The Effects of Deliberative Polling in an EU-wide Experiment: Five Mechanisms in Search of an Explanation

DAVID SANDERS[*]

Deliberative Polls simulate public opinion in a given policy domain when members of the relevant mass public are better informed about the issues involved. This article reports on the results of a three-day Deliberative Poll, conducted before the June 2009 European Parliament elections, to evaluate the effects of deliberation on a representative sample of EU citizens. Findings show that, compared with a control group, deliberators changed their views significantly on immigration (becoming more liberal), climate change (becoming greener) and the EU itself (becoming more pro-European). Five different explanations of why deliberation appears to work are tested: sampling bias, increased political knowledge, discussion quality, small group social conformity pressure and the influence of other Deliberative Poll actors, but none is satisfactory.

Deliberative Polls (DPs) aim to simulate what public opinion in a given policy domain would be if members of the relevant mass public were better informed about the issues involved. As such, DPs offer an important new mechanism for democratic practice, one that has been adopted in a variety of institutional contexts crossnationally.[1] The procedures of DPs vary according to specific circumstances, but the general principles involved are simple: (1) in the chosen issue domain, the state of opinion among the target mass public is assessed; (2) a representative sample of that public is then invited to a 'deliberative event' at which the participants are exposed to 'deliberation' – to information and debate about the issue(s) involved; and (3) opinion is re-assessed to determine the extent to which the deliberative process might have changed people's views.[2] The record of a growing number of DP events is that deliberation does indeed change median opinions.[3] The implication is that policy makers who wish to take note

[1] For specific examples, see Robert C. Luskin, James S. Fishkin and Roger Jowell, 'Considered Opinions: Deliberative Polling in Britain', *British Journal of Political Science*, 32 (2002), 455–87.

[2] See James S. Fishkin, *The Voice of the People* (New Haven, Conn.: Yale University Press, 1997); James S. Fishkin and Robert C. Luskin, 'Bringing Deliberation to Democratic Dialogue', in Maxwell McCombs and Amy Reynolds, eds, *A Poll with a Human Face* (Mahwah, N.J.: Lawrence Erlbaum, 1999), pp. 3–38; James S. Fishkin and Robert C. Luskin, 'Broadcasts of Deliberative Polls: Aspirations and Effects', *British Journal of Political Science*, 36 (2006), 184–8.

[3] See Jason Barabas, 'How Deliberation Affects Policy Opinions', *American Political Science Review*, 98 (2004), 687–701; James S. Fishkin, *When the People Speak: Deliberative Democracy and Public Consultation* (Oxford: Oxford University Press, 2009).

of public opinion in arriving at their decisions might wish to take more notice of post-deliberative, 'informed' opinion and less of the relatively 'uninformed' views of mass publics that are typically expressed in other public opinion arenas.[4]

This article reports the results of a three-day Deliberative Poll, conducted in advance of the June 2009 European Parliament elections, a poll or DP that sought to evaluate the effects of deliberation on a representative sample of citizens of the European Union (EU).[5] This 'Europolis' Poll focused on two issues of relatively high salience according to contemporary Eurobarometer polls: the question of extra-communitarian *immigration* into the European Union and the need for policy makers seriously to confront the issue of *climate change*. Participants in the DP were given briefing materials in advance of the event. At the event itself, they were given the opportunity to listen to, and to question, politicians and experts in these two policy areas. These plenary sessions were followed by small-group discussions. The views of 'deliberators' were measured both before and immediately after the DP event and again some two months later. The results from this 'test' group were compared systematically with those of a 'control' group who did not attend the DP event, but whose views were recorded at the same times as those of the test group. These comparisons enable the estimation of the extent to which participation in the DP affected participants' attitudes towards immigration and climate change. Perhaps more interestingly, however, they also allow for the analysis of the possible collateral effects of DP participation on attitudes towards the European Union itself. The reported findings show that participants in this EU-wide DP event developed more pro-EU attitudes than their control group counterparts.

The key substantive question investigated here is *why* Deliberative Polling appears to affect its participants directly and also collaterally. Five possible mechanisms – or explanations – for these observed DP effects are examined: the possible 'self-selection' consequences of DP participation; changes in participants' knowledge about the issues involved; the quality of small-group discussions; the 'social desirability' effects of small-group discussions; and the strength of the arguments articulated by other actors in the deliberation process itself. These mechanisms have all been suggested in previous work on deliberative polling.[6] The results reported show that none of these explanations is compelling or operates across all attitude/behaviour sets. Rather, the effects of DP are variegated and inconsistent, with different factors affecting different individuals in different contexts.

Section 1 of the article outlines the precise design of the DP experiment conducted in 2009. It also summarizes the main theoretical ideas that underpin each of the mechanisms that are hypothesized to produce the 'deliberative poll effect'. Section 2 presents the mean changes in opinion that occurred in the test and control groups in the period between March and August 2009. The results show that across a wide range of attitudes, mean opinion changed significantly within the test group but not within the control group. Section 3 specifies and tests a series of individual-level multivariate panel models that demonstrate the robustness of the DP effect across a range of different attitude domains.

---

[4] John Gastil, *By Popular Demand: Revitalising Representative Democracy through Deliberative Elections* (Berkeley: University of California Press, 2000).

[5] The details of the Europolis project are described at http://www.europolis-project.eu/ and in Pierangelo Isernia and Kaat Smets, 'Democracy in Hard Times: Does Deliberation Affect Attitude Strength' (CIRCaP, University of Siena, Europolis Working Paper, 2010).

[6] James S. Fishkin and Robert C. Luskin, 'Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion', *Acta Politica*, 40 (2005), 284–98. See also Fishkin, *When the People Speak*.

TABLE 1   *Timing of Interviews, Test- and Control-Group Panels*

	Wave 1 April 2009	Wave 2 23 May 2009	Wave 3 26 May 2009	Wave 4 August 2009
Test-Group Panel ($N = 348$)	t_1	t_2	t_3	t_4
Control-Group Panel ($N = 729$)	t_1	*	*	t_4

Note: Deliberative Polling Event: 29–31 May 2009. * Signifies no interview conducted in this wave.

Section 4 develops and tests models that assess the role of the five mechanisms identified above as possible explanations of the DP effect.

1. THE EXPERIMENTAL DELIBERATIVE POLL DESIGN AND POSSIBLE EXPLANATIONS FOR DP EFFECTS

The core requirement of any experimental design is to ensure that a randomly selected group of test participants can be compared systematically with members of an equivalent control group who are not subjected to the experimental test. This condition was satisfied in the present study by conducting two parallel panel surveys (see Table 1): one aimed at recruiting participants in the DP event (initial $N = 3,000$) and the other aimed at recruiting members of the control group (initial $N = 1,300$). Members of both groups were recruited by telephone, using random digit dialling, in April 2009. The initial test and control-group samples were based on national quotas in all EU member states and were both designed to be representative of the EU population as a whole. A random subset of the initial test group ($N = 600$) was invited to participate in the DP event. Some 348 of these respondents attended the event.[7] They constitute the DP *test-group panel* in the analysis here. As Table 1 indicates, their views on a range of topics were ascertained on four occasions: at the time of the recruitment interview (panel wave 1 = t_1); when they arrived for the DP event (t_2); at the end of the DP event (t_3); and two months after the European Parliament elections (t_4). The views of members of the *control-group panel* were gathered on two occasions: at the time of the recruitment interview (t_1) and two months after the European elections (t_4). Of the original 1,300 respondents to the first wave control-group survey, some 729 responded at t_4. These 729 respondents constitute the control group analysed here. The analysis consists in making comparisons (a) across panel waves within the test group panel and (b) between the changes in attitude that are observed in the test and control groups over time. The core expectation is that, if deliberation makes a difference, then the test group's views – in contrast to those of the control group – should change systematically across the four panel waves.

As noted above, the empirical analysis reported here – like many previous studies – confirms that 'deliberation' makes a difference to people's views. Why should this be the case? Extensive research has been conducted into the sources of political attitude change

---

[7] Respondents' expenses for attending the three-day DP event, which was held in a hotel on the outskirts of Brussels, were fully reimbursed and a modest per diem fee was also paid. The event was funded by the European Ccommunity under Framework Programme VI.

over many decades.[8] Aggregate time-series analyses show, for example, that average support for the incumbent government tends to move in line with the performance of the domestic economy.[9] Individual-level survey-based analyses show that people's attitudes towards a wide range of political objects vary, *inter alia*, according to their changing life circumstances, to their exposure to new social groups or to new messages that they receive from the mass media and/or from friends and acquaintances.[10] Political psychology experiments show that attitude change in relation to a particular political object results from a complex mixture of assessments of the object's salience, rational evaluations of its performance or consequences (which may include the use of heuristics or cognitive shortcuts) and emotional reactions to it.[11]

One obvious difficulty associated with trying to assess the mechanisms of attitude change in Deliberative Polls is that a DP event itself is not a laboratory-based experiment. In the case of the Europolis Poll, 348 individuals were brought together for three days of discussion about immigration and climate change. They participated both in small-group discussions, in which (in groups of around twenty-five individuals) they were encouraged to articulate their own views, and in plenary sessions, in which a variety of experts and politicians stated their views and were questioned about the topics under consideration.[12]

---

[8] See, for example, Philip E. Converse, 'The Nature of Belief Systems in Mass Publics (1964)', *Critical Review*, 18 (2006); Paul M. Sniderman, Richard A. Brody and Phillip E. Tetlock, *Reasoning and Choice: Explorations in Poltical Psychology* (Cambridge: Cambridge University Press, 1991); John Zaller, *The Nature and Origins of Mass Opinion* (Cambridge: Cambridge University Press, 1992); John Zaller and Stanley Feldman, 'A Simple Theory of the Survey Response', *American Political Science Review*, 36 (1992), 579–616.

[9] See, for example, Michael S. Lewis-Beck, *Economics and Elections: The Major Democracies* (Ann Arbor: University of Michigan Press, 1988); Helmut Norpoth, Michael Lewis-Beck and Jean Dominique Lafay, eds, *Economics and Elections: The Calculus of Support* (Ann Arbor: University of Michigan Press, 1991); Robert S. Erikson, Michael B. McKuen and James A. Stimpson, *The Macro Polity* (New York: Cambridge University Press, 2002); Han Dorussen and Michael Taylor, eds, *Economic Voting* (London: Routledge, 2002).

[10] See, for example, Benjamin I. Page and Robert Shapiro, *The Rational Public: Fifty Years of Trends in America's Policy Preferences* (Chicago: Chicago University Press, 1992); R. Michael Alvarez, *Information and Elections* (Ann Arbor: University of Michigan Press, 1997); R. Michael Alvarez and John Brehm, *Hard Choices, Easy Answers* (Princeton, N.J.: Princeton University Press, 2002); Robert Huckfeldt and John Sprague, *Citizens, Politics and Social Communication* (Cambridge: Cambridge University Press, 1995); Richard E. Petty and Duane T. Wegener, 'Attitude Change: Multiple Roles for Persuasion Variables', in Daniel T. Gilbert, Susan T. Fiske and Gardner Lindzey, eds, *The Handbook of Social Psychology*, 4th edn (Boston, Mass.: McGraw Hill, 1998), pp. 323–90; Lawrence R. Jacobs, Fay Lomax Cook and Michael X. Delli Carpini, *Talking Together: Public Deliberation and Political Participation in America* (Chicago: Chicago University Press, 2009).

[11] Samuel Popkin, *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns* (Chicago: Chicago University Press, 1991); Jon A. Krosnick and Robert P. Abelson, 'The Case for Measuring Attitude Strength in Surveys', in Judith M. Tanur, ed., *Questions about Questions: Inquiries into the Cognitive Bases of Surveys* (New York, Russell Sage Foundation, 1992), pp. 177–203; Richard E. Petty and Jon A. Krosnick, *Attitude Strength: Antecedents and Consequences* (Mahwah, N.J.: Lawrence Erlbaum, 1995); Arthur Lupia and Matthew D. McCubbins, *The Democratic Dilemma: Can Citizens Learn What They Really Need to Know?* (New York: Cambridge University Press, 1998); James H. Kuklinski and Buddy Peyton, 'Belief Systems and Political Decision Making', in Russell J. Dalton and Hans-Dieter Klingemann, eds, *Oxford Handbook of Political Behavior* (Oxford: Oxford University Press, 2007), pp. 45–63.

[12] Participants spoke and listened to the discussion in their own native language: simultaneous translation was provided as necessary for all who required it. Note that the small group discussions *preceeded* the plenary sessions. This allowed participants first to debate the issues for themselves, and then to seek further clarification or elucidation from specialists with different viewpoints in the plenaries.

Participants were supplied with detailed briefing materials both before and during the DP event.[13] They also discussed the DP topics (and presumably many other topics) in informal social exchanges throughout the DP event. Given the impossibility of directly observing the activities and responses of 348 individuals in all these contexts, we rely here on their answers to survey questions (posed in waves 3 and 4) that sought to elicit their reactions to different aspects of the DP event. Thus, for example, although we cannot know for sure how much attention individuals paid to the briefing materials they were sent in advance of the event, we can use their responses to questions about roughly how much of these materials they read as a broad guide to their actual consumption of them. Similarly, we cannot 'know' how far a particular individual's opinions were swayed by the views of a particular expert or politician. Nonetheless, we can use participants' responses to questions about the *usefulness in clarifying their thinking* of different aspects of the DP event in order to analyse the potential impact of each of these aspects on attitude change.

Bearing in mind this need to measure *indirectly* the underlying sources of any DP exposure effects on attitude change, we consider five possible mechanisms or explanations. The first relates to the possible 'selection bias' of the individuals who constitute the test group – the people who actually attended the DP event.[14] We can, of course, check to see whether or not members of the test group differed significantly in their wave 1 attitudes and characteristics from members of the control group. Indeed, this task is undertaken explicitly in Section 2. What we cannot know is whether or not the sort of people who are prepared to give up a weekend (albeit in a luxury hotel) deliberating about immigration and climate change are intrinsically more susceptible to changing their minds about political matters than are people who prefer to stay at home. The best we can do in these circumstances is, first, to weight the data as necessary so that the profile of the test sample, on any relevant characteristics, resembles the control sample as closely as possible; and, second, to control statistically for as many possible extraneous influences on attitude change as we can. In short, although we cannot test for selection bias effects directly, we can do so indirectly by testing to see if weighting the data to the key characteristics of the control sample makes any difference to any observed DP test effect. We can further mitigate any possible effects of selection bias by incorporating potentially confounding exogenous variables into any models that seek to estimate the size and significance of the 'deliberation effect'. Tests of this sort are undertaken explicitly in Sections 3 and 4.

The four other possible explanations for any observed DP test effect can be investigated rather more directly. A second mechanism that could conceivably account for DP attitude change is the acquisition of *greater political knowledge*, which could in principle change either an individual's assessment of the importance of a given issue or the position on that issue that she/he adopts.[15] The suggestion that increased knowledge of a particular policy domain should be associated with an increased sense of its *importance* or *salience* seems fairly unexceptional. However, in relation to the possible role of political knowledge in attitude *position* change, two rival hypotheses present themselves. The first derives from cognitive mobilization theory.[16]

---

[13] All materials supplied are available at the website cited in fn. 5.

[14] J. Heckman, 'Sample Selection Bias as a Specification Error', *Econometrica,* 47 (1979), 153–61.

[15] See Scott L. Althaus, 'Information Effects in Collective Preferences', *American Political Science Review*, 92 (1998), 545–58; Jason Barabas, 'How Deliberation Affects Policy Opinions', *American Political Science Review*, 98 (2004), 687–701.

[16] Ronald Inglehart, 'The Silent Revolution in Europe: Intergenerational Change in Post-Industrial Societies', *American Political Science Review*, 65 (1971), 991–1017; Ronald Inglehart, *The Silent*

The core claim is that increased political knowledge, other things being equal, serves to make people more post-materialist – more tolerant and cosmopolitan – in their general outlook. In terms of the experiment conducted here, this in turn implies that increased political knowledge as a result of DP exposure should render participants more liberal in terms of immigration, greener in terms of climate change and more pro-EU in terms of general European orientations. In short, there should be a *positive correlation between increased knowledge and cosmopolitan attitudes*. The second – counter-hypothesis – derives from Zaller's finding that increases in political knowledge can serve to *reinforce* existing attitudes rather than to change them.[17] The implication here, for example, is that (a) cosmopolitan individuals who acquire greater relevant knowledge will tend to become more convinced cosmopolitans, while (b) non-cosmopolitans in similar circumstances will tend to become more non-cosmopolitan. This in turn implies that, in 'mixed' groups such as those considered here, the effects of (a) and (b) will probably cancel each other out, producing *no observable overall effect of increased knowledge on cosmopolitan attitudes*.

In order to test these various knowledge-related hypotheses, the same battery of political knowledge questions was presented to DP participants in all four interview waves and to the control group in the wave 1 and wave 4 interviews. The items covered both specific knowledge about each of the two deliberative issues discussed – immigration and climate change – and also, given that the Poll was conducted at the time of European Parliament elections, about more general knowledge of the institutions of the European Union. It is a relatively straightforward task to ascertain whether or not increases in political knowledge were associated with attitude change and whether or not any knowledge effects were restricted solely to the DP test group. The relevant findings are reported in Sections 2 and 4 below.

A third mechanism that could underpin any DP-induced attitude change relates to the *quality of the organized small-group discussions* that took place among respondents. Deliberation is intrinsically a process that requires full consideration of the relevant ideas and issues that constitute and surround a given policy arena.[18] All of the discussion groups at the DP event were chaired by experienced moderators who had been trained to maximize the fullest possible deliberation in each group. This said, it was inevitable that there would be some variation across discussion groups in terms of the quality of the discussion or deliberation. Respondents were asked to rate the quality of the small-group deliberation on a number of criteria. By calculating the average scores on these rating scales for each discussion group, we can estimate the intersubjective quality of the deliberation in each group. If the *quality* of deliberation affects the extent of attitude change, then we would expect to find that attitude change is greater (or less) in those groups where, on average, the participants thought that the discussion quality was higher (or lower).[19] This hypothesis is also tested in Section 4.

(F'note continued)

Revolution: Changing Values and Political Styles among Western Publics (Princeton, N.J.: Princeton University Press, 1977).

[17] Zaller, *The Nature and Origins of Mass Opinion.*

[18] James S. Fishkin and Robert C. Luskin, 'Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion', *Acta Politica*, 40 (2005), 284–98; Diana C. Mutz, *Hearing the Other Side: Deliberative versus Participatory Democracy* (Cambridge: Cambridge University Press, 2006). See also Benjamin I. Page, *Who Deliberates? Mass Media in Modern Democracy* (Chicago: Chicago University Press, 1996).

[19] This use of aggregated intersubjective measures of discussion quality here is the result of necessity rather than choice. One important part of the Europolis DP project was the construction of objective measures of discussion quality based on content analyses of the actual discussion sessions. Unfortunately,

A fourth possible mechanism of attitude change under deliberation relates to *social desirability*. Social desirability pressures are known to produce a range of responses in face-to-face survey interviews and in small-group situations.[20] In interviews, respondents will sometimes give responses that they think they 'ought' to give or that they believe the interviewer 'expects' of them. In small groups, some participants may adjust their views on a particular topic in order to bring them into line with what they perceive to be majority opinion within the group. *In extremis*, this kind of thinking produces the well-known 'groupthink' phenomenon, in which members of a group converge on a single view of a key problem and fail to articulate plausible criticisms of it.[21] It is clearly possible that social desirability pressures, notwithstanding the core intentions of the deliberative process, could have similar effects in DP small groups. This would imply a sort of regression to the mean over time: individuals who score more highly than the relevant small-group mean on a given attitude scale at wave 1 would be expected to shift their positions on the scale downwards by wave 4; those below the small group mean would be expected to shift their positions upwards. This in turn implies that, for any given attitude dimension, a variable that measures each respondent's position as a *deviation from the small-group mean* at wave 1 should have a positive effect on the *change* in the attitude between waves 1 and 4. Again, this idea is tested explicitly in Section 4.

A final mechanism that could induce attitude change under deliberation is the simple *power of argumentation* to which participants are subjected. In normative terms, this should be the core mechanism through which opinion change occurs.[22] Deliberation should engender changes of mind in particular individuals because they are now convinced by arguments either that they had not heard before or that they now, having been subjected to a new, more lucid exposition, find more compelling. The immediate difficulty here, of course, is specifying criteria that might enable a 'compelling' argument to be distinguished from a 'non-compelling' one. Nonetheless, although we cannot measure 'argument strength' objectively, it is possible to use participants' perceptions of the usefulness of different aspects of the DP process for *clarifying their own thinking* on key issues. In this respect, participants in the event were asked to rate four different aspects of the DP process on 0–10 scales, where zero denoted 'a complete waste of time' and 10 denoted 'extremely useful'. These aspects were: reading the briefing documentation provided; listening to 'experts'; listening to 'politicians'; and 'talking to other participants outside the formal discussions'. By considering participants' own estimates of the utility of each aspect for clarifying their thinking, an assessment can be made of the extent to which participants felt that the arguments articulated in these different vehicles were compelling *for them*. The empirical implication is that the more (or less) an individual

---

(F'note continued)
financial constraints on the project meant that it was possible to produce objective measures of deliberation quality for only seven of the twenty-five discussion groups. Moreover, these groups were restricted to a limited set of languages and, therefore, cannot be regarded as representative of the full range of deliberation discussion that took place. In addition, to have objective data on only seven groups does not provide sufficient cases for a robust analysis of the effects of objective discussion quality to be undertaken. In these circumstances, I reluctantly restrict the analysis of 'deliberation quality' to subjective, as opposed to objective, measures.

[20] R. J. Fisher, 'Social Desirability Bias and the Validity of Indirect Questioning', *Journal of Consumer Research*, 20 (1993), 303–15.

[21] Irving L. Janis, *Victims of Groupthink* (Boston, Mass.: Houghton Mifflin, 1972).

[22] Jon Elster, *Deliberative Democracy* (Cambridge: Cambridge University Press, 1998).

found a particular vehicle useful for clarifying her/his thinking, the more (or less) likely she/he is to change her/his views on any given attitude dimension. Again, this claim is tested explicitly in Section 4.

2. DESCRIBING THE DP EFFECT: CHANGES IN TEST- AND CONTROL-GROUP OPINIONS ACROSS INTERVIEW WAVES

As noted above, the Europolis Deliberative Poll experiment focused on two core issues: immigration from outside the European Union and climate change. Each of these issues was measured in terms of both *salience* and *position*. *Salience* was assessed by asking respondents to rate the importance of each issue on a 0–10 scale.[23] *Position on immigration* was measured by combining the responses to a range of questions that sought to establish the respondent's degree of tolerance (on a 0–10 scale) towards immigrants.[24] *Position on climate change* was assessed by asking respondents to place themselves on a 0–10 scale where high values denoted support for strong measures to combat climate change even if this damages economic growth, and low values denoted support for promoting economic growth even if it hurt efforts to combat climate change.[25] Table 2 reports mean values on these four sets of measures for both the test panel and the control panel. For the test group, means are reported for waves 1 through 4. For the control group, they are reported for waves 1 and 4 only, as the control group was not interviewed in waves 2 and 3. The table also reports the probability values for difference of means *t*-tests. For each attitude measure, these *t*-tests make three sets of comparisons: between test and control groups means at wave 1; between test group means at waves 1 and 4; and between control group means at waves 1 and 4.

Several important patterns are revealed by the table. First, the scores for the test group on all four measures increase significantly between waves 1 and 4. For example, the mean score on the pro-immigration index increases from 5.47 to 5.79 ($p = 0.002$), and the mean on the climate importance score increases from 7.53 to 8.26 ($p = 0.000$). Second, the test group scores all increase progressively during the actual test exposure – from wave 1 to wave 2 and from wave 2 to wave 3. Three of the scores then decline slightly from wave 3 to wave 4 (the exception being the pro-immigration index, which continues to rise slightly through to wave 4), suggesting that the test effect tends to decay only moderately after the DP event. A third feature of the table is that, while the changes in the *test*-group mean scores between waves 1 and 4 are all significant, the changes in the *control*-group mean scores with one exception are not. Although the control group 'Combat Climate Change' index changes significantly between waves 1 and 4 ($p = 0.027$), the changes in the three

---

[23] The question used was: 'On a scale from 0 to 10, where 0 is "no problem at all", 10 is "the most serous problem we face", and 5 is "exactly in the middle" how serious a problem or not would you say [immigration/climate change] is?'

[24] The Pro-immigration index was constructed, after extensive dimensional testing, as a 0–10 constant range scale from twelve questions that sought to measure respondents' attitudes towards immigrants to the European Union. Details on index construction are available from the author on request.

[25] The combat climate change scale was based on the question: 'On a scale from 0–10, where 0 means that we should do everything possible to combat climate change even if that hurts the economy, 10 means that we should do everything possible to maximize economic growth, even if that hurts efforts to combat climate change and 5 is exactly in the middle, where would you position yourself on this scale, or haven't you thought much about that?' The ordering of the scale was reversed so that a high score signified a more pro-environmental position.

TABLE 2   *Changes in Core Immigration and Climate Change Attitude Mean Scores, Wave 1 to Wave 4*

| | Range | Mean score | | Difference of means *t*-test *p*-value | | |
		Test Group	Control Group	Test versus Control Group, w1	Test Group w1→w4	Control Group w1→w4
Pro-Immigration Index	0–10			0.000	0.002	0.706
Wave 1		5.47	5.08			
Wave 2		5.56				
Wave 3		5.64				
Wave 4		5.79	5.10			
Immigration Importance	0–10			0.187	0.000	0.925
Wave 1		5.35	5.60			
Wave 2		6.12				
Wave 3		6.57				
Wave 4		5.97	5.61			
Combat Climate Index	0–10			0.104	0.004	0.027
Wave 1		5.87	5.56			
Wave 2		6.29				
Wave 3		6.65				
Wave 4		6.32	5.80			
Climate Importance	0–10			0.807	0.000	0.161
Wave 1		7.53	7.49			
Wave 2		7.94				
Wave 3		8.31				
Wave 4		8.26	7.62			

remaining measures all fail to achieve significance at conventional levels. The crucial point here is that this pattern of significant change in the test group and (generally) non-significant change in the control group strongly confirms the idea that the Deliberative Poll experience did have a powerful effect on participants' opinions. Test-group subjects on average became more liberal on immigration (mean increase: 5.47 to 5.79); increased their average rating of the importance of both immigration and climate change (which increased, respectively, from 5.35 to 5.97 and from 7.53 to 8.26); and became more convinced of the need to combat climate change even at the expense of economic growth (mean change: 7.53 to 8.26). The final feature of Table 2 that merits attention is the comparison between the average test-group and control-group scores at wave 1. The difference of means tests here indicate that in relation to three of the measures (the two importance measures and the combat climate change index), the test- and control-group means are statistically indistinguishable. This said, the test group is more liberal on immigration at wave 1 than the control group (5.47 in comparison with 5.08; $p = 0.000$), implying that the test- and control-group panels may differ from each other in important ways. This possibility is considered explicitly in Section 3 below.

It is clear from Table 2 that exposure to deliberation affected the immigration and climate-change attitudes of the test-group panel respondents. Given that test respondents spent three days explicitly discussing these two core issues, these manifest effects of deliberation are perhaps unsurprising. It is possible, however, that a deliberative poll that brings together participants from all over the European Union could also have important latent effects on people's attitudes. Simply exposing people to others from other parts of the Union in a friendly and relaxed atmosphere could 'collaterally' change people's perceptions of the European Union itself. Table 3 reports on the equivalent changes for collateral effects on EU-related attitudes. The table distinguishes between four different sets of EU-related attitudes:[26] people's sense of European identity;[27] their preferences for EU policy competence or 'policy scope';[28] their degree of satisfaction with EU-level democracy;[29] and their overall evaluations of the European Union as a whole.[30] The results are very similar to those observed in Table 2. First, on three out of four of the

---

[26] These measures are not highly intercorrelated. The highest bivariate correlation is $r = 0.36$. This reinforces the idea that the measures reflect distinct sets of attitudes towards Europe. This interpretation is supported by extensive dimensional analysis of EU-related attitudes reported in David Sanders, Paolo Bellucci, Mariano Torcal and Gábor Tóka, eds, *The Europeanization of National Polities? Citizenship and Support in a Post Enlargement Europe* (Oxford: Oxford University Press, 2012 forthcoming).

[27] The EU identity measure was taken from responses to the question: 'On a scale from 0 to 10, where 0 is "not at all", 10 is "completely", and 5 is exactly in the middle, how much would you say you think of yourself as European?'

[28] The EU Policy Scope measure was constructed additively from questions which asked respondents' preferences for EU-level (as opposed to national or regional level) decision making in four policy areas – 'fighting unemployment', 'climate change', 'immigration policy' and 'fight against crime'.

[29] The EU Democracy Satisfaction or 'EU Representation' measure was constructed from responses to: 'On the whole, how satisfied or not are you with the way democracy works in the European Union?' Response options were Very Satisfied, Somewhat Satisfied, Neither Satisfied nor Dissatisfied, Dissatisfied, Very Dissatisfied.

[30] The EU Evaluations index combined responses to two questions: '(1) Generally speaking, do you think that [country's] membership of the European Union is a Very Good Thing, a Fairly Good Thing, Neither Good nor Bad, a Fairly Good Thing, A Very Bad thing; (2) On a 0 to 10 scale, where 0 means that [country] has "not benefited at all" from being a member of the EU, 10 means "has benefited enormously", and 5 is "exactly in the middle", using this scale, would you say that on balance [country] has benefited or not from being a member of the EU?'

TABLE 3   *Changes in 'Collateral' EU-Related Attitude Mean Scores, Wave 1 to Wave 4*

| | Range | Mean score | | Difference of means *t*-test *p*-value | | |
		Test Group	Control Group	Test versus Control Group, w1	Test Group w1→w4	Control Group w1→w4
European Identity	0–10			0.072	0.007	0.533
Wave 1		7.11	6.76			
Wave 4		7.53	6.83			
EU Policy Scope	0–4			0.536	0.367	0.277
Wave 1		1.58	1.53			
Wave 4		1.64	1.48			
EU Representation	0–4			0.432	0.011	0.570
Wave 1		2.11	2.06			
Wave 4		2.25	2.08			
EU Evaluations	0–10			0.024	0.000	0.248
Wave 1		4.68	4.47			
Wave 4		4.99	4.53			

measures (the exception is EU Evaluations), the wave 1 test-group mean is indistinguishable from the wave 1 control-group mean. Second, for the *test* group, the average change between wave 1 and wave 4 is generally significant (the only exception is in relation to EU Policy Scope). In contrast, for the *control* group, there is *no* significant change between waves 1 and 4. This again supports the idea of a Deliberative Poll effect: opinion in the test panel, exposed to deliberation, changes and opinion in the control panel, not so exposed, does not.

Overall, two general conclusions are suggested by the results presented in Tables 2 and 3. First, across both tables there is evidence of some sort of test-exposure effect. In both tables, the test-group mean scores tend to increase significantly, whereas the control-group scores tend either to fall or not to change at all. This is clear evidence of a Deliberative Poll or DP effect. Second, with a relatively small number of exceptions, at wave 1, the test and control panels were indistinguishable from each other on most of the measures reported. This suggests that the test and control panels were fundamentally similar, with the implication that any observed test effect is due to the operation of the test stimulus (the DP event) rather than the result of sampling bias. This said, the 'relatively small number of exceptions' suggests the need for continuing caution about the direct comparablility of the test and control samples. Appropriate controls are accordingly introduced in the analysis reported here later.

3. TESTING THE ROBUSTNESS OF THE DP EFFECT: SPECIFYING AND ESTIMATING MULTIVARIATE MODELS

It is clear from the results reported above that the views of the participants in the Europolis Deliberative Poll changed much more than those of the non-participant control group: the simple experimental control of exposure to deliberation *did* change the way people thought across a range of different issues. In addition to this experimental control, however, it is also possible to apply non-experimental controls – to use multivariate statistical controls to test the robustness of the bivariate relationships that have thus far been described. The general form of the model employed to test the robustness of the bivariate DP effect is

$$Y_t = a + b_0 Y_{t-1} + \Sigma(b_k X_{kt}) + b_j Test + u_{it}, \tag{1}$$

where Y_t is the dependent variable attitude measured at wave 4; Y_{t-1} is the lagged dependent variable measured at wave 1; X_{kt} is a set of independent control variables measured at wave 1; *Test* is a dummy variable, where 1 indicates membership of the DP test group and 0 membership of the control group; and u_{it} is a random error term.

Table 4 provides an illustration of this model specification and the general set of control variables that were employed. The dependent variable in the table is the respondent's score at wave 4 on the pro-immigration index described earlier. The inclusion of a lagged dependent variable on the right-hand side of the equation effectively means that the (weighted) *change* in pro-immigration attitudes between waves 1 and 4 is being modelled. The larger (or smaller) this coefficient is, the more (or less) stable the measured dependent variable attitude is between waves 1 and 4. There are no general expectations about the signs or significance levels of the coefficients on the independent control variables. They are included because any or all of them could in principle have effects on the change in the dependent variable. They thus include controls for gender, age, education, class, religion, religiosity, feelings of national identity, left–right ideology (including a term for left–right

TABLE 4   *Model of Change in Pro-Immigration Position Score, Wave 1 to Wave 4*

	Coeff.	Std	Prob.
Pro-Immigration Position Score, wave 1 (0–10)	0.69	0.03	0.000
Male/not (0/1)	−0.03	0.07	0.641
Age Category (1–6)	−0.02	0.02	0.476
Education (1–4)	0.09	0.04	0.021
Catholic/not (0/1)	0.01	0.08	0.947
Protestant/not (0/1)	−0.04	0.11	0.710
Working Class/not (0/1)	−0.13	0.08	0.108
Religiosity (1–3)	−0.01	0.08	0.933
National Identity (0–10)	−0.01	0.01	0.494
Left–Right Position (0–10)	−0.13	0.04	0.004
Left–Right Squared	0.01	0.01	0.082
Left Party Grouping/not (0/1)	−0.04	0.09	0.693
Right Party Grouping/not (0/1)	−0.11	0.09	0.198
Test Exposure – deliberation/not (0/1)	0.38	0.07	0.000
Constant	2.02	0.30	0.000
N	1,077		
Adjusted R^2	0.52		

Notes: Dependent Variable: Pro-Immigration Position Score, Wave 1. All independent variables measured at Wave 1. Age categories are: 18–25; 26–35; 36–45; 46–55; 56–65; 65+. Religiosity based on frequency of attendance at religious services. Education is based on years in education: up to 15; 16–18; 19–24; 24+. National Identity is respondent's self-placement on a 0–10 scale. Left–Right Position is respondent's self-placement on a 0–10 scale. Left (Right) Party Grouping indicates whether or not, in Wave vote intention, the respondent specified a party that was part of a left (right) of centre grouping in the European Parliament. Left groups were defined as: PES, Far Left or Green; Right groups as EPP Far Right or Libertas; the Control category was Other/None.

ideology squared to incorporate the possibility that the effects of ideology may be non-linear), and party preference (to take account of the possibility that, over and above any effects of ideology, the cues provided by either left-leaning or right-leaning parties might affect attitude change). The key theoretical expectation for the model reported in Table 4 is that the coefficient for the *Test* variable will be positive and significant. If this expectation is fullfilled, it suggests that the DP test effect is sufficiently robust to withstand these various non-experimental controls – that the test effect endures even when there is control for these other possible influences on attitude change.

The results reported in Table 4 are reassuring. First, the coefficient on the lagged dependent variable is relatively high ($b = 0.69$), suggesting that, although there is some change in pro-immigration attitudes across waves, these attitudes are also fairly stable. Second, only two of the control variables are significant at conventional levels. The positive coefficient on the education term ($b = 0.09$) indicates that more educated respondents were slightly more likely to become more pro-immigration across waves. The negative coefficient on the left–right scale term ($b = -0.13$) implies that more right-wing respondents were less likely to become more pro-immigration than their more left-wing counterparts. The remaining control variables all yield non-significant coefficients, suggesting that these other variables had no effects on the change in pro-immigration attitudes between waves 1 and 4. Third, the model is reasonably well determined, with an adjusted $r^2 = 0.52$, although this is to be expected, given the inclusion of the lagged

TABLE 5   *Significance of Test Effect in Multivariate Models of Change in Core and Collateral Dependent Variables, Wave1 to Wave 4*

	Test *vs* Control Effect (0/1)		
	Coeff.	Std	Prob.
Core Attitude Change Model			
Pro-Immigration Position Index Model (0–10)	0.38	0.07	0.000
Immigration Importance Model (0–10)	0.54	0.16	0.001
Combat Climate Change Position Model (0–10)	0.37	0.16	0.020
Climate Change Importance Model (0–10)	0.62	0.12	0.000
Collateral EU-Attitude Change Model			
European Identity (0–10)	0.44	0.14	0.002
EU Policy Scope (0–10)	0.11	0.07	0.162
EU Representation (0–4)	0.14	0.06	0.016
EU Evaluations (0–4)	0.78	0.10	0.000

endogenous variable. The key finding of the table, however, relates to the coefficient for the Test term. This is, as expected, both positive and highly significant. Controlling for all the other variables in the model, exposure to deliberation still exerts a large, significant effect on people's attitudes towards immigration. Indeed, the coefficient of $b = 0.38$ ($p = 0.000$) indicates that, in comparison with the control group, those exposed to deliberation on average increased their scores on the 0–10 pro-immigration index by 0.38 of a point between waves 1 and 4.

Table 5 reports the consequences of estimating a series of models, based on the specification described in Equation 1 and Table 4, for each of the dependent variables identified in Tables 2 and 3. The results confirm that the inferences about the DP test effect on pro-immigration attitudes drawn from Table 5 can be generalized both to the remaining core attitudes towards immigration and climate change and also to three of the four collateral EU attitude sets. With the exception of attitudes towards EU Policy Scope (which in any case, as seen in Table 3, produced no *bivariate* test effect), all of the Test coefficients retain their significance in the face of extensive multivariate controls. These findings strongly support the idea that Deliberative Polls do change people's attitudes – both in relation to the core issues that are explicitly being deliberated upon but also, in this case, in relation to collateral attitudes relating to the European Union as a whole.[31]

---

[31] This strong collateral effect on EU attitudes was not observed in the only other EU-wide deliberative poll that has been conducted – Tomorrow's Europe (see http://cdd.stanford.edu/polls/eu/ *Tomorrow's Europe: the first EU wide Deliberative Poll*). One possible explanation for the difference in these two sets of DP findings could derive from the differential character of the issues debated. In the Europolis poll covered here, the issues of immigration and climate change perhaps intrinsically lend themselves to EU-wide *solutions* and accordingly invoked a more pro-EU response among participants. In contrast, the Tomorrow's Europe DP focused on Turkey's entry into the European Union. This could have served to focus participants' attention to the *difficulties* likely to confront EU citizens in the future, which in turn could have invoked less cosmopolitan attitudes among the Tomorrow's Europe participants. A second possibility is that the set of EU attitude measures employed in the Europolis event were more comprehensive than those used in Tomorrow's Europe. Future studies should enable these potential factors to be considered more explicitly.

4. WHY DO DELIBERATIVE POLLS CHANGE PEOPLE'S ATTITUDES?

The results presented in the two previous sections show that there is a clear Deliberative Poll effect. In Section 1, a series of possible mechanisms were advanced that might help to explain why this effect occurs. This section develops and tests models that seek to assess the possible role played by these various mechanisms in generating the DP effect.

The first possible mechanism investigated here can be summarized under the general heading of 'the effects of (self) selection bias'. The idea here is that any observed DP effect is not really the result of actual deliberation. Rather, it reflects the fact that the type of person who allows herself or himself to be exposed to a DP is *a priori* more likely to change her or his views on any given issue compared with someone who is not interested in participating in a DP. Thus, if the test group changes its views during its deliberation more than the control group does during its non-deliberation, than this is not because of the deliberation itself. If the non-deliberators had somehow been induced to deliberate, they would not necessarily have changed their views under deliberation in the same way as the deliberators did. This claim obviously cannot be tested directly. However, it is possible to use information about the test- and control-group samples to *weight* the test-group cases so that, as far as possible, the profile of the test group – in terms of both demographics and key attitudes that are not being used as dependent variables – matches that of the control group. Using this weighted version of the test-group data, we can simulate how the test group would have behaved if, counterfactually, they had possessed the same demographic and attitudinal profile as the control group. Although not the same as subjecting the non-deliberators to deliberation, it is nevertheless the nearest that we can get to the counterfactual position.

Is such a weighting justified? And if so, how should it be effected? It will be recalled from Tables 2 and 3 that there were certain respects in which the test- and control-group samples at wave 1 were significantly different from one another: in Table 2, in terms of pro-immigration attitudes; and in Table 3, in terms of EU Evaluations. In addition, *t*-test results (not reported) comparing the demographic profiles of the test and control groups showed that they were significantly different from one another in terms of interest in politics, political knowledge, gender and education (the test group being disproportionately interested, knowledgeable, female and well educated). Making decisions about how to weight data is always a difficult process. If too many weighting variables are used, then the analytic results can become unstable because of the excessive overweighting assigned to particular individuals. If too few are used, then the profile of the weighted sample fails to correspond sufficiently to the target sample. It is also standard practice to avoid applying weights determined by the dependent variable. Here, we report the consequences of weighting the test group data by the three (non-dependent) variables that displayed the greatest differences between the test and the control groups: interest in politics, gender and education.[32]

Tables 6 and 7 report the equivalent findings to those outlined in Tables 4 and 5, but with the data weighted by interest, gender and education. The estimated models also take account of the clustering of the data that is implied by the separate collection of the test- and control-group samples. Table 6 indicates the consequences of weighting and clustering the model shown in Table 4. The results show that the two independent variables that were significant in Table 4 model (left–right ideology and education) are no longer significant at conventional levels. The core results, however, are very similar to

[32] Other weighting combinations in fact produced similar results to those reported here.

TABLE 6   *Model of Change in Pro-Immigration Position Score, Wave 1 to Wave 4: Data Weighted and Clustered to Take Account of Possible Selection Bias*

	Coeff.	Robust Std	Prob.
Pro-Immigration Position Score, wave 1 (0–10)	0.66	0.04	0.000
Male/not (0/1)	−0.02	0.10	0.867
Age Category (1–6)	−0.02	0.02	0.527
Education (1–4)	0.11	0.07	0.367
Catholic/not (0/1)	0.01	0.08	0.951
Protestant/not (0/1)	−0.07	0.14	0.695
Working Class/not (0/1)	−0.14	0.05	0.202
Religiosity (1–3)	0.02	0.10	0.846
National Identity (0–10)	−0.01	0.01	0.712
Left–Right Position (0–10)	−0.14	0.01	0.066
Left–Right Squared	0.01	0.00	0.076
Left Party Grouping/not (0/1)	−0.05	0.12	0.757
Right Party Grouping/not (0/1)	−0.15	0.03	0.112
Test Exposure – deliberation/not (0/1)	0.41	0.00	0.002
Constant	2.12	0.16	0.047
Weighted N	1,076		
Adjusted R^2	0.50		

Note: Data weighted by Gender, Interest in Politics and Education. Standard errors adjusted for two clusters (one test; one control) in sample. Dependent Variable: Pro-Immigration Position Score, Wave 1. All other variables as defined in Table 4.

TABLE 7   *Significance of Test Effect in Multivariate Models of Change in Core and Collateral Dependent Variables, Wave1 to Wave 4: Data Weighted and Clustered to Take Account of Possible Selection Bias*

	Test *vs* Control Effect (0/1)		
	Coeff.	Std	Prob.
Core Attitude Change Models			
Pro-Immigration Position Index Model (0–10)	0.41	0.00	0.002
Immigration Importance Model (0–10)	0.60	0.02	0.022
Combat Climate Change Position Model (0–10)	0.44	0.01	0.018
Climate Change Importance Model (0–10)	0.67	0.01	0.007
Collateral EU-Attitude Change Models			
European Identity (0–10)	0.52	0.02	0.018
EU Policy Scope (0–10)	0.11	0.02	0.092
EU Representation (0–4)	0.16	0.01	0.017
EU Evaluations (0–4)	0.79	0.04	0.030

Note: Data weighted by Gender, Interest in Politics and Education. Robust Standard Errors adjusted for two clusters (one test; one control) in sample.

those reported in Table 4. The coefficients on the lagged dependent variable and on the *Test* variable are both positive and significant. This again suggests that, in spite of weighting and clustering, actual participation in the DP event was crucial in changing opinions on the pro-immigration index between waves 1 and 4. A similar set of general conclusions is in order in relation to Table 7. As in Table 4, the results in Table 7 show

that the *Test* coefficient is significant and positive in all the attitude models except for EU Policy Scope. The clear implication of all this is that weighting and clustering the data make no difference to the quasi-ubiquity of the DP *Test* effect. This in turn suggests that the DP effect cannot be plausibly explained by the distinctive character of the test group. In sum, (self) selection bias appears to be an unconvincing explanation of the DP *Test* effect or, at a minimum, an explanation that is 'not proven'.

The analysis conducted so far has involved making either explicit or implicit comparisons between test- and control-group respondents. It has been established that there is a clear DP effect and that that effect is not explained by the composition of the test-group sample. In examining other possible explanations for the DP effect, the analysis now focuses solely on the test group. It will be recalled from Section 1 that four putative mechanisms underlying deliberators' changing views mechanisms were hypothesized to operate in this context, as follows:

— *Increased political knowledge*. This was measured as the change in an individual's knowledge quiz score (minimum score = 0; maximum = 6) between waves 1 and 4. As noted in Table 2, it is clear that, while political knowledge increased significantly among test-group respondents, among the control group it changed barely at all. Recall that one background 'salience' hypothesis and two rival 'position' hypotheses are of relevance if greater political knowledge represents a mechanism through which the DP effect operates. The *salience* hypothesis is simply that greater knowledge of an issue induces people to attribute greater importance to it. The first *position* hypothesis (based on Inglehart's work) is that increased knowledge should make people more tolerant on immigration, more inclined to support measures to combat climate change, more likely to regard immigration and climate change as important and more sympathetic towards the European Union. In short, changes in knowledge should exert significant, positive effects on each of our core- and collateral-dependent variables. The second *position* hypothesis (based on Zaller's work), in contrast, suggests that the reinforcement effects of greater knowledge should cancel each other out and that knowledge change should accordingly have no effect on our core- and collateral-independent variables.

— *Small-group discussion quality*. The twin assumptions here are that *average perceptions* of the quality of discussion can be used as a surrogate for the actual quality of the small-group discussions at the DP event, and that the quality of discussion, thus measured, can be used to assess the quality of deliberation. It is hypothesized that, if discussion quality is one of the mechanisms through which the DP effect operates, it should exert significant, positive effects on each of our core- and collateral-dependent variables.

— *Social desirability pressures*. This is the tendency for individuals to adjust their own preferences to conform to the social environment in which they find themselves. In the context of the Deliberative Poll, we use average opinion in each of the small discussion groups to represent these environments and then use the extent to which the individual respondent's views deviate from her or his small-group mean as a measure of the extent of social desirability pressures. The general hypothesis here is that, for any given attitude, individuals who score *above* the small-group mean at wave 1 should *reduce* their scores on the attitude by wave 4; and that individuals *below* the small-group mean at wave 1 should *increase* their scores by wave 4. This in turn implies that a variable that measures *above*-average scores on a given attitude at wave 1 as deviations from the small-group mean, but which is otherwise zero, should have a *negative* effect on

attitude change between waves 1 and 4. Concomitantly, a variable that measures *below*-average scores as deviations from the small-group mean, but which is otherwise zero, should have a *positive* effect on attitude change between waves 1 and 4. These symmetrical social pressure effects are accordingly operationalized below as a pair of predictor variables – one that measures positive deviations from the group mean and one that measures negative deviations.

— *Articulation of arguments that clarify an individual participant's thinking.* As noted above, participants were asked to rate the importance of four factors in clarifying their thinking: the DP documentation provided and the arguments expressed, respectively, by experts and politicians in the plenary sessions and by fellow participants in informal discussions. The hypothesis here is that, for each of these possible factors, attitude change between waves 1 and 4 should be positively and significantly affected by the strength of the perceived clarifying effect.

The possible effects of these various measures on the views of deliberators were estimated as:

$$\begin{aligned}
Y_t = a &+ b_0 Y_{t-1} + b_1 \delta Knowledge + b_2 \ Discussion \ Quality \\
&+ b_3 \ Social \ Desirability \ Pressure \ (above - group \ mean)_t \\
&+ b_4 \ Social \ Desirability \ Pressure \ (below - group \ mean)_t \\
&+ b_5 \ Documentation + b_6 \ Experts + b_7 \ Politicians \\
&+ b_8 \ Informal \ Conversations + \Sigma(b_k X_{kt}) + u_{it},
\end{aligned} \qquad (2)$$

where Y_t is the dependent variable attitude measured at wave 4; Y_{t-1} is the lagged dependent variable measured at wave 1; $\delta Knowledge$ is the change in the respondent's political knowledge score between waves 1 and 4;[33] *Discussion Quality* is the average perceived discussion quality of each small-discussion group; *Social Desirability Pressure$_t$* is the extent to which the respondent deviates from the small discussion-group mean score on the issue under consideration, at wave 1; *Documentation*, *Experts*, *Politicians* and *Informal Conversations*, respectively, measure the respondents's perceptions of the 'clarifying effects' of each of these factors; X_{kt} is a set of independent control variables measured at wave 1; and u_{it} is a random error term. Note that there is now no *Test* variable as in Equation 1 because this specification relates to deliberators only: the control group is excluded.

Tables 8 and 9 report the results of estimating Equation 2 for each of our core- and collateral-dependent variables. To avoid overburdening the reader with detail, the tables do not report coefficients for the independent control variables even though they were included in the estimation.[34] The estimations cluster the respondents by small-discussion group, to ensure the correct estimation of (robust) standard errors. Table 8 presents the results for attitudes towards immigration and climate change. In the *immigration position* equation (column 1), the only significant mechanism variable is 'talking informally to other participants'. The coefficient ($b = 0.08$) is significant and correctly signed. The remaining coefficients, however, all fail to achieve accepted levels of significance. In the *immigration importance* equation (column 2), only the *Social Desirability* (above group mean) coefficient

---

[33] There are other possible ways of specifying and estimating knowledge acquisition effects. See fn. 19 and Tables A1 and A2 in the Appendix for an analysis of the main alternative specifications.

[34] Dropping these controls from the various specifications makes no difference to the magnitudes, signs or significance levels of the key explanatory variables.

($b = -0.57$) is significant and correctly signed. In the *climate change position* equation (column 3), *none* of the putative explanatory mechanism variables achieves significance at conventional levels. In the *climate change importance* model (column 4), the *Experts* term is correctly signed and almost significant ($p = 0.06$). The *Social Desirability* (above group mean) term is significant but is incorrectly signed. Contrary to theoretical expectations, the positive coefficient of $b = 0.12$ implies that a climate change position score above the small-group mean at wave 1 produced an *increase* in the respondent's wave 4 score. This clearly contradicts the idea that social-desirability pressures led respondents to adjust their own climate change attitudes to the group mean.

These results are obviously disappointing in terms of the extent to which they shed light on the reasons why Deliberative Polling affects participants' attitudes. A similar conclusion is in order in relation to the findings shown in Table 9, where models for EU-related attitudes are presented. In the *EU Representation* equation (column 3), the *Experts* term is positive and significant in the *EU Evaluations* equation (column 4) and the same is true of the *Knowledge* coefficient. However, in the *European Identity* and *EU Policy Scope* equations (columns 1 and 2 respectively), none of the mechanisms produces a statistically significant effect.

What do these findings tell us about the possible mechanisms that underpin the effects of deliberative polling? The brutal conclusion has to be 'nothing very positive'. The pattern of significant effects across Tables 8 and 9 is far from consistent. The *Experts* term features significantly in two of the seven equations (for climate change importance and for EU Representation) while *Social Desirability* and 'talking to others' each appear correctly signed in only one each (respectively, the immigration importance and position models).[35] It would probably be unreasonable to demand that a given predictor variable should figure significantly in all seven equations in order to conclude that the mechanism it represents really does help produce the 'Deliberative Poll effect'. However, the feeble pattern of effects that is actually observed in Tables 8 and 9 clearly offers very little systematic explanation of why the DP effect occurs. Indeed, there are actually three compelling reasons for concluding that the models presented in Tables 8 and 9 do not identify the mechanisms responsible for the DP effect. First, there is no obvious theoretical reason (or set of reasons) why *Social Desirability* should affect 'immigration importance' while the clarifying effects of talking to others should affect 'immigration position' – rather than, say, vice versa – or why these same variables should leave all other measured attitudes unaffected. Second, the amount of variation in the respective dependent variables that is explained by the significant mechanisms identified in Tables 8 and 9 is extremely limited. The tables report the reductions in r^2 values that occur if each significant putative mechanism variable is dropped from the specification. In all four instances in which a significant effect is thus dropped, the fall in r^2 is extremely modest: by 0.03 (from 0.25 to 0.22) in the immigration-importance equation and by 0.01 in the other three.

The third reason is perhaps the most damning of all. The models in Tables 8 and 9 estimate the effects of nine 'mechanism coefficients' across each of eight equations – a total of seventy-two possible effects. At the $p = 0.05$ significance level, we would expect,

---

[35] The political knowledge term produces no significant effects at conventionally accepted levels. This suggests that there is no systematic tendency for increased knowledge to change DP participants' attitudes in terms of either salience or position. This conclusion seems to support Zaller's claims about the impact of political knowledge rather than Inglehart's.

TABLE 8   *Models of Change in Core Immigration and Climate Change Dependent Variables Wave 1 to Wave 4, Deliberators Only*

	Immigration Position		Immigration Importance		Climate Position		Climate Importance	
	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
Immigration Position wave 1	**0.68**	**0.000**						
Immigration Importance wave 1			**0.25**	**0.000**				
Climate Change Position wave 1					**0.50**	**0.030**		
Climate Change Importance wave 1							**0.32**	**0.000**
Change in Knowledge waves 1–4	−0.07	0.181	−0.07	0.505	0.03	0.758	−0.04	0.375
Group Discussion Quality	0.13	0.245	0.36	0.121	−0.01	0.969	−0.18	0.301
Social Conformity Pressure (Above)	−0.03	0.816	**−0.57**	**0.015**	−0.18	0.367	**0.12**	**0.034**
Social Conformity Pressure (Below)	−0.10	0.498	−0.27	0.167	−0.31	0.081	−0.00	0.918
Clarify thinking – experts	0.02	0.670	0.05	0.569	0.02	0.876	0.13	0.069
Clarify thinking – politicians	−0.01	0.612	0.01	0.07	0.07	0.444	−0.02	0.660
Clarify thinking – other participants	**0.08**	**0.036**	0.09	0.135	−0.04	0.625	−0.00	0.959
Clarify thinking – documentation	0.04	0.298	−0.12	0.127	0.11	0.163	0.01	0.846
Constant	0.41	0.724	2.38	0.263	2.34	0.429	7.42	0.000
Weighted N	347		347		347		347	
Adjusted R^2	0.55		0.25		0.19		0.32	
Adjusted R^{2*}	0.54		0.22		n/a		0.31	

Notes: Controls included but not reported (all measured at wave 1) for Gender, Age, Education, Catholic, Protestant, Working Class, Religiosity, Left–Right Ideology, Left–Right Squared, Left Party, Right Party.

Data weighted by Gender, Interest in Politics and Education. Standard errors adjusted for twenty-five clusters (the twenty-five separate small discussion groups) in sample.

Social Conformity Pressure (Above) indicates the extent to which, at wave 1, the respondent scored above the mean of the small group of which s/he was part on the dependent variable in question. Social Conformity Pressure (Below) indicates the equivalent for individuals below the mean.

Significant coefficients at $p=0.05$ or better in **bold**.

*Dropping significant non-endogenous predictors.

TABLE 9   *Models of Change in Collateral EU-Related Dependent Variables Wave 1 to Wave 4; Deliberators Only*

	European Identity		EU Policy Scope		EU Representation		EU Evaluations	
	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
European Identity, Wave 1	**0.69**	**0.000**						
EU Policy Scope, Wave 1			**0.60**	**0.001**				
EU Representation, Wave 1					**0.45**	**0.041**		
EU Evaluations, Wave 1							**0.45**	**0.000**
Change in Knowledge waves 1–4	−0.02	0.762	0.08	0.060	−0.01	0.856	0.10	0.064
Group Discussion Quality	0.24	0.425	0.05	0.799	−0.19	0.176	0.10	0.516
Social Conformity Pressure (Above)	−0.28	0.080	−0.13	0.570	−0.11	0.579	−0.20	0.214
Social Conformity Pressure (Below)	−0.30	0.205	−0.34	0.068	−0.13	0.602	0.15	0.382
Clarify thinking – experts	0.12	0.134	−0.01	0.903	**0.07**	**0.016**	−0.05	0.308
Clarify thinking – politicians	−0.07	0.346	−0.01	0.652	−0.02	0.388	0.02	0.715
Clarify thinking – other participants	−0.08	0.237	0.03	0.341	0.02	0.566	0.09	0.111
Clarify thinking – documentation	0.03	0.772	0.06	0.146	0.02	0.650	0.06	0.152
Constant	1.20	0.585	−0.39	815	2.22	0.166	2.45	0.079
Weighted N	347		347		347		347	
Adjusted R^2	0.38		0.22		0.24		0.45	
Adjusted R^2 *	n/a		n/a		0.23		n/a	

Note: Data, weighting, estimation as in Table 8. * Dropping significant non-endogenous predictors.

even if there were no real relationships in the data, to observe one significant effect for every twenty effects estimated. With seventy-two estimated effects, we would therefore expect, even if there were no 'real' effects operating in the data, to observe between three and four significant effects. Viewed in this light, the four significant mechanism effects observed in Tables 8 and 9 look singularly unimpressive. In short, notwithstanding the efforts made here to identify the mechanisms that underpin the effects of Deliberative Polling, the broad conclusion suggested by the evidence is that, whatever it is that produces the observed – and real – DP effect, it is not any of or any combination of increased knowledge, discussion quality, social-desirability pressure or the clarifying effects of exposure to documentation, experts, politicians and other DP participants.[36]

[36] One of the surprising features of the analysis presented above, given the findings of other DP studies, is the apparent absence of an explanatory role for changes in political knowledge. Previous studies (e.g. Fishkin and Luskin, 'Experimenting with a Democratic Ideal'; Fishkin and Luskin, 'Broadcasts of Deliberative Polls'; Luskin, Fishkin and Jowell, 'Considered Opinions') have found that increases in knowledge constitute an important part of the DP effect on participants' attitudes. It could be argued that the lack of knowledge-effects observed here results more from a failure of model specification than from a lack of 'real effects'. There are two important ways in which the specification and estimation of possible knowledge effects here differs from those used in other studies.

First, increases in knowledge here are captured explicitly in terms of observed individual-level change, rather than in terms of the 'implied change' that results from the use of a lagged dependent variable model. In 'Broadcasts of Deliberative Polls', Fishkin and Luskin argue that post-test knowledge *level* is the best proxy for knowledge *change*. Here, this would imply that knowledge change could best be measured using the level of knowledge at wave 4 rather than the change in knowledge between waves 1 and 4. The use of this term 'level' is typically justified on the grounds that people with high knowledge at time 1 by definition cannot increase their knowledge level at time t_2 as much as can people with low knowledge at time t_1. Moreover, with a lagged dependent variable this 'level of knowledge at wave 4' specification can in any case be regarded as capturing the *implied* effects of knowledge change. However, it should be noted that this is only the *modelled* effect of change – not the actual effect of observed individual change. In any event, substituting 'wave 4 level of knowledge' for the 'change in knowledge between waves 1 and 4' in the models reported in Tables 8 and 9 produces virtually identical results to those reported in those two tables. Using the 'level of knowledge at wave 4' specifications, the knowledge term fails to produce a significant effect for almost all the dependent variables examined here. The only exception is in the immigration importance equation, where the knowledge term produces a significant (but negative) effect. All of this suggests that the use of an explicit knowledge change measure in Tables 8 and 9 does not mask effects that would be otherwise revealed if an implicit change (wave 4 level) measure were used instead.

A second way in which knowledge effects could be specified and estimated is to regard them as interaction effects rather than as simple direct effects. This is the approach taken in Fishkin and Luskin's 'Broadcasts of Deliberative Polls'. The intuition is that the effects of knowledge acquisition on attitudes vary according to the levels of other explanatory variables in the model. Thus, for example, it might be expected that the effects of inceased knowledge on people's climate change positions will be greater if there is a relatively high (as opposed to a relatively low) level of small group discussion quality. The logic of this approach is that we should be more interested in the effects of the *interactions* between knowledge and each of the other explanatory variables in the model than in the direct 'main' effects of increased knowledge itself. This approach is tested explicity in the results reported in Tables A1 and A2 in the Appendix. The tables are based on the 'full' model specifications as shown in Tables 8 and 9, but with all possible interactions between knowledge change and the core explanatory variables added to the models. The tables report the results for the interaction effects only. (The pattern of 'main' effects remains virtually identical to that reported in Tables 8 and 9.) The results lend little credence to the idea that knowledge change effects can be discerned using this interaction approach. Across the two tables, only four out of fifty-six possible knowledge interaction effects are statistically significant – barely more than would be expected on the basis of chance alone. This again suggests that it is *not* the particular model specification design adopted here that is somehow masking the 'real' 'knowledge acquisition' mechanism that underpins the DP effect.

SUMMARY AND CONCLUSIONS

Deliberative Polling is a specialized form of social scientific enquiry that is increasingly being used by public authorities and policy makers to inform their decision making. The research enterprise analysed here constituted a classic Deliberative Poll. It involved comparing the changing views of a randomly selected group of deliberators with those of a randomly selected control group of non-deliberators. Since both the deliberating test group and the non-deliberating control group were designed to be representative of the population of the European Union, the results should in principle be generalizable to the population of the European Union as a whole. The core result was that deliberation over two important EU-wide issues – immigration and climate change – did indeed change the views of the deliberating test group. In comparison with the control group, deliberators, perhaps unsurprisingly, attached increased importance to the two issues after the DP event. Deliberators also, on average, became more *liberal* on the question of immigration – they became more tolerant of immigrants – and more *green* on the issue of climate change – they became more prepared to accept policies to combat climate change even though they might damage the economy. There was also a collateral DP effect on deliberators' views about the European Union itself: they increased their sense of European identity; they felt increasingly satisfied with democracy at the EU level; and they developed more positive evaluations of the European Union itself. Again, given that the DP event had brought together participants from all over the European Union in a friendly, well-organized and congenial atmosphere, perhaps this collateral effect might be thought unsurprising.

That the Deliberative Poll produced this kind of result, with significant attitude change associated with the test group but with not the control group, is consistent with the findings of previous such Polls. The substantive implication here is that EU policy makers could probably assume that, if there were more public deliberation on the issues of immigration and climate change, European public opinion would be more liberal and greener than it is at present. But is this implication justified? The second, and most important, part of the analysis conducted here involved trying to assess what the actual mechanisms of attitude change among the deliberators might be. Our analysis of the comparison between the test and control groups showed that weighting and clustering the data and controlling for a range of theoretically relevant demographic and other dispositional variables continued to produce findings consistent with a strong 'DP test' effect. Taken together, these findings suggest, however, that it is *not* the distinctive character of the deliberators themselves that produces this DP effect. Our analysis then went on to make use of the individual-level and group-level variance in key measures associated exclusively with the test group. This stage of the analysis involved testing four further hypotheses about the possible mechanisms that could account for the deliberators' changing attitudes: increased knowledge, discussion/deliberation quality, social conformity pressure and the role of documents, experts, politicians and other participants in clarifying people's thinking. The key finding in this connection is that none of these hypothesized alleged mechanisms successfully explained why deliberators changed their views during the deliberative process. A small number of the putative mechanisms appeared to have significant statistical effects in a small number of very limited contexts, but these statistical effects were fragmented and so infrequent as to be consistent with random variations in the data.

This leaves the observed 'effects' of deliberation almost entirely unexplained. The analysis shows what seems *not* to matter – knowledge, discussion quality, group conformity and the role of other actors in clarifying thinking – but it does not show what

does matter. On the face of it, this appears a very negative conclusion. But it is nonetheless an important one. The fictional detective Sherlock Holmes allegedly claimed that when all the plausible explanations of a crime have been eliminated then only the implausible remain. The analogy here is that if individual-level and group-level explanations comprehensively fail to identify the underlying mechanisms of the DP effect then we need to look elsewhere. One possibility that perhaps needs to be considered systematically in the course of future DP studies is that, even though there may be no small-group social desirability effects in Deliberative Polling, there may be present a much larger, event-level social-desirability effect. DP events are generally organized by liberal-minded academics and like-minded collaborators, who undoubtedly strive to be as neutral as possible in their dealings with DP participants. It is conceivable that unintentionally this liberal, even slightly left-of-centre ethos pervades DP events. The strong collateral effects of increased sympathy for the European Union that were observed in this study are certainly consistent with the idea that the ethos of the 2009 event and the character of organizers themselves may have had unintended, diffuse effects on deliberators' attitudes. There is clearly scope for the construction of an experimental DP design that consciously attempted to vary the overall presentational ethos of otherwise identical DP events. Any differences in observed outcomes under experimental conditions such as those would provide a powerful insight into the possible explanation of the DP effect. It might turn out that the observed 'DP effect' was not a DP effect at all but a DP-event effect, that the key causal variable was not the process of deliberation itself, but the style, values and outlook of those organizing the deliberation. There are certainly precedents for such findings in other social science research.

Finally, and more optimistically, it should be stressed that the experiment reported here was only one of a wider set of experiments conducted as part of the Europolis Deliberative Poll. The analysis here has focused solely on the comparisons that can be made across waves 1 and 4 of the Europolis data and across the narrowly defined test and control groups. The broad conclusion that must be drawn from the analysis reported here (with the possible exception of Zaller's observations about knowledge acquistion and attitude reinforcement) is that the mechanisms of attitude change at the Europolis DP event have yet to be identified. This said, further analysis of the much more detailed measures that were collected during waves 2 and 3 of the DP event may well reveal far more about the character of and causal mechanisms underlying attitude change among its participants.