variable is knowledge (which, shrugging off some subtle distinctions, we treat as synonymous with "information" and "sophistication"). The data afford a good many relevant items.

Policy Attitudes

We have used questions about what the U.S. should or should not do, should give greater or lesser priority to, or do more or less of to construct nine policy indices. Each index makes use of every item seeming to both measure the relevant concept and to cohere with the others

seeming to do so. In a couple of cases, the "index" is only trivially an index, consisting of just a single item. In one other case, it consists only of two items. We wish in these cases that the data afforded more relevant, coherent items, but they do not, and it is better to let these one or two items sing solos or duets than to cancel the concert or permit the cacophony of more disparate ensembles only loosely related to the underlying concepts or to one another. The largest number of items is ten.

The response categories are linearly scored on a [0, 1] scale, giving all the items a common metric. The indices, averaging the individual items composing them, thus also range from 0 to 1. Items that are "missing data"—either DK or NA— for a given respondent are excluded for that respondent.[6]

In a number of cases, we "pre-average" subsets of items prior to averaging the whole set (effectively giving the individual items in the pre-averaged subset less weight than the rest). We do this for batteries of items significantly reflecting attitudes besides the one being measured. An example is a battery about the priority that should be given to foreign aid, increased trade, and a number of other possibilities as ways of promoting democracy. One might value democracy but think little of any specific proposal for promoting it. These items reflect attitudes toward foreign aid, increased trade, etc, as well as toward the goal of promoting democracy. The *average* response, however, *across* all six asked-about ways of promoting democracy may be taken as a reasonable measure of the respondent's attitude toward the goal of promoting democracy.[7]

The indices and their component items are:

Environmental Protection. This pre-averages two items asking whether the respondent supports or opposes "requiring higher mileage from automobiles even if that means less

powerful automobile engines" and "requiring cleaner production of electricity, even if that means higher electricity rates" as ways of reducing greenhouse gases, then averages that with items asking what priority, on a scale from 0 to 10, should be accorded "protecting the global environment" as "as a long range foreign policy goal" and to what extent "global warming is not really a problem so there is no need to do anything about it" versus "a serious problem [about which] we need to act now (both initially on 0 to 10 scales). Chronbach's alpha is .78 in the face-to-face sample and .73 in the online sample.

Fighting Terrorism. This index pre-averages four items asking how much importance should the U.S. place on "encouraging more democracy in Middle East countries like Egypt and Saudi Arabia," "increased foreign aid to countries that may be breeding grounds for terrorism," "working with other countries to identify and combat terrorism," and "building up our intelligence capabilities" as "ways of reducing future terrorism directed against the U.S." The remaining ingredients are items asking what priority, again on a scale from 0 to 10, should be accorded "preventing the spread of weapons of mass destruction," "fighting terrorism," and "protecting the U.S. from attack" as "long range foreign policy goal[s]" and to "discouraging countries from trying to develop nuclear weapons" as a reason for "provid[ing] foreign aid to other countries." Cronbach's alpha is .82 face to face and .83 online.

Increasing Foreign Aid. This is a single item asking whether "the amount of money the U.S. is now devoting to foreign aid should be increased, reduced, or kept about the same." Some of the items in the next index, "global altruism," may be thought relevant, but empirically this item, with its explicit reference to spending and silence about what exactly the money may be spent on, stands apart.

Fighting Poverty and Suffering. This variable is also about foreign aid but specifically about foreign aid for food, medicine, and the alleviation of poverty. It is not explicitly about spending. The index averages the priorities the respondent would give, as "long range foreign policy goal[s]," to "providing food and medical help to poor countries" and "reducing world poverty," again on a scale from 0 to 10, and the pre-averages, first, of the priorities he or she would give to "reducing hunger and disease in poor countries" and "helping poor countries develop their economies" as reasons for "provid[ing] foreign aid to other countries" and, second, of the responses to questions asking whether the U.S., "as a global leader" should spend more money to help fight world hunger in developing countries" and "to help fight the AIDS epidemic in developing countries" versus "concentrate[ing] on dealing with problems at home first." Cronbach's Alpha is .84 and face-to-face and .82 online.

Protecting Human Rights. This is a single item asking what priority should be given to "protecting human rights in other countries."

Internationalism. This too is a single item, in this case asking how much one agrees or disagrees that "this country would be better off if we just stayed home and did not concern ourselves with problems in other areas of the world."

Multilateralism. Two of this index's ingredients are the differences between the strength with which the respondent supports or opposes American military action "with United Nations support" versus "acting alone" to "prevent mass killings in foreign countries" and to "prevent the spread of weapons of mass destruction to countries that might use them." These differences then averaged with questions asking how strongly the respondent agrees or disagrees that "the only way to solve environmental problems like global warming is through international agreements, requiring countries to work together"; how important it is to "work with other countries to

identify and combat terrorism"; whether it is best to "work with groups of countries through international institutions like the World Trade Organization," "to work with other countries one by one to establish agreements," or "to leave things as they are" in dealing with international trade; and whether "the U.S. acting by itself," "the U.S. acting with close allies like NATO," "the U.S. & its close allies acting through the United Nations," "the United Nations," or "nobody" should "take the lead" in "trying to resolve international conflicts" and in "providing foreign assistance to other countries."[8] Cronbach's Alpha is .63 face-to-face, .75 and online.

Democratization. This index averages items asking whether the respondent agrees more that "the U.S. should be promoting democracy in other countries" or that "how other countries are governed is not our concern," the priority that should be given to "helping newly democratic countries develop their democratic institutions" as a reason for "provid[ing] foreign aid to other countries," and the pre-average of the importance that should be placed each of six "possible ways the U.S. might promote democracy outside of the U.S." ("help[ing] with building democratic institutions," "increasing trade," "trade penalties for human rights violations," "foreign aid," "provid[ing] U.S. troops to help keep the peace," and "increased support for organizations like the Peace Corps that send Americans abroad to help other countries"). Cronbach's alpha is .77 face to face and .72 online.

Trade. This is a single item asking whether the U.S. should "repeal the North American Free Trade Agreement, called NAFTA," "leave NAFTA the way it is," "adopt agreements like NAFTA but with more Latin American countries."

Knowledge

Our data include eleven foreign policy knowledge items to gauge the extent of learning from participating in the Deliberative Poll. These run as follows:

The Democratic and Republican parties' positions on global warming (2 items). Respondents were asked to place both parties on a 0-to-10 scale from "global warming is not really a problem so there is no need to do anything about it" (0) to "it's a serious problem and we need to act now" (10). We score respondents as correct if and only if they place the Democrats to the "act now" side of the midpoint on the one item and as correct if and only if they place the Republicans to the "do nothing" side of the midpoint on the other one.[9]

Bush's position on foreign aid. The question asks whether President Bush wants to increase foreign aid, decrease it, or keep it the same. The correct response is that he wants to increase it.

Bush's position on international agreements to control greenhouse gases. This question asks whether President Bush supports or opposes recent international agreements to control greenhouse gases. The correct response is that he opposes them.

The percentage of the federal budget going to military spending. This question asks whether about 1 dollar, 5 dollars, 10 dollars, 20 dollars, or 30 dollars or more, out of every $100 in the federal budget, go to military spending. The correct answer is 20 dollars or more.

The percentage of the federal budge going to foreign aid. This question asks whether about 1 dollar, 5 dollars, 10 dollars, 20 dollars, or 30 dollars or more, out of every $100 in the federal budget, go to foreign aid. The correct answer is 1 dollar.

The percentage of U.S. goods and services sold abroad. This question asks whether 4 dollars, 8 dollars, 12 dollars, 24 dollars out of every 100 dollars in goods and services produced by the U.S. are sold to customers abroad. The correct answer is 12 dollars.

The incidence of AIDS in Africa. This question asks whether "fewer than 5," "about 10," "about 20," or "30 or more" of every 100 adults in "those African countries with the highest rates

of infection " have "AIDS or the AIDS virus"? The correct answer is "30 or more".

Whether the U.S. has a veto in the WTO. Respondents are asked if it is true or false that "The U.S. has a veto on World Trade Organization decisions." The correct answer is "false."

Whether the U.S. has a veto on the Security Council. Respondents are asked if it is true or false that "The U.S. has a veto on the United Nations Security Council." The correct answer is "true."

The causes of global warming. Respondents are asked whether global warming is "caused mostly by human activities like driving cars and burning fuel," "caused by naural changes in the climate," or "not occurring at all." The correct answer is human activities.

In each case, we score the correct answer as 1 and all other responses, including Don't-Know's (DKs), as 0.[10] We average the items so scored to form a knowledge index also ranging from 0 to 1. These items all measure domain-specific foreign policy knowledge but we also have two items measuring more general political knowledge, unfortunately present in the face-to-face questionnaire only:

The Democratic and Republican parties' positions on the liberal-conservative dimension (2 items). Respondents are asked to place the Democratic and Republican parties on a 0-10 scale, "where 0 is about as liberal as they come, 10 is about as conservative as they come and 5 is exactly in the middle." As with the placement items on global warming, we treat these as separate items. The correct answers put the Democrats to the liberal side of the midpoint and the Republicans to the conservative side of it.

Deliberation's Effects on Policy Attitudes

The first question concerns deliberation's effects on these policy attitudes. We have two sorts of estimates of deliberation's effect in each mode and thus also of the mode effect (the difference of these effects). The first revolves around change over time. The differences between the post-deliberation (T2) and pre-deliberation (T1) means for the face-to-face and online participants estimate the effects of deliberating in each mode, and the difference of those differences estimates the mode effect.

The results, in Table 1, show that deliberation affected many policy attitudes in both modes. In the face to face treatment, seven of nine policy attitude indices showed statistically

(Table 1 about here)

significant change. In the online treatment, six of the nine did so. On two issues in particular, the face to face sample changed dramatically. They became far more internationalist and came to favor far greater spending on foreign aid. The online sample also changed a great deal, and in the same direction, on foreign aid spending.

More modestly but still significantly, both face-to-face and online deliberation increased support for democratization, global development, and protecting human rights. Face-to-face but not online deliberation significantly increased support for anti-terrorism measures and multilateralism. Online but not face-to-face deliberation significantly increased support for environmental protection and free trade.

Generally speaking, as this description suggests, the online and face to face results are quite similar—indeed remarkably so, given the design differences previously noted and further, inescapable, innumerable differences of detail. These were parallel but different events. The samples were different. The moderators were largely different. The content and tenor of the

small group discussions were different. The expert panelists were different, as were the questions they fielded and answers they gave.

Unsurprisingly, therefore, there were statistically significant differences in the magnitude of the policy attitude change on eight of the nine indices. Yet the broad pattern of change was, as we say, strikingly similar. On four of the nine issues, there was statistically significant change in the same direction. On none of the nine was there statistically significant change in opposing directions.

The second sort of estimate of deliberation's effect in each mode and thus also of the mode effect revolves around the post-deliberation differences between treatment and control groups. The post-deliberation (T2) differences between the treatment and control group means estimates the effects of deliberating in each mode, and the difference of those differences estimates the mode effect.

This analysis is regrettably complicated by some differences in the control group questionnaires, which omitted a number of items in the face-to-face case. Indeed three of our nine indices can no longer be computed at all. Thus Table 2 presents the results just for the

(Tables 2 and 3 about here)

online mode, using all the items we used in Table 1, while Table 3 presents the results for both online and face-to-face modes side by side, confining both the treatment and control group in each mode to those items available for the face-to-face control group.

The picture here is similar. In both modes, there is a deliberative effect. Online, using all the items, there is a statistically significant difference between treatment and control group on six of the nine indices. The results with the truncated indices in Table 3 are somewhat attenuated but show differences that at least scrape conventional standards of significance on three of the

six available indices on line and on five of the six face-to-face. There are statistically significant differences of differences (mode effects) on only one of the six, and again the pattern of change is broadly similar. On three of the six, there is statistically significant difference in the same direction. On none are there statistically significant differences in opposite directions.

Something about diffs?

Knowledge Gains

The next question is, what is producing these attitude changes? We are particularly interested in the extent to which they are driven by the increases in information and thought the process is designed to stimulate. Thought is difficult to capture, but we can much more easily gauge information, specifically by the foreign policy knowledge items described above.

Regrettably, not quite all eleven foreign policy knowledge items were asked of everyone. Two items—about the Democratic and Republican parties' positions on global warming—were put only to the face-to-face treatment group, not to the face-to-face control group, nor to either the treatment nor the control group online. Two additional items—about Bush's posture toward global agreements to control greenhouse gases and about the causes of global climate change—were put to the online treatment and control groups as well as the face-to-face treatment group but not to the face-to-face control group. Thus only seven of the all eleven foreign policy knowledge items can be used to compare the face-to-face treatment group (who were asked all eleven) with the face-to-face control group (who were asked only the seven) at T2, and only the nine put to both the online treatment and online control group can be used to compare them at T2 or to gauge the online information gain from T1 to T2.

We follow the same two strategies as in examining deliberation's effects on policy attitudes. First we look at the changes—the knowledge gains—from T1 to T2, then at the differences between the treatment and control groups at T2. Table 4 presents the participants'

(Table 4 about here)

knowledge gains—first, in Table 4A, for just the face-to-face participants, using all eleven knowledge items, then, in Table 4B, for both the online and face-to-face participants, using only the nine foreign policy knowledge items in the online questionnaire. Table 4B also shows the difference of differences, estimating the mode effect.

As can been seen, the participants learned significantly. Whether based on all eleven foreign policy items or only on the eight asked online, the face-to-face participants averaged answering about 12% more of the items correctly after deliberating. The online gain, by contrast, was far more limited. The online participants emerged getting an average of only about 4% more of the foreign policy knowledge items correct. Even this more modest gain, however, was still statistically significant. The difference between the face-to-face and online gains was also statistically significant. The online treatment conveyed information, but the face-to-face treatment conveyed more.

Table 5 presents the contrasts between treatment and control groups—in Table 5A, for the online experiment only, using the nine foreign policy knowledge items available there; and in Table 5B, for both the online and face-to-face experiments, using the seven items available

(Table 5 about here)

for the face-to-face control group and thus for both treatment and control groups in both modes.

The results, in tune with those in Table 4, show statistically significant contrasts between the treatment and the control group in each mode. Depending on whether all nine items

available for the online comparison or only the seven also available for the face-to-face comparison are included in the reckoning, the online participants average getting somewhere between 6.3 and 8.7% more of the foreign policy knowledge items than the online control group members. The face-to-face participants average answering 14.6% more of the items correctly than the face-to-face control group members. The mode effect of 14.6 – 6.3 = 8.3% is sizable and significant. From this perspective, too, it seems, the participants learned online but learned more—indeed a great deal—face-to-face.

It may also be worth noting the effect on general political information, as measured by a two-item index averaging the items based on the placements of the Democratic and Republican parties on a liberal-conservative scale. Even though we made no effort to acquaint our participants with terms like "liberal" or "conservative" or the conventional assignments of the two parties to the sides of this dimension, our participants did in fact seem to acquire a much clearer sense of the parties' locations. The only contrasts we can form in this case are for the face-to-face experiment. There we find that the average percentage of our participants getting these two items right increased by 10.1% from T1 to T2 and that on average 18.0% more of the participants than of the control group got them right at T2.

Policy Attitude Change in the Small Groups

Another possible explanation for the policy attitude changes lies in the small groups. Two hypotheses suggest themselves. One is that opinions homogenize. Whether from blind conformism or as a function of the balance of arguments, the variance of opinions within the

(Table 6 about here)

small groups shrinks. Table 6 presents the proportions of the 24 face-to-face and 15 online small groups showing a decrease in within-group variance, both issue by issue and across all 9 issues.

The results show that the variance does more often than not diminish both face-to-face and online but slightly more often face-to-face than online (72.7 versus 63.7% of the time). The stronger social component of face-to-face deliberation may make it more centripetal. There is some notable variation across issues, especially online, in the extent to which the variance declines. On some issues, particularly foreign aid spending in the online condition, more groups show increases than decreases. The issues for which the decreases are most pervasive tend to be the same whether the discussion is face-to-face or online: U.S. security, multilateralism, and democracy. The members of given small groups came to agree distinctly more on these issues. So do the issues for which the decreases are least common: the environment and foreign aid spending. The members of given small groups did *not* come to agree distinctly more, indeed in some cases came to agree less, on these issues.

Another hypothesis, in keeping with Sunstein (2000) and the jury studies he draws on, is that the groups tend to "polarize," in the sense of becoming more extreme. A group's T2 opinions look like its T1 opinions, only more so. The mean opinion moves further out on the same side as it started on. To test this notion, we must specify the origin (point of reference), with respect to which the group can be said to be becoming more or less extreme. The most natural origin is the midpoint—.5 on a 0 to 1 scale. An arguably more appealing alternative, however, is the grand mean (of the whole participant sample). A group averaging .7 on the U.S. Security index counts as wanting to beef up security if the midpoint (.5) is the point of reference, but so is every group, since the grand mean is roughly .8 (both face-to-face and online). In relation to the grand mean, by contrast, a group whose mean is only .7 is relatively reserved about beefing up security.

Table 6 presents the proportions of the small groups becoming more polarized under both definitions. If the point of reference is the midpoint roughly two thirds of the small groups do polarize face-to-face, although only about half of them do so online. If the point of reference is the grand mean, however, only about 40% of the small groups become more polarized, either face-to-face or online. There is some modest tendency, at least with the midpoint as the point of reference, for the polarization to be greater face-to-face, again perhaps in keeping with the stronger social component of the face-to-face interactions. Again, too, there is some considerable variation by issue, particularly if the point of reference is the midpoint. Here, however, there is not much resemblance between the face to face and online modes in the issues showing polarization versus moderation, especially if the midpoint is the point of reference.

Since Sunnstein has made much of the alleged tendency of small group deliberations to be polarizing in his sense, presenting it as a drawback of deliberative democracy, it is worth underscoring that these results do relatively little to sustain his gloom. Moderated small group deliberations in which no decision, much less any unanimous decision is required, and with a moderator making sure that all the major arguments are aired and considered, are a reasonable if still distant approximation of the democratic ideal we are aiming at but quite different from juries. Balanced deliberation with the sole purpose of helping the participants clarify their own thinking does not seem to produce any strong polarization.

Knowledge Gains, Small Group Influences, and Policy Attitude Change

These last results suggest some modest small group influences on policy attitude change. We have also already seen that there were major information gains. It remains to combine these two sorts of influences in a more explicit, individual-level model of policy attitude change. This model is simply:

$$P_2 - P_1 = \gamma_0 + \gamma_1 I_2 + \gamma_2(P_1 - G_1) + u,$$

where P_1 and P_2 are the participant's policy attitudes before and after deliberation (at T1 and T2), I_2 is the level of knowledge he or she emerges with at T2, G_1 is the mean T1 attitude of the other members of his or her small group, and u is the customary disturbance or error term. We use observed T2 knowledge rather than the observed knowledge gain because it is a better measure of actual knowledge gain (see Luskin, Fishkin, and Jowell 2002, Luskin 2002). People who emerge with a lot of information have gained a lot—either observably, if they started off low at T1, or unobservably, if they started off high at T1.

Table 7 shows the ordinary least squares (OLS) estimates for both the online and face-to-face participant samples. The signs in parentheses beside the regressors' names are those of the mean attitude change in the whole sample, first online, then face-to-face. These are also the signs we normatively expect for the knowledge coefficients. If the attitude changes in whole sample are as we hope information driven, they should be greatest for those who learn and thus emerge knowing the most. That is, γ_1 should have the same sign as $\overline{P}_2 - \overline{P}_1$, the mean attitude change in the whole sample. To be sure, some of the attitude changes in the whole sample are statistically insignificant, and in these cases we may consider that there is no strong expectation as to the sign of γ_1. Table 7 denotes statistically significant attitude changes in the whole sample by double plus or minus signs. We also expect γ_2 to be negative, meaning that the participants to tend to narrow the initial gap between their own and their small group's position.[11]

The model appears to fit very well. The adjusted R^2s are respectable to large for models of this sort, confined to survey measures of psychological variables, involving only two

(Table 7 about here)

regressors, and having a change score as the dependent variable. None of the coefficient estimates carries a statistically significant but anomalously signed estimate. Though always good, the fit is better, sometimes much better, for the face-to-face sample.

The results also confirm the expectation regarding the small group coefficient in spades. The estimate is always negative, always significant. The more important regressor for us, however, is information. Online, there are six issues showing a statistically significant net attitude change in the whole sample and thus engendering a prediction about the sign of the information coefficient. Face-to-face, there are seven. In this critical respect the model fares much better in the online than in the face-to-face experiment. Face-to-face, only one of the seven expectations as to sign is met; online, five of six are. We hasten to add, since one-for-seven seems an extremely low batting average, that six of the seven face-to-face coefficient estimates for which we had an expectation were of the right sign; it was just that five of those six did no attain the conventional but arbitrary .05 level of significance.

But the small group coefficients in this model may give small group mechanisms too much credit. Both the small group regressor and the dependent variable include the participant's T1 attitude, the former with a positive sign, the latter with a negative one, which builds in a certain amout of negative covariance. What happens if we break the small group regressor apart into the participant's T1 attitude and the T1 mean of his or her small group, entered separately. We should then expect a negative sign on the coefficient of the participant's T1 attitude (indicating "regression toward the mean") and a positive one on the coefficient of the T1 mean of participant's small group. In other studies, we have also found that the lion's share of the composite small group variable's estimated effect is actually just regression toward the mean.

Table 8 shows the results of this divorce here, and the story is the one we have heard before. It is the same story both online and face to face. The information effects are scarcely affected, but the group mean variable is now significant only for one of the nine regressions in the online sample for none of the nine in the face-to-face sample. Information actually has the more important effect, by a wide margin online and a very slender one face-to-face.

Why the face to face opinion changes rested less firmly than the online opinion changes on learning is unclear, but one speculation revolves around our impression that the online and face to face deliberations were qualitatively different in focus and emotional charge. The latter happened to fall on the eve of the decision to go to war, and the discussion seemed to concentrate very heavily on that and related issues and to be relatively partisan. The former, spread over the preceding weeks, seemed to be less concentrated on the war and emotionally cooler. Since only two of the dozen items in the information index—the two asking what percentage of the federal budget goes for military spending and whether the U.S. has a veto on the U.N. Security Council—had anything at all to do with the war, military security, or multilateralism, the information captured by the index just may not have been as relevant for the face to face as for the online deliberations.

Conclusion

These results establish online Deliberative Polling as a viable process with significant potential both for improving practices of public consultation and for illuminating the role of deliberation in policy attitude formation and change. This experiment is only the initial launch of the process. But already we can see that the online and face-to-face deliberation appear to

produce mainly similar results. In both of the two parallel Deliberative Polls we examine, the participants became more informed and changed their views in a generally more internationalist direction. If anything, the participants' changes of attitude seemed *more* information driven online than face to face. In particular, the information based model, which we have used to explain change in other Deliberative Polls, worked extremely well in the online case—explaining change for five of the six indices for which there was significant change.

This project also provides an empirical platform for many other comparisons of online and face-to-face Deliberative Polls. Does online voice-based discussions allow for the same kind of mutual understanding and apparent "empathy" that we find in the face to face projects? Do they have the same positive effects on civic engagement? Where we use ranking questions, will we find that online Deliberative Polls depress the likelihood of voting cycles (as in List et al. 2000)? Such questions await further analyses and further projects.

In the meantime, we can say that the online Deliberative Poll offers a practical tool for public consultation. While deliberation's effects were generally somewhat smaller online than face-to-face, it is worth emphasizing that this was a maiden voyage and hardly exhausted the online possibilities. For instance, we arbitrarily limited the deliberations to eight hour-long sessions over four weeks (with the Christmas holidays forming a hiatus in the middle), but there is no reason online deliberations could not extend far longer. An online sample could continue deliberating for eight weeks or eight months. Eventually the cumulative effects of protracted enough online deliberation should surpass from a weekend of face-to-face discussion.

The online design's biggest cost, that of providing computers and internet access, will only diminish, as the digital divide narrows and access to computers become more and more like access to telephones. In the not too distant future, online Deliberative Polling can become a cost

effective but deliberative alternative to conventional polling and a truly widespread institution for improving public consultation. In this sense, the aspiration for online Deliberative Polling parallels the aspiration for conventional polling when it was young. Gallup thought that the public opinion poll might bring something like the New England town meeting to the large scale nation state.[12] This dream, we now know, was unrealistic for conventional polling, but online Deliberative Polling may have the potential to achieve it.

References

Sunstein, Cass R. 2000 "Deliberative Trouble? Why Groups Go to Extremes" Yale Law Journal, Vol 110, pp. 71-1119.

List Christian, McLean, Iain, Fishkin, James and Luskin, Robert C. 2000 "Deliberation, Preference Structuration and Cycles: Evidence from Deliberative Polls" Paper presented at the meetings of the American Political Science Association.

Luskin, Robert C., Fishkin, James S., and Roger Jowell. 2002. "Considered Opinions: Deliberative Polling in Britain." British Journal of Political Science, 32, 455-87.

Table 1
The Effects of Face-to-Face Versus Online Deliberation:
T1-T2 Differences, Participants Only

	T1		T2		T2 – T1		Mode Diff.
	Face-to-Face	***Online***	***Face-to-Face***	***Online***	***Face-to-Face***	***Online***	
Environment	.778	.686	.771	.715	-.007	.029***	.037***
	(.011)	(.014)	(.012)	(.014)	(.008)	(.010)	(.013)
U.S. Security	.800	.805	.821	.800	.021**	-.005	.026**
	(.009)	(.009)	(.008)	(.009)	(.007)	(.007)	(.010)
Human Rights	.701	.584	.727	.630	.026**	.046***	.020
	(.012)	(.015)	(.010)	(.013)	(.011)	(.013)	(.017)
Multilateralism	.733	.735	.786	.748	.053***	.012	.041***
	(.007)	(.009)	(.005)	(.008)	(.006)	(.008)	(.010)
Global Altruism	.589	.448	.683	.478	.094***	.030***	.064***
	(.010)	(.009)	(.009)	(.009)	(.008)	(.007)	(.011)
Internationalism	.744	.665	.849	.684	.105***	.019	.086***
	(.016)	(.019)	(.014)	(.019)	(.016)	(.019)	(.025)
Foreign Aid spending	.478	.312	.724	.411	.246***	.098***	.148***
	(.019)	(.024)	(.018)	(.026)	(.019)	(.021)	(.029)
Democracy	.633	.511	.687	.536	.054***	.025***	.029**
	(.011)	(.010)	(.009)	(.009)	(.009)	(.010)	(.014)
Trade	.492	.348	.478	.396	-.014	.047**	.061**
	(.016)	(.019)	(.014)	(.019)	(.018)	(.020)	(.027)

*p < .10; **p < .05; ***p < .01.

Table 2
T2 Differences between Treatment and Control Groups
(Complete Indices, Online Only)

	Participants		Control Group		Difference	
Environment	.715	(.014)	.648	(.016)	.067***	(.021)
U.S. Security	.800	(.009)	.784	(.011)	.016	(.014)
Human Rights	.630	(.013)	.583	(.016)	.046**	(.020)
Multilateralism	.748	(.008)	.720	(.010)	.028**	(.013)
Global Altruism	.478	(.009)	.426	(.010)	.051***	(.014)
Internationalism	.685	(.019)	.650	(.020)	.036	(.028)
Foreign Aid Spending	.414	(.024)	.289	(.024)	.125***	(.034)
Democracy	.535	(.009)	.489	(.011)	.047***	(.014)
Trade	.393	(.018)	.360	(.020)	.033	(.027)

*p < .10; **p < .05; ***p < .01.

Table 3
T2 Differences between Treatment and Control Groups
(Indices Confined to Shared Items, Both Modes)

	Participants		Control Group		Difference		Mode Diff.
	Face-to-Face	***Online***	***Face-to-Face***	***Online***	***Face-to-Face***	***Online***	
Environment	.744	.687	.712	.660	.031*	.027	.004
U.S. Security	.828	.813	.809	.796	.011	.017	.006
Human Rights	.727	.630	.665	.583	.062***	.046**	.016
Multilateralism	.777	.752	.739	.720	.039***	.032**	.007
Global Altruism	.723	.606	.642	.561	.080***	.045***	.035*
Internationalism	.848	.685	.682	.650	.167***	.036	.131***

Table 4
Knowledge Gains, T2 –T1

A. Face-to-Face Participants (All Items)

	T1	**T2**	**T2 – T1**
Knowledge	.373	.496	.123***
	(.009)	(.009)	(.009)

B. Online vs. Face-to-Face Participants (Shared Items)

	Face-to-Face			**Online**			**Mode Diff.**
	T1	***T2***	***T2 – T1***	***T1***	***T2***	***T2 – T1***	
Knowledge	.348	.467	.119***	.367	.406	.038***	.081***

*p < .10, ** p < .05, ***p < 01.

Table 5
T2 Knowledge Differences between Treatment and Control Groups

A. Online (All Nine Available Items)

	Online		
	Treatment	***Control***	***Diff.***
Knowledge	.406	.319	.087***

B. Online vs. Face-to-Face (Shared Items)

	Face-to-Face			Online			Difference
	Exp. Group	***Control Group***	***Diff.***	***Exp. Group***	***Control Group***	***Diff.***	
Knowledge	.427	.282	.146***	.354	.291	.063***	.083***

*p < .10; **p < .05; ***p < .01

Table 6
Small Group Behaviors

	Face-to-Face			Online		
	Moving further out from the *midpoint*	Moving further out from the *grand mean*	Decreasing variance	Moving further out from the *midpoint*	Moving further out from the *grand mean*	Decreasing variance
Environment	50.0	45.8	50.0	80.0	40.0	46.7%
U.S. security	75.0	37.5	75.0	40.0	26.7	100.0
Human Rights	58.3	25.0	58.3	93.3	40.0	66.7
Mulitlateralism	95.8	54.2	95.8	80.0	40.0	80.0
Global altruism	100.0	45.8	100.0	20.0	46.7	60.0
Internationalism	79.2	45.8	79.2	60.0	26.7	53.3
Foreign aid Spending	25.0	41.7	25.0	6.7	40.0	33.3
Democracy	79.2	37.5	79.2	40.0	40.0	80.0
Trade	25.0	29.2	25.0	6.7	53.3	46.7
Total	65.3	40.3	65.3	48.1	39.3	63.7

Note: Entries are percentages of the **xx** small groups.

Table 7
Attitude Change as a Function of t_2 Information and t_1 Distance from Group Mean

Policy Index		Online			Face-to-Face		
		Coeff.	**S. E.**	***p***	**Coeff**	**S. E.**	***p***
	Constant	-.017	.024	.46	-.066	.024	.01
Environment	t2 Info	.117	.053	.02	.125	.049	.01
(++, -)	t1 Att. Distance*	-.228	.040	.00	-.261	.038	.00
	Adj. R^2	.115			.124		
	N	245			340		
	Constant	.002	.016	.88	.023	.018	.20
U.S. Security	t2 Info	-.020	.036	.29	-.004	.035	.46
(-, ++)	t1 Att. Distance	-.253	.049	.00	-.432	.034	.00
	Adj. R^2	.094			.322		
	N	245			340		
	Constant	.044	.025	.08	.039	.026	.14
Human Rights	t2 Info	.005	.056	.47	-.028	.053	.30
(++, ++)	t1 Att. Distance	-.514	.044	.00	-.555	.038	.00
	Adj. R^2	.359			.380		
	N	244			340		
	Constant	-.004	.016	.80	.039	.014	.01
Multilateralism	t2 Info	.043	.036	.11	.029	.028	.15
(+, ++)	t1 Att. Distance	-.480	.045	.00	-.590	.034	.00
	Adj. R^2	.318			.470		
	N	245			340		
	Constant	.004	.015	.77	.079	.021	.00
Global Altruism	t2 Info	.063	.034	.03	.031	.043	.23
(++, ++)	t1 Att. Distance	-.322	.042	.00	-.389	.037	.00
	Adj. R^2	.191			.243		
	N	245			340		
	Constant	-.014	.041	.73	.071	.039	.07
Internationalism	t2 Info	.078	.092	.20	.070	.078	.19
(+, ++)	t1 Att. Distance	-.467	.057	.00	-.604	.044	.00
	Adj. R^2	.221			.363		
	N	237			334		
	Constant	.012	.050	.82	.033	.047	.48
Foreign Aid Spending	t2 Info	.208	.112	.03	.457	.094	.00
(++, ++)	t1 Att. Distance	-.270	.058	.00	-.566	.045	.00
	Adj. R^2	.101			.341		
	N	194			313		
	Constant	-.009	.019	.64	.042	.023	.07
Democracy	t2 Info	.084	.044	.03	.026	.047	.29
(++, ++)	t1 Att. Distance	-.505	.051	.00	-.486	.039	.00
	Adj. R^2	.284			.314		
	N	245			340		
	Constant	-.054	.046	.23	-.211	.039	.00
Trade	t2 Info	.238	.098	.01	.418	.078	.00
(++, -)	t1 Att. Distance	-.531	.057	.00	-.719	.045	.00
	Adj. R^2	.296			.458		
	N	216			327		

Table 8
Attitude Change as a Function of t_2 Information, t_1 Group Mean, and t_1 Attitude

Policy Index		Online			Face-to-Face		
		Coeff.	**S. E.**	***p***	**Coeff**	**S. E.**	***p***
Environment (++, -)	Constant	.143	.196	.47	.017	.123	.09
	t2 Info	.128	.055	.01	.135	.049	.00
	t1 Group Mean	.001	.279	.50	-.026	.151	.43
	t1 Att	-.243	.044	.00	-.285	.039	.00
	Adj. R^2		.114			.131	
	N		245			340	
U.S. Security (-, ++)	Constant	.315	.132	.02	.364	.129	.00
	t2 Info	-.028	.036	.22	-012	.035	.37
	t1 Group Mean	-.107	.159	.25	.041	.152	.39
	t1 Att.	-.278	.049	.00	-.463	.036	.00
	Adj. R^2		.110			.334	
	N		245			340	
Human Rights (++, ++)	Constant	.094	.158	.55	..394	.134	.00
	t2 Info	.005	.056	.45	-.012	..053	.41
	t1 Group Mean	.431	.259	.05	-.555	.074	.09
	t1 Att.	-.518	.046	.00	-.591	.040	.00
	Adj. R^2		.364			.392	
	N		244			340	
Multilateralism (+, ++)	Constant	.119	.230	.60	.376	.089	.00
	t2 Info	.044	.036	.10	.025	.028	.18
	t1 Group Mean	.322	.298	.14	.165	.116	.08
	t1 Att.	-.490	.049	.00	-.622	.034	.00
	Adj. R^2		.316			.490	
	N		245			340	
Global Altruism (++, ++)	Constant	.197	.086	.02	.314	.082	.00
	t2 Info	.069	.034	.02	.044	.043	.15
	t1 Group Mean	-.086	.186	.32	.012	.133	.46
	t1 Att.	-.348	.043	.00	-.420	.038	.00
	Adj. R^2		.204			.260	
	N		245			340	
Internationalism (+, ++)	Constant	.183	.130	.16	.583	.120	.00
	t2 Info	.104	.093	.13	.095	.076	.10
	t1 Group Mean	.172	.193	.19	-.048	.151	.38
	t1 Att.	-.483	.057	.00	-.653	.044	.00
	Adj. R^2		.226			.397	
	N		237			334	
Foreign Aid Spending (++, ++)	Constant	.104	.078	.19	.320	.100	.00
	t2 Info	.235	.113	.02	.456	.093	.00
	t1 Group Mean	-.038	.208	.43	.010	.178	.48
	t1 Att.	-.294	.060	.00	-.613	.047	.00
	Adj. R^2		.101			.362	
	N		194			313	

Table xx - *Continued*

Democracy (++, ++)	Constant	.240	.147	.11	.430	.089	..00
	t2 Info	.093	.044	.02	.025	.028	.18
	t1 Group Mean	.037	.279	.44	.165	.116	.08
	t1 Att.	-.531	.053	.00	-.622	.034	.00
	Adj. R^2		.290			.490	
	N		245			340	
Trade (++, -)	Constant	.015	.103	.88	.208	..086	.02
	t2 Info	.239	.098	.01	.390	.075	.00
	t1 Group Mean	.345	.254	.09	-.046	.147	.38
	t1 Att.	-.543	.059	.00	-.775	.044	.00
	Adj. R^2		.295			.502	
	N		216			327	

NOTES

*An earlier version of this paper was presented at the annual meeting of the American Political Science Association, Philadelphia, PA, August 27-31, 2003. We are grateful to to Kyu Hahn and Jennifer McGrady for research assistance, the Center for Deliberative Democracy at Stanford University for support, and **xxxx** for comments. The research was planned 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. Both the online and face-to-face Deliberative Polls received major financial support from the William and Flora Hewlett Foundation.

[1]The design differs from those of Citizen Juries, Consensus Conferences, and assorted other, vaguely similar deliberative fora in a number of important ways, but perhaps most critically in the use of sizable random samples. Deliberative Polling can claim to be showing the views of a more informed and thoughtful *public*, not just a narrow, self-selected stratum.

[2]Since the proceedings and results receive media coverage, it is also, less centrally, a vehicle for educating the public about policy issues and electoral choices, and it may also be seen as a demonstration project for making real-world democracy more deliberative (Ackerman and Fishkin 2004).

[3]The posttest survey in both experiments occurred during the same week in January of 2003.

[4]If participation were mandatory, like jury service, face-to-face gatherings could be convened for longer periods, but this is not a possibility we advocate at this point.

[5] In the case of the war in Iraq, for example, we were able to re-assemble a sample that who had deliberated about American foreign policy, including the possibility of war, in January of 2003 for a series of follow-up discussions in September of 2004.

[6]Equivalently, respondents receive their average score on the items they do answer.

[7]We have tried using these items without pre-averaging them, but the index generally coheres less well than with the pre-averaging.

[8]Nobody's taking the lead, chosen by only a few percent anyway, is treated as missing data.

[9]These two items could be folded down into one, based on the relative placements of the two parties (scoring respondents putting the Democrats to the "act now" side of the Republicans as correct and all other respondents as incorrect), but other research indicates that knowledge measures constructed from absolute placements—of each of each object (party, in this case) individually—fare better (Luskin, Cautrès, and Lowrance 2004, Luskin and Bullock 2004).

[10]Note that this treatment of DKs is deliberately at odds with Mondak's recommendation to treat DKs and incorrect answers as differently (effectively, by giving the latter part-credit), which other research shows to be ill-advised (Bennett 2001, Luskin and Bullock 2004).

[11]More precisely, the greater the (necessarily nonnegative) I_2, the larger the opinion change, and the more positive (negative) the time 1 difference between the individual and his or her group, $P_1 - G_1$, the more negative (positive), the opinion change.

[12]George Gallup "Public Opinion in a Democracy" (Princeton: the Stafford Little Lectures, 1938), p. 6.