

DELIBERATIVE POLLS: TOWARD IMPROVED MEASURES OF “INFORMED” PUBLIC OPINION?

Vincent Price and Peter Neijens

ABSTRACT

New research techniques have recently been developed to gather measures of public opinion that is better informed or more deliberative than that recorded in typical mass opinion surveys. These techniques include deliberative polls, educational surveys, and citizen planning cells. In view of what they set out to accomplish, what can we say from a scientific perspective about the utility of these methods? How are we to best interpret the data they produce? To address these questions, this paper reviews several of the most prominent and well-developed examples of deliberative or educational polling. We argue two main points. First, these new methods of assessing public opinion must be evaluated in terms of *specific* quality criteria that apply to different phases and/or participants in the democratic decision making process. Some techniques attempt to maximize several distinct qualities at once, making it difficult to identify specific objectives for evaluating success. Second, at least five important core methodological elements of educational or deliberative polls can be identified, each of which can theoretically alter results. To date, however, data bearing upon the effects of these methodological elements are in limited supply. Lack of knowledge about how method influences individual and collective opinion outcomes thus renders several of these techniques problematic.

Survey researchers and analysts of public opinion have long debated the conceptual relationship between public opinion and mass survey data. Although public opinion was classically conceived as an emergent product of widespread popular deliberation and discussion (e.g. Bryce 1888, Blumer 1946, Habermas 1989), many have come over time to accept mass survey data as its standard rendering (Key 1961, Converse 1987). However, in the face of evidence that opinions given to pollsters and survey researchers are often unorganized, isolated, uninformed responses (e.g. Bishop *et al.* 1980, Neuman 1986), arguments over

* Portions of this manuscript were presented at the World Association for Public Opinion Research thematic seminar on Quality Criteria in Survey Research, Cadenabbia, Italy, June 1996. Work on this project was supported in part by the Marsh Center for the Study of Journalistic Performance in the Department of Communication Studies at the University of Michigan.

the adequacy of mass survey data as indicators of deliberative public opinion persisted.

In recent years 'deliberative polls' and variants of focused group discussions have been advanced as supplements or alternatives to conventional mass opinion surveys. These new efforts are intended to correct perceived problems of superficiality in mass opinion data and deficiencies in public communication. Some researchers have attempted to assemble scientifically representative samples of citizens who are immersed in an extensive program of deliberation and discussion, in order to produce measures of better-informed public opinion grounded in meaningful public discourse (e.g. Fishkin 1991). Other researchers have proposed less drastic modifications of standard survey research techniques in 'educational surveys' that attempt to inform respondents about complex public issues in an effort to gather assessments of more thoughtful opinion (e.g. Neijens 1987).

How can the success of these efforts to gather measures of higher quality public opinion be judged? We attempt here to address the following questions. First, to what ends do these new methods aspire? What quality objectives, in particular, do they have in mind? Second, in view of what they set out to accomplish, what can we say from a scientific perspective about the utility of these methods? How well have they been evaluated in light of their quality objectives? How are we to best interpret the data they produce? What methodological conditions must be met to permit sensible interpretation of results? Since these techniques all, in one way or another, involve manipulation of respondents, what strictures must accompany interpretation of the resulting data?

NEW METHODS FOR ASSESSING PUBLIC OPINION

In a companion piece (Price and Neijens 1997), we have argued that conceptions of quality in public opinion are inextricably tied to broader conceptions of quality in democratic decision making, a complex process involving multiple phases and collective participants. Judgments of quality in public decision making can be conceptualized along a variety of dimensions.

Decision theorists, who systematically examine and seek to improve human decision making, have suggested at least two dimensions relevant to judging the quality of any decision (e.g. Rohrman 1986). One can judge the quality of the decision making *process* employed, or the quality of the decision *outcome* that is eventually attained. In the case of collective decision making, one can also apply quality considerations at both the *individual* level (e.g. How well do people understand the problems or proposed solutions at hand? Do they evaluate them rationally? Do they hold views that are in their best interests?); and at the *collective* level (e.g. How completely has a society examined its options?

How rationally is the collective choice determined? Is the choice in the best collective interest?). We can thus evaluate the quality of public opinion along each of these dimensions, which we have attempted to catalogue (a descriptive review of these criteria is provided by Price and Neijens 1997).

Outcome-oriented criteria focus on the quality of opinions and decisions themselves. Many of these—for example, opinion stability, consistency, conviction, and knowledge of and acceptance of a decision's consequences—can be applied to both individual opinions and to aggregate public opinion. Other criteria—for instance the representativeness of an opinion—make sense mainly in connection with collective opinion. Process-oriented criteria can similarly be applied at both the collective and individual level. These include the extensiveness of the information search supporting an opinion, the care of deliberation and analysis, the extent to which the opinion formation process has been extensive and wide-ranging, free of censorship and independent of social pressures. At the collective level, the decision making process should ideally represent those affected, generate many different viewpoints and decision options, and be perceived as legitimate by all participants (see Price and Neijens 1997).

Deliberative polls seek to maximize select quality criteria, as we will soon illustrate. But it is important to recognize that they are not the first attempts to do so. Indeed, deliberative polls historically follow a long line of efforts by pollsters and survey researchers to counter the tendency of mass opinion surveys to collect and disseminate opinions that may be ill-informed 'non-attitudes' or 'pseudo-opinions' (Price and Neijens 1997). Survey researchers and pollsters have long attempted to 'qualify' their measures of opinion in various ways. Examples include the use of 'don't know' response options, probes and question filters designed to reduce manufactured opinions, or the addition of measures of opinion intensity to accompany directional measures of opinion (e.g. Schuman and Presser 1981, Converse and Presser 1986). Application of 'likely voter' models to qualify poll results in election contexts (e.g. Traugott and Tucker 1984) or the use of a 'mushiness index' that attempts to assess the relative stability, consistency, and thoughtfulness of mass opinion data (Yankelovich 1991), are some other illustrations.

Survey and polling methods aimed at increasing the quality of public opinion range from these sorts of attempts at filtering standard opinion data (some of which have become widely accepted as good research practice) to more concerted efforts to inform or educate survey respondents—what we will call *surveys of informed public opinion*. Even more ambitious are new techniques that aim to inform through lengthy programs of active deliberation—what we will call *deliberative polls*. Here we will examine both approaches, selecting in each case several of the most prominent and well-developed examples for review and

analysis. Among surveys of informed opinion we will single out recent work by the Americans Talk Issues Foundation (ATIF, e.g. Kay *et al.* 1994b) and by Dutch researchers using the Choice Questionnaire (e.g. Neijens 1987). Among deliberative polls we will focus on the American and British Deliberative Polls organized by James Fishkin (e.g. Fishkin 1995) and the Planning Cell method developed in Germany by Peter Dienel (e.g. Dienel 1978, 1989).

AMERICANS TALK ISSUES

Since 1987 the Americans Talk Issues Foundation (ATIF) has conducted a series of 'educational' public interest surveys on a variety of issues. The surveys are designed and conducted by small, politically balanced teams of issue experts and public opinion researchers (Kay *et al.* 1994a). The teams work to insure that both the questions and the analysis are fair and balanced and, within the limits of telephone interviewing, provide accurate contextual and factual information.

The ATIF program of surveys has a number of quality objectives. One main goal is *consensus location*: 'A search for the most widely supported proposals by testing various features of the proposals separately and in combination, as well as verifying support by exposing respondents to pro and con arguments, including cost, benefits, and probable consequences of the most preferred proposals' (Kay *et al.* 1994a). More generally, the ATIF surveys aim at a variety of desired outcomes: achieving *consensus* (a collective outcome), increasing the *extensiveness of the information base* (an individual or collective outcome), and encouraging respondents to confront and *accept the consequences* of their opinions (an individual or collective outcome). They seek to *optimize collective interests* and better *represent collective desires* (collective outcomes). In that they are intended to assist policy makers, they also aim to improve the *clarity of the path to action* (a collective outcome).

In terms of opinion-formation processes, the ATIF efforts aim at improving the *extensiveness of the information search* by political leaders as they formulate public policy proposals. The surveys are designed to *extend mass participation* in decision making by gathering public views of a broad range of policy options and potential solutions to social problems. Conventional opinion polls and surveys assess popular responses to a narrow range of proposals favored by elites, in the latter stages of decision making. In contrast, the ATIF program seeks ways of injecting mass input during the earlier, developmental phase when policy options are first being assembled (see Price and Neijens 1997).

Some examples will illustrate how the surveys are conducted. In March 1993 and January 1994, ATIF examined public responses to a total of 50 proposals for government reform (ATIF Surveys 22 and 24). Three different approaches

to gathering opinions were employed. First, some of the proposals were presented in pairs or triplets, randomly ordered, of slightly different versions (e.g. 'limiting terms of office of members of Congress and other elected officials,' versus 'limiting terms of elected officials, registered lobbyists, and government bureaucrats,' (Kay *et al.* 1994a, p. 2). This was intended to allow analysts to detect the impact of small variations in wording, ideas, or the possibility of bias (p. 3). The second approach involved asking opinions on a baseline proposal (e.g. 'partial public funding to candidates running for Congress') followed by a series of limiting conditions (e.g. 'How would you rate this proposal . . . if you knew that: . . . Funding would come in part from a new check-off box on IRS tax returns'; Kay *et al.* 1994a, p. 5). The third approach consisted of asking respondents their opinion on a particular proposal (e.g. to 'require candidates for Congress to raise, at least, one-half of their campaign funds from individual voters in their districts') followed by a series of *persuasive arguments* supporting and opposing the proposal, most citing one or another desired or undesired consequence should the proposal be enacted. Respondents were asked to evaluate how convincing they found each argument to be. After these ratings, they were once again asked their opinion of the proposal (e.g. 'Now, that you have considered all these arguments, rate the original proposal again . . . Do you favor or oppose making this proposal into law?'; Kay *et al.* 1994a, p. 7). Thus, the ATIF formats are designed to assess what difference it makes for measured public support of particular proposals if survey respondents are fully aware of multiple options and encouraged to consider possible outcomes of each.

THE CHOICE QUESTIONNAIRE

The Choice Questionnaire (Neijens 1987, Neijens *et al.* 1992) resembles in some respects the third approach employed by the ATIF program and shares many of its quality objectives. However, it eschews the argumentative format used by ATIF and employs a more detailed, face-to-face survey format.

The Choice Questionnaire also limits its efforts to improving the evaluation phase of collective decision making rather than attempting to inject mass opinion into an earlier, policy-development stage. It attempts to provide citizens with a broad base of reasonably objective information, summarizing a full range of viable policy options and the probable consequences of each, provided by technical and policy experts. One of the Choice Questionnaire's distinguishing features is that it aims to insure, by design, that respondents *fully understand the consequences of various options*, think *carefully* about them, and *make trade offs among consequences* when deciding their choice. In doing so, the Choice Questionnaire provides a more highly structured information and response format than do the ATIF surveys.

The instrument was developed for use in connection with the General Social Debate (GSD) in the Netherlands. The GSD, carried out from 1982 to 1984, focused on alternative national programs for electricity production. Initiated by the Dutch Government because of substantial extra-parliamentary opposition to the use of nuclear power, the GSD sought to involve the Dutch population in determining a national energy policy. A Steering Committee, consisting of widely respected members recruited from parties with different standpoints in the debate, was given the task of organizing the debate. The Steering Committee prepared an Interim Report containing information pertinent to the problems at hand, after consultations with interested organizations, popular action groups, and citizens attending organized public hearings and discussions. The Committee next attempted to gather the opinions of citizens from all strata of society. The Choice Questionnaire, which contained information based on the Interim Report, was one of the instruments used in this popular opinion-gathering process.

Following the work of decision analysts (e.g. Keeney and Raiffa 1976, Edwards 1977, 1983), the Choice Questionnaire provided encapsulated information about principal energy policy options and their consequences, as described by a number of attributes—for example, costs, environmental impact, and social and health-related consequences. Respondents were presented with information cards describing six alternative energy strategies, five relying upon different sources of power generation (natural gas, oil, coal, nuclear power, and wind energy) plus a conservation strategy. A set of six to nine consequences was offered for each strategy. When experts disagreed over potential outcomes, their conflicting views were each represented.

After reviewing the information, respondents had to select a *combination* of any three strategies, under a set of guidelines insuring that the three strategies together would provide sufficient energy. Twenty combinations were thus possible. Because of concerns about respondent 'information overload' (Slovic and Lichtenstein 1971), processing of the supporting material was aided by a serial evaluation procedure. Respondents were asked to evaluate the attractiveness of each consequence, with the expectation that they would absorb the information more thoroughly as a result of the series of judgments they had to make. Respondents were first asked whether they considered a given consequence to be important or unimportant. If they believed the consequence to be important, they were asked to indicate whether they considered it advantageous or disadvantageous, and to estimate the magnitude of the advantage or disadvantage. They were then asked to add up the advantageous and disadvantageous evaluations for each option, respectively (assuming that such a 'book-keeping system' would help them make their choice without having to recall all the information). They were, finally, asked to choose the preferred combination of

three energy strategies. No prescription—that is, no decision rule—was given for this final choice.

The procedures involved in application of the Choice Questionnaire, then, are more elaborate than those used in the ATIF telephone surveys. In a face-to-face general population survey, field interviewers gathered background information from respondents, presented the Choice Questionnaire, explained how to complete the evaluation packet, and assisted respondents in filling out a practice questionnaire. Interviewers then left respondents with a booklet containing the Choice Questionnaire, requested that it be filled out at home within a week, and later returned to collect it.

THE DELIBERATIVE POLL

More elaborate than either of the two surveys of informed opinion so far described is the Deliberative Poll (Fishkin 1991, 1995). According to Fishkin, ‘an ordinary poll models what the electorate thinks, given how little it knows. A deliberative opinion poll models what the electorate *would* think if, hypothetically, it could be immersed in intense deliberative processes’ (Fishkin, 1991, p. 81; italics in original). The basic plan is to select a national, random sample of the voting-age population and transporting them as ‘delegates’ to a single site for several days of debate and deliberation. The delegates debate issues with political leaders and each other, and are then polled on their preferences.

The general design is modeled on the ancient, Athenian practice of selecting, by lottery, citizen juries numbering around 500 to judge accusations or charges against political leaders. This sort of ‘court of public opinion’ enjoyed great power in Athens, even in certain cases the right to review and overturn decisions of the Athenian Assembly (Fishkin 1991, pp. 87–8).¹ The process is intended ‘to adapt the deliberative possibilities of small-scale politics to the problem of selecting candidates and launching issues in a large-scale nation state’ (Fishkin 1991, p. 8). The exercise would combine the thoughtfulness and depth of face-to-face politics with the representative character of a national event (p. 51).

The Deliberative Poll focuses on most of the same quality objectives as the ATIF surveys and the Choice Questionnaire, but *adds* to these the goals of advancing *discussion and debate* as both individual and collective processes, and to create a communicative forum that is fully *representative* of the population at large. The Deliberative Poll aims at improving the *extensiveness of the information search* among political leaders and the public alike. Like the ATIF surveys, it

¹ It is perhaps worth noting, (as does Fishkin 1991), that it was just such a randomly constituted citizen jury that condemned Socrates to death, a fact that underscores the complicated relationship between the quality of a decision making process and the quality of decision outcomes.

seeks to increase *mass participation* in the proposal development process, by creating an event that would allow ordinary citizens to penetrate the 'hocus pocus' and 'amplification' of press-mediated political debate and engage directly with their leaders (Fishkin 1991, p. 8). This discussion, argues Fishkin (1991, p. 92) cannot merely be one-way. 'The basic point,' says Fishkin,

is that deliberative opinion polls offer direct democracy among a group of politically equal participants who, as a statistical microcosm of the society, represent or stand in for the deliberations of the whole. The institution is, in that sense, a direct face-to-face society for its participants and a representative institution for the nation-state (1991, p. 93).

Fishkin's Deliberative Poll (see Fishkin 1994) was employed in Britain in April 1994, produced by Granada Television and *The Independent*, and was broadcast nationally by Channel 4 in May. Granada television had previously conducted similar exercises in connection with their *Granada 500* programs. A national sample of 896 citizens was drawn by the Social and Community Planning Research (SCPR) and interviewed on the question: 'Rising Crime: What Can We Do About It?' Each respondent was invited to attend a two-day meeting in Manchester, and 302 elected to do so. They received carefully balanced briefing materials, participated in small group discussion about crime and crime prevention, and addressed questions to competing experts and politicians. At the end of the session, they were again polled on their views.

To date the most ambitious application of the Deliberative Polling method has been the National Issues Convention (NIC) held in the United States at the University of Texas at Austin in January 1996. The event was supported by the Public Broadcasting System (PBS), the University of Texas, the Presidential Libraries and a number of other public and private sponsors. As with the British effort, the NIC was developed purposely as a media event, was well publicized by its organizers, and produced over five hours of television coverage aired nationally on PBS. To select delegates for the convention, the National Opinion Research Center (NORC) at the University of Chicago interviewed a national sample of 915 citizens, age 18 and older, in late November and early December of 1995. All respondents were invited to serve as delegates to the NIC in Austin. Participants' transportation and housing expenses were paid, and they received a \$300 honorarium and \$25 toward the cost of telephoning their families during the convention. In cases of financial hardship, additional funding was provided. In the end, 459 delegates attended the event.

Delegates received non-partisan briefing materials prepared by Public Agenda and the National Issues Forums of the Kettering Foundation, on three topics: 'Mission Uncertain: Reassessing America's Global Role,' 'Pocketbook Pressures: What Should be Done about Them?' and 'The Troubled American Family:

Which Way Out of the Storm?’ Upon their arrival in Austin, delegates were randomly assigned into one of 30 small groups. On the first day of the program, they spent a total of nine hours (three on each topic) discussing the issues within their groups, led by moderators supplied by the National Issues Forums, and preparing three questions to ask presidential candidates. Over the following two days, delegates met as a single body with issue experts, four candidates for the Republican presidential nomination (three via satellite), and with Vice President Al Gore. On the last day they met within their smaller groups once again to discuss candidates and issues, after which they were polled individually via a survey instrument with 81 questions repeated from the initial survey. As a control, those initial survey respondents who agreed to come but could not participate and those who refused to participate were also interviewed using the same questionnaire. A separate random-digit dialing sample was also later contacted and interviewed, again with the same post-convention questionnaire.

THE PLANNING CELL

Although it was developed completely independently from and much earlier than the Deliberative Poll, the Planning Cell shares with the later many common quality objectives and methodological features. The Planning Cell (*Planungszelle*) is a technique devised over twenty years ago in Germany by Dienel (1978) for use in conjunction with technological planning. As with the Deliberative Poll, the technique aimed at increasing the possibilities for popular participation in decision making. More specifically, it sought to improve the accuracy and efficiency of societal decision making on complex technological issues.

In general design, the Planning Cell is quite similar to the Deliberative Poll. A group of citizens are selected at random, given paid leave from their working obligations, and are engaged in collaborative, face-to-face planning sessions with other citizens with the assistance of technical advisors (Dienel 1978, Renn *et al.* 1984). The seriousness of the planning process is underlined by an honorarium paid to citizens, who are assigned the esteemed role of ‘consultant’ in public planning (Renn *et al.* 1984). Unlike the Deliberative Poll, however, planning cells are designed to be limited in size (roughly 25 persons per cell); and rather than being focused around a single, full-sample meeting for the purposes of broadcast to a national audience, planning cells are intended to be replicated a number of times, in a number of locales, in order to increase their validity and legitimacy. Dienel (1989) reports that the technique has been employed in dealing with local problems—city planning, highway routing, the planning of recreation areas, choosing local energy sources and siting of waste disposal facilities—and with national and regional issues—social decisions on energy planning or implementation of new information technologies.

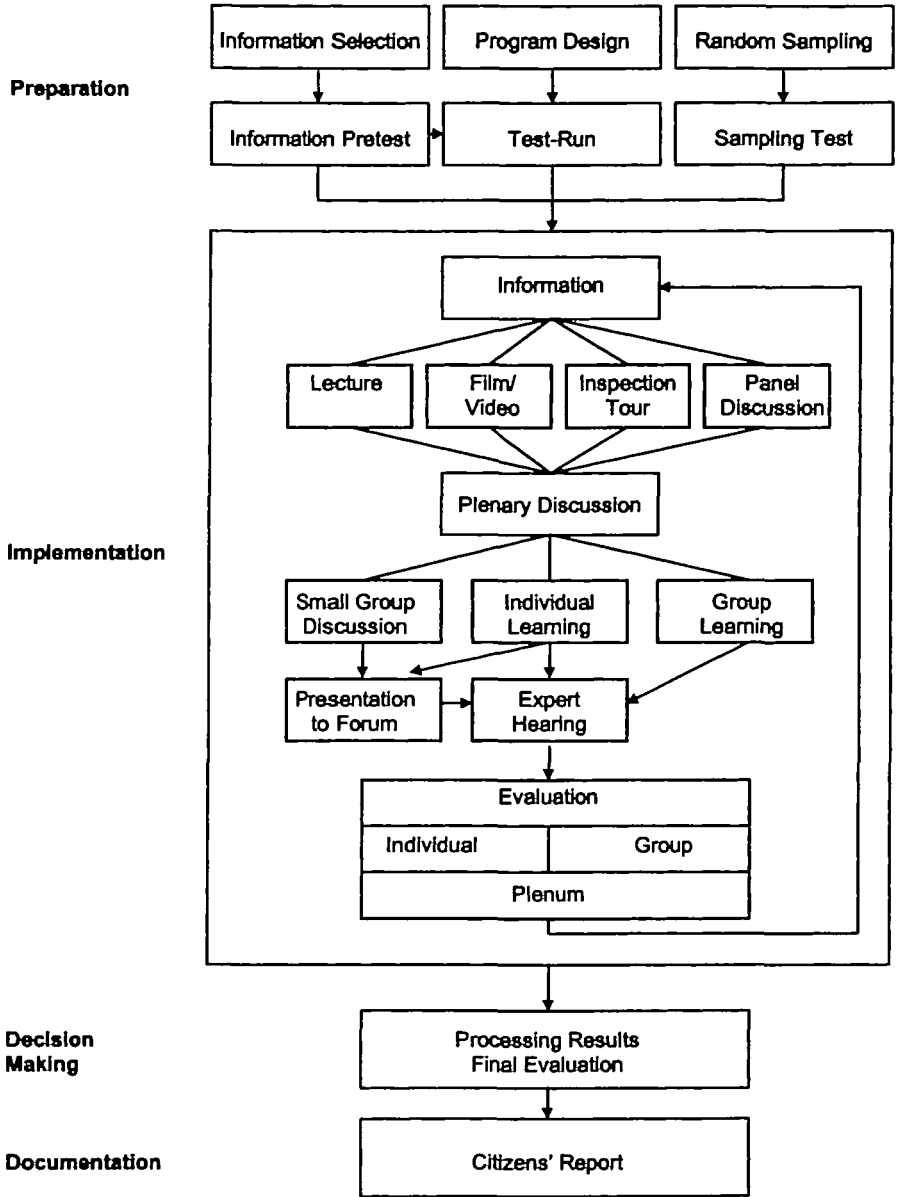


FIGURE 1 Design of the Planning Cell (taken from Renn *et al.* 1984, p. 18).

To guarantee replicability and comparability of results, the Planning Cell procedure is more highly standardized than the Deliberative Poll. Figure 1, taken from Renn *et al.* (1984), illustrates the general design. In the preparation

stage, informative materials are prepared by scientists of differing points of view and pre-tested; a general program of citizen deliberation is designed; and random selection procedures are developed and pre-tested. A test run is then conducted to work out any procedural difficulties. As implemented in a study of public energy supply systems in Germany (Renn *et al.* 1984, Diemel 1989), the planning cell programs involved alternating periods of information provision, discussions, small-group working sessions, and survey evaluations. To prevent the emergence of group hierarchies, membership in small-group sessions is rotated from session to session (Diemel 1989).

The first day of the four-day program involved inspection of a local energy plant, followed by group discussion of local energy problems, a lecture, lunch, group discussion of national energy problems, and a closing session introducing eight major criteria (e.g. cost, supply, economic effects, environmental impact, health and safety, social impact, political impact, and international implications) and nearly 40 minor criteria for evaluation of alternative energy sources. The second day involved a series of sessions on major energy alternatives built around expert presentations (live or on video), each followed by survey questionnaires, and closed with small groups of 3–5 citizens attempting to apply the evaluative criteria to each energy alternative (e.g. nuclear, coal, hydro, oil, solar and wind power). Later, small-group representatives presented and explained their overall evaluations to the entire group. Day three involved more such sessions, followed by meetings with politicians from the German Parliament representing major political parties. On the fourth and final day, participants each generated their own list of evaluative criteria, then discussed the local energy situation again, and finally spent a closing session individually evaluating energy alternatives both intuitively and by application of a calculated weighted sum of criteria judgments applied to each alternative. They also collected political arguments to ‘sell’ their own decision to the general public, evaluated their entire experience as a ‘consultant’ and were offered their honorarium check and a celebratory glass of champagne. Between June 1982 and April 1983, four test sessions and 20 full-scale planning cells were conducted in seven urban, suburban, and rural locations across Germany. A total of 427 citizens participated, in cells ranging in size from 12 to 32.

ASSESSING THE NEW DELIBERATIVE METHODS

Up to this point, then, we have described four noteworthy attempts to gather measures of informed public opinion. Two of these—the ATIF program of educational surveys and the Choice Questionnaire—involve modest adaptations of standard survey methodology, while the Deliberative

Poll and the Planning Cell chart a somewhat different course, attempting to constitute representative deliberative caucuses, or 'citizen juries' through random selection procedures.

Each approach attempts to advance a variety of specific qualities in public opinion and its role in collective decision making. The Choice Questionnaire aims at the most limited range of objectives, seeking to expand the information base of survey respondents, force them to confront and trade off consequences of alternative proposals or policies, and thus to gather more thoughtful and consistent individual opinions that, when aggregated, better optimize collective interests and collective desires. The ATIF survey program shares these objectives, adding to them an interest in injecting mass input into elite development of policy options. The Deliberative Poll and the Planning Cell attempt the broadest range of objectives, aiming at creating laboratory-type conditions that permit direct or participatory democracy to operate.

These are each impressive and potentially rewarding ventures for public opinion research. But good intentions are not enough. How are we to judge the relative success of these ambitious efforts? On the one hand, they do seem viable as possible ways of registering more informed public opinion than that offered by typical polls or surveys. On the other hand, they each clearly *manipulate* respondents to a significantly greater degree than do typical surveys and thus raise important questions about their methodological effects.

We should note at this juncture that methodological effects are not in and of themselves undesirable. On the contrary, *every* measurement (a method for systematic observation) produces a methodological effect (that is, a response from the phenomenon observed). That is precisely the intention. But valid measurement requires that we understand and intend the effect observed. How well do they match objectives? Having discussed the various objectives involved, let us now take a closer look at the specific techniques employed with the aim of learning how well they match intentions.

Analytically, we can identify several main methodological dimensions distinguishing these techniques. In conducting a survey of informed public opinion or a deliberative poll, a researcher must, as a minimum, decide: (1) the extent to which the researcher provides substantive information to respondents; (2) the way that information is structured (e.g. whether it is overtly persuasive or not, whether it attempts a balanced exposition of 'expert' opinion); (3) the extent to which a representative sample is intended, and the means by which one is gathered; (4) the extent to which the technique encourages or allows private deliberation, or public discussion, or both; and (5) the degree to which and manner in which the decision making process of the respondents is structured by the researcher. What evidence do we now have about the impact of any of these five methodological elements?

PROVISION OF INFORMATION

Each of the techniques reviewed provides substantial information to respondents in the hope that this will alter, for the better, the opinions gathered. Demonstrating that providing information in fact alters expressed opinions is thus a logical first step to demonstrating the utility of these techniques. On this count, there is clear evidence that changes are induced. Fishkin (1994) reports several clear changes in opinions of participants in the British Deliberative Poll on Rising Crime. Generally, support for the use of imprisonment for fighting crime was reduced, while support for alternatives like compulsory training and counseling increased, particularly for juvenile offenders. More highly educated participants demonstrated larger changes than did their less well-educated counterparts. Among the pre/post-deliberation changes identified in the 1996 American Deliberative Poll were a decline in support for a flat-tax proposal and increased support for turning federal low-income security programs over to state governments.

ATIF survey findings illustrate the effects of providing even limited information to respondents. In January 1994, 26 percent of Americans surveyed supported a general proposal for partial public funding of Congressional elections. However, more than half (56 percent) of respondents supported a more complex proposal, namely, that:

Funding would come from the radio/TV industry who would provide free air time for debates required of candidates receiving these funds AND who would offer reduced advertising rates to these candidates, low enough to substantially cut their campaign costs (Kay *et al.* 1994b, p. 5).

These data suggest large information effects, but they are not especially compelling. The additional 'information' offered in the various ATIF proposals is not actually background information that might illuminate respondents' opinions of the baseline proposal (for example, information about how much incumbents and challengers typically spend on Congressional elections) but rather alterations to the proposal itself—thus, respondents are presented with *different* questions, not in fact with different pieces of information relevant to the *same* question. So it is difficult to know whether changes that result are a product of introducing background information or of question framing.

Other data provide more convincing evidence of the impact of providing information. In four cases on the same January 1994 survey, respondents were asked their opinions on a given proposal, then a brief series of supportive and opposing arguments, and were then asked their opinion once again. Before/after differences in support of these four proposals ranged from -2 percent to +9 percent, with the largest difference reported for a proposal limiting outside-of-district campaign contributions (Kay *et al.* 1994b, pp. 6-9). In three

of these cases, ATIF also conducted experiments that randomly altered the order of arguments and the inclusion/exclusion of the original opinion question. Results indicated no significant primacy or recency effects due to the order of presentation, nor of 'lock-in' effects due to respondents maintaining the opinions they expressed prior to receiving the arguments (Kay *et al.* 1994b, pp. 39–41). As noted above, similar changes in opinion have been registered by Deliberative Polls, and even in cases where little net change in the balance of opinion has been produced, analysts have pointed to large amounts of gross change (Fishkin 1994).

Demonstrating some change in opinion, however, only establishes that one obtains a distinctive result when providing information to respondents. It is another matter entirely to establish that this result is distinctively 'better.' Data bearing on this question are in short supply. Among many possible quality criteria (Price and Neijens 1997), only the *consistency of opinions* and their *correspondence with understood consequences* have been empirically assessed and related to the provision of information. In examining an application of the Choice Questionnaire to citizens' views on national energy policies, Neijens (1987) studied the consistency between respondents' preferred choices among energy alternatives and their evaluations of the positive and negative consequences of each alternative. He found that 68 percent of the respondents who filled out the Choice Questionnaire completely made a choice that was consistent with their evaluations of consequences (compared with only 37 percent responding to a standard opinion survey providing no information). Respondents did so by trading off desired and undesired consequences of available options in a compensatory way. Cognitive ability and involvement with the issue were, as one might expect, associated positively with the consistency of the choice. But together these variables explained no more than 7 percent of the variation in consistency.² After examining respondent characteristics, prejudices, and pre-existing levels of opinion consistency, Neijens (1987) concluded that between 24 percent and 36 percent of the respondents were aided in their decision making by the Choice Questionnaire—that is, were influenced by the information provided to change their choices, bringing them into line with their evaluations of desired or undesired consequences (see also Van der Salm *et al.* 1997).

² Rates of consistency were calculated based on an equally weighted 'addition of utilities' model, but the same pattern holds when other models are applied as well (Neijens 1987, p. 185): A general assumption underlying all such calculations is that respondents' evaluations of the consequences were actually based on the information provided. If this were not so, consistency between choices made and evaluations would be of little value. For example, halo effects occur when people who favor an option rate it highly on all desirable attributes, while people who dislike the option rate it poorly (Beckwith and Lehmann 1975). In this case, the consequences of an option are judged in the light of pre-existing preferences and become mere proxies for the choice itself. Analyses, however, showed no evidence of halo effects (Neijens 1987, Neijens *et al.* 1988). Other experimental research further demonstrates that changes produced by the Choice Questionnaire stem from, and agree with, the information provided (Van der Salm *et al.* 1997).

Providing information about social, environmental, or political consequences of some policy to survey respondents, then, does appear to elevate the consistency between their stated preferences and their evaluations of those consequences. With these sorts of findings, one can begin to build a case for increased quality, at least as it pertains to acceptance of consequences and consistency. But at the same time, the provision of information—precisely because it *is* so influential—injects potential quality problems, particularly as they relate to *freedom from control*. Who is to decide what information is germane to a decision? In discussing the planning cell, Renn *et al.* (1984) note that there is a great danger of manipulating the results through the selection and provision of information. ‘The quality of the output relevant to planning,’ they observe, ‘is decisively dependent upon the amount of given information and the preliminary work by the sponsoring agency’ (p. 19).

Work in cognitive psychology shows that the *formulation (or framing)* of a problem may have decisive effects on a subject’s perceptions and choices (e.g. Tversky and Kahneman 1982). This problem is not only apparent in the provision of information; it is also observed in the formulation of questions in more standard survey questionnaires. A large number of ‘irrelevant’ factors, such as the order in which the information is presented, the choice of anchor points and response modes, etc. may all play a role in perception and decision processes (see, for example, Schuman and Presser 1981, Kahneman *et al.* 1982, Molenaar 1982, Converse and Presser 1986, Petty *et al.* 1987). The framing problem seems unavoidable. Information has to be formulated in a certain way, and there is not always an exclusively ‘correct’ way of doing so. This is true even when individuals attempt to formulate a problem for themselves (Fischhoff *et al.* 1980). Mass opinion surveys do not tap ‘unframed’ choices, but choices framed by a wide array of personal and environmental factors, not the least of which is media coverage of the issue in question (see Iyengar 1991, Price and Tewksbury 1997).

The question is whether providing much more information—as each of these newer procedures does—increases or decreases the power of the survey instrument to frame respondents’ choices. One might upon first reflection assume that the greater the ‘dose’ of information, the larger the probable framing effect. But the selecting of just a few bits of information, as is inevitably done in connection with crafting standard survey questions, might in fact place much more focused attention on limited dimensions of a problem. Provision of more information might conceivably mitigate framing effects, in that, when confronted by a large amount of information and many *differing* perspectives on a particular issue, people might have greater flexibility in choosing to frame the problem along whatever lines they prefer. That is in many respects the argument in favor of the Deliberative Poll—that it immerses delegates in such a rich

information environment that they are in some sense freer to choose their own 'terms of the debate.'

How should social problems be framed? 'Making that decision,' comment Slovic and colleagues (1980, p. 35), 'takes one out of psychology and into the domains of law, ethics, and politics.'

Extreme care must be taken to select knowledgeable and trustworthy designers and program coordinators. We cannot propose a general solution here, as a competent and credible program staff would have to be put together in consultation with representatives of the people who were to be informed. (Slovic *et al.* 1980, p. 18, see also Fischhoff 1984).

This requirement would seem to be met reasonably well by most of the methods reviewed above. In each case, care has been taken to select 'blue ribbon' planning committees that are both politically and scientifically 'balanced.' But the ultimate question is one of political legitimacy. Unfortunately, these well-intended efforts may be insufficient to guard against charges of bias leveled by those who find the eventual outcome of the exercise objectionable.

STRUCTURING OF INFORMATION

Implicit in the forgoing discussion are questions, not just about *what* information is provided, but also about the way that information is *structured*. Some of the techniques reviewed above (the ATIF program, the Deliberative Poll, and the Planning Cell) present information to participants in a persuasive context, in the form of arguments favoring or opposing particular proposals. These techniques not only create the opportunity for information effects, but also for *source effects* stemming from perceived credibility, trustworthiness, or expertise (see, e.g. Petty and Cacioppo 1981). By comparison, the designers of the Choice Questionnaire attempted to present information about estimated consequences of various policy options in a non-argumentative format and in that way to minimize such effects.

Little evidence is presently available to discern the effects of selecting one of these formats over another. But some results reported by Renn *et al.* (1984) are perhaps instructive (see also below). In their application of the Planning Cell to citizen views of energy policies, they noted that even after four days of discussion surrounding various consequences of alternative energy strategies, participants sometimes made final selections that deviated rather significantly from those implied by all of their comparative evaluations—regardless of whether or not their comparative evaluations were weighted to take account of those factors they deemed most important. Renn *et al.* concluded that non-cognitive factors, mainly loyalty to parties and other reference groups siding

with particular energy proposals, explain the discrepancies. When characteristics of a source of 'facts' are known, it is very likely that perceived credibility and trustworthiness enters into judgments about the veracity of those facts. Is this 'rational?' Would it be more rational to arrange a system by which the identities of all participants were obscured? This would focus all attention upon the *content* of messages, but would hardly constitute a 'public' forum in the commonly understood sense, or an uncensored and free-flowing debate. Quality criteria come into conflict. On the one hand, we often wish to insulate debate from social pressure, but on the other hand, we wish to make the decision process *social*. Ironically, we cannot have one without the other (Price and Neijens 1997).

What is clear is that the Deliberative Poll and Planning Cell, by subjecting participants to very complex argumentative stimuli, seriously confound the impact of information *per se* with a variety of social and normative influences (see the extensive literature on informational and normative social influence in small-group decision making, e.g. Deutsch and Gerard 1955, Burnstein and Vinokur 1977, Turner 1982, Moscovici 1985). Given their commitment to discussion and debate, perhaps they could hardly be expected to do otherwise. But it leaves us, scientifically, without any means of determining what the *unique* contribution of information might be. Even the ATIF surveys, in presenting lists of pro- and con-arguments for various proposals that are extracted from the arena of contemporary political debate, affords respondents the opportunity to easily identify the 'sides' of the issue—many arguments echo Republican or Democratic Party political rhetoric—and to use a 'who's for it/who's against it' heuristic in responding.

Turning just to informational influence, we can see that in any complex presentation of information a number of presentational factors can alter perceptions of the problem at hand or the particular 'pieces' of information given the greatest attention. These include the *order* of presentation, which can lead to primacy or recency effects, as well as the *structural format* employed. For example, information can be presented per proposal, or per attribute, or as an information matrix (proposals by attributes). Alternative proposals and their attributes may be presented in fully-factorial form, or in some reduced form where different attributes are presented for different options. Experimental decision making research has indicated that subjects do indeed adapt their information processing to the presentation format (Bettman and Kakkar 1977). In a face-to-face deliberative forum, the same presentational issues surface: Should each candidate appear separately before a citizen panel? Or should multiple candidates appear in debate format? Should all representatives of competing proposals be forced to address the same questions, or not? Social-scientific theory and research addressing each of these issues exist, but they

TABLE 1 Percentage of consistent decisions under four conditions—Choice Questionnaire^a

<i>Condition</i>	<i>Percentage offering consistent decision^b</i>	<i>N^c</i>
No information ^d	37	300
Information only ^d	48	299
Information + evaluations of the consequences	57	300
Information + evaluations of the consequences + overall evaluations	68	1243

^aData are drawn from Neijens (1987, pp. 184–7).

^bA decision to select a particular option was deemed consistent if it agreed with a respondent's summed positive and negative evaluations of consequences across options. By chance alone, only 5 percent of respondents would have offered such a consistent decision.

^c*N* includes only respondents who carried out the tasks given in the various conditions completely.

^dIn the 'no information' and 'information only' conditions, respondents completed their evaluations of consequences *after* selecting their favored option.

have not yet received systematic empirical attention with respect to the techniques reviewed here (excepting some order-of-presentation experimentation by the ATIF program, as noted above, and some pretesting work done in development of the Choice Questionnaire format).

Complicating methodological assessments profoundly, then, is the practice of offering information programs to respondents in all-or-nothing 'packages' that structure information in complex ways, combining a number of facets that could each have independent effects. Without experimental isolation of separate components of such a package, one has no analytical capacity for judging whether a particular effect stems from one or another aspect of the design.

The limited exception to this pattern is again the evaluative work completed on the Choice Questionnaire (Neijens 1987). As noted above, the large majority of respondents filling out the 'complete' version of the questionnaire (68 percent) made energy-policy choices that were consistent with their evaluations of the consequences of each energy alternative (compared with 37 percent in a standard survey format). Yet this change cannot be attributed entirely to the provision of information, since the complete version included not only information about consequences, but also a procedure asking respondents to make serial evaluations of these consequences when reading the information and a 'book-keeping' system for respondents to add up the advantageous and disadvantageous evaluations for each option to assist in an overall evaluation. Random subsamples completing different versions of the questionnaire indicate that *each* facet of the Choice Questionnaire's structure has significant, and in this case cumulative, effects (see Neijens 1987, Table 1). Compared to the baseline 'no information'

condition—where a majority of respondents actually give policy preferences that are inconsistent with their views about the consequences of various energy alternatives—each element of the Choice Questionnaire procedure increases the rate of opinion-evaluation consistency by about 10 percent. Other analyses of open-ended arguments provided by respondents for their choices also indicated that they *retained relatively little information* unless they completed the evaluation procedure.

REPRESENTATIVE SAMPLING

Each of the four techniques, in keeping with conventional opinion polls, aims at producing an assessment of public opinion based upon a representative sample of citizens. But variations in the nature of each technique render them more or less likely to succeed in accomplishing this objective. At one end of the spectrum, the telephone surveys conducted by ATIF differ little from most random-digit dialing surveys; at the other extreme, the Planning Cell and Deliberative Poll demand rather extraordinary contributions from participants (and attempt to match those demands with relatively extraordinary levels of subject compensation).

The Choice Questionnaire resembles most face-to-face general population surveys in many respects, but asks respondents to complete a set of procedures that are, by comparison, very demanding. In the project conducted for the Dutch General Social Debate, the Choice Questionnaire was completed, in its full version, by 81 percent of respondents who participated in the initial face-to-face interview. Seven percent had failed to complete any part of the questionnaire when interviewers returned to collect it, while another 12 percent partially completed the questionnaire or were eliminated because someone other than the eligible respondent completed it (about 3 percent). Analyses of non-response found that participation was not biased with respect to sex, place of residence, or the composition of the respondent's household. However, participation was systematically related to education and age. Participants were somewhat more educated and younger than the full sample of respondents, although the differences were slight. Similarly, those who were more involved in the energy debate, who reported greater understanding of the procedure and reading books more often were also slightly more likely to participate. On the other hand, the proportion of variance in participation explained by all these variables together was less than 10 percent, so the profile of participants in the Choice Questionnaire probably differs only to a limited extent from the profile of the Dutch population as a whole.

By comparison, rates of compliance with requests to leave home and participate in planning or discussion sessions lasting several days are, not surprisingly,

much lower. Out of the 869 British respondents surveyed by SCPR in the spring of 1994 and invited to attend a three-day Deliberative Poll on Rising Crime, only 302 (35 percent) accepted the offer. Still, comparisons of the group attending with the full sample indicated a very close match on demographic variables, measures of knowledge about crime, and concern about crime (Fishkin 1994). The recruitment efforts mounted by NORC for the 1996 National Issues Convention in Austin, Texas probably come close to defining the maximum effort that might be applied toward recruitment of attendees to a multi-day, nationally representative caucus or planning session. In that case, 50 percent of the respondents initially surveyed and interviewed elected to participate (representing an overall response rate of 36 percent, due to the fact that the initial survey garnered a 72 percent rate of response). As noted in connection with other techniques, the group of final participants was slightly younger and better educated than the full sample, and included fewer retirees and widows.

The problem, of course, is that matching variables in this fashion does not guarantee that participants are indeed representative in all relevant respects—some of which may not have been measured. Data from Renn *et al.* (1984) illustrate that close matches with population statistics on some variables is no guarantee that possibly important biases on other variables can be safely disregarded. Renn *et al.*, like the NORC planners of the National Issues Convention, report extensive efforts to garner random samples for their planning cells—including personal visits with participants and their employers, provision of baby-sitting services, and other herculean efforts. These efforts resulted in response rates of only 20 percent. Refusals were reportedly based upon family or job commitments, lack of interest, lack of satisfaction with the honorarium amount, or opposition to the procedure (but not perceived incompetence). There was, in spite of the low response rate, again a very close match of sample demographics with German population statistics (with the largest shortfall being an under-representation of persons older than 65—only 6 percent of the sample compared to 20 percent of the population). On the other hand, respondents showed a very clear bias in professional affiliation, with the sample attracting far too many white-collar workers, particularly government employees, and too many students (Renn *et al.* 1984, p. 28).

So deliberative techniques can come with significant sampling problems, and it is not entirely clear that they can be overcome even with enormous time and money (\$4 million dollars were spent on the 1996 NIC). Nevertheless, there is no doubt that the use of random sampling procedures can improve upon purely voluntary attendance. For example, the Dutch General Social Debate involved 18,000 participants in informational meetings and discussion meetings, 12,000 of whom completed survey questionnaires. But the profile of participants was far from representative of the Dutch population (Neijens 1987). They were

much better educated, less likely to live in large cities, and more likely to vote for 'radical' left-wing parties. Men and middle-aged people were also significantly overrepresented in the discussion groups. Along similar lines, Merkle (1997) has found that attendees of the 1996 NIC were more politically involved than were non-attendees (pp. 600–1, Table 2).

PRIVATE DELIBERATION/PUBLIC DISCUSSION

All of the techniques we have reviewed make concerted efforts to encourage private deliberation among respondents prior to registering their opinions, but once again the extent to which each technique encourages such deliberation clearly varies. The ATIF surveys attempt, in a telephone interview situation, to lead respondents through a series of questions structured in such a way that they help people 'think through' various consequences of policies; but they do not in fact leave respondents with any time to think. As with most polls and surveys, the ATIF techniques gather top-of-the-head responses (albeit in such a way as to give them a more thoughtful cast). Respondents in the Choice Questionnaire, on the other hand, were given a full week to read over the informational materials and to complete their booklets. So they are at least provided with the opportunity for private deliberation, whether or not they took full advantage of it. They were also permitted, if they wished, to discuss their task with others; but discussion was neither endorsed nor discouraged. What exactly is the impact of private deliberation *per se* on opinion formation? What would be the outcome of simply providing time for thinking about a problem without any information, as compared with providing information without time for deliberation (the ATIF surveys), or providing both information and time to consider it carefully (the Choice Questionnaire)? Although private deliberation is often assumed to improve opinions, some recent research in psychology on the effects of 'mere thought' (e.g. Wilson *et al.* 1993, Schooler and Wilson 1991) has in fact suggested that introspection can sometimes adversely affect the quality of preferences and decisions.

Both the Deliberative Poll and the Planning Cell, of course, aim to immerse participants not just in private deliberation but also in public discussion. Yet a wide variety of possible structural arrangements for discussion are possible, and they are certainly not equivalent. In the Planning Cell design, several sessions are deliberately set aside for individual deliberative work, small-group discussions (in a series of rotating, differently constituted groups of a half-dozen or less) and whole-group meetings numbering up to 30 participants. In contrast, the Deliberative Poll incorporates more large meetings, of 450 or so participants, placing participants into working groups of 15–20 that remain together for most

of the program. Perhaps even more critical is the fact that group activities in the Deliberative Poll are videotaped for national broadcast.

Unfortunately, data are not available to assess with any degree of reliability the differential impact of any of these fundamental design features, although there are very good reasons to suspect that each could well have important consequences. To take just a single example, although one might imagine that the addition of public discussion to private deliberation significantly alters the nature of participants' opinions, we are in fact unable to discern whether this is indeed the case, or to what extent it may be true. Take for instance the differences in opinion on the flat tax (in America) or on incarceration of criminals (in Britain) produced by the Deliberative Poll. Would comparable differences have been observed through an ATIF-style educational survey or the Choice Questionnaire? For that matter, what share of the effects produced might be traceable to the National Issues Forums information booklets provided and attendant private deliberation (which might well have occurred prior to any public engagement), rather than to the discussion *per se*? Again, when programs are offered in all-or-nothing packages, without incorporation of experimental methodologies to disaggregate effects and attribute them to various components of the package, we are left with little guidance for evaluating the validity of the exercise—or for improving it through modification.

If indeed we plan to ascribe political importance to decisions rendered by citizen juries, it only makes sense to examine the ways in which structural jury composition, or the deliberation procedures and 'rules of order' adopted, systematically alter decisions (see, for example, existing research on criminal and civil trial juries, e.g. Tindale *et al.* 1990, Tanford and Penrod 1986, Estlund 1994). The considerable research literature on small-group dynamics has recorded significant evidence of structural effects on group outcomes, so we have good reason to expect that the particular way in which deliberative interaction is organized will be consequential.

STRUCTURING OF DECISION MAKING PROCESS

Just as information, deliberation, and opportunities for interaction must be structured, so must the final decision making process employed by any respondent. Survey researchers are already familiar, for example, with response-option effects created by the manner in which respondents are allowed or encouraged to report their answers to questions (e.g. Schuman and Presser 1981).

The degree to which each of the techniques reviewed attempt purposely to structure the decision making of participants, and the ways in which they do so, again clearly vary. Most highly structured is the Choice Questionnaire,

where information presentation and the decision task are both organized according to theories of decision analysis (Keeney and Raiffa 1976, Edwards 1977). Not only is the information structured so as to draw attention to trade-offs between different desired and undesired consequences of various choice options, but the response format is designed to simplify and amplify that structure. As we have already seen (in Table 1) this structuring of the decision making process—serial evaluation of various consequences, followed by an additive book-keeping system prior to the final choice—had significant effects on the quality of opinions, at least as operationally defined.

Renn and colleagues, in their application of the Planning Cell to citizen views of energy policies, assessed participants' opinions on alternative energy programs in two ways—through intuitive 'holistic' evaluations and by calculating preferences based on the addition of weighted (or unweighted) respondent evaluations across eight different criteria (e.g. environmental impact, cost, social implications, etc). Some large differences in citizen preferences were reported. For example, 16 percent indicated an intuitive preference for a strong conservation/solar energy program, while calculated preferences indicated 37 percent favoring the same scenario; and while 41 percent indicated an intuitive preference for a moderate nuclear program, that figure dropped to 37.2 percent when preferences were calculated on the basis of weighted criteria ratings. As mentioned earlier, the authors concluded that loyalty to reference groups siding with particular energy proposals explained these discrepancies, although it is possible that it is also a product of the way respondents were asked to report their views.

Another important variation in the method of gathering opinions concerns whether people are asked to make *serial judgments* about various proposals independently, to *choose one option* from a set of competing proposals, or to *rank-order* proposals. There is an established literature on the differences between the preference structures revealed through rank-order questions and those based on serial ratings (e.g. Alwin and Krosnick 1985). Krosnick and Alwin (1988) offer some evidence that, when provided with serial ratings, survey respondents avoid making difficult trade offs by rating valued alternatives equally. Thus, forced rankings often supply a better view of respondent preferences. The issue with respect to collective decision making is fundamental, as it concerns the structure of collective and often incompatible options faced by the community. If it must choose *some* option for dealing with a pressing problem (e.g. reforming the U.S. social security system) with virtually all options having well-understood drawbacks, then offering survey respondents individual proposals in serial fashion allows them to reject proposals that they might otherwise see as the best among available options. In such a case, forcing respondents to choose between viable options (as in the Choice Questionnaire) may best represent 'considered' public

judgment—in the sense that it more directly forces people to confront the consequences of their choices (Yankelovich 1991).

ASSESSING METHODOLOGICAL EFFECTS ON OPINION QUALITY

The foregoing discussion has identified a host of methodological variables manipulated by these new research techniques, very few of which have been systematically examined to date. Clearly, isolation of each facet of these alternative methods through experimental design is in order. Even then, we would be left without much empirical guidance as to the quality of these methods unless specific opinion-quality criteria are themselves measured in conjunction with the application of each method—and preferably in such a way as to allow *comparative* assessments about the success of each approach. That is, we should ideally evaluate each technique, not just through comparisons to ‘conventional’ opinion surveys as a baseline, but to each other as well.

Again, we are struck by the extent to which operational measures of opinion quality are generally lacking, or if they exist, have yet to be analyzed and presented for review. Results from overtime measurements needed to assess the *stability* of opinions, once formed in response to these methods, have yet to be published (although follow-up surveys of participants were collected in connection with the Deliberative Poll).³ As our review here indicates, researchers implementing some techniques—most especially the Choice Questionnaire and the Planning Cell—have gathered sufficient data to examine some limited but key aspects of *outcome-oriented* opinion quality. In deploying the Choice Questionnaire during the Dutch Social Debate, its designers did in fact develop a rather extended program of evaluative research, including assignment of subjects to receive different components of the Choice Questionnaire (as noted above). The evaluation study was designed with specific and measurable objectives, focusing on (1) the extent to which non-response could be minimized and (2) the extent to which the information presented to respondents was in fact employed in making their decisions (Neijens 1987, p. 66). The latter was operationalized as the extent to which evaluations of the problem were consistent with the information provided, and as the extent to which respondents’ choices were consistent with the evaluations (see Table 1). Possible indirect effects of the Choice Questionnaire (e.g. stimulation of discussion or interest, or improvement of decision making skills) were, on the other hand, not examined.

³ A recent study by van Knippenberg and Daamen (1996) did assess the stability, over a span of two months, of the opinions given by respondents to the choice questionnaire. They found that more motivated and highly educated respondents showed greater opinion stability. Yet because they did not assess the stability of opinions given in response to a more conventional survey format, they could not evaluate the impact of the choice questionnaire itself, by itself or in interaction with motivation or education, on opinion stability.

Still, the data recorded by the evaluation research allowed at least some leverage on several quality criteria, including opinion *consistency*, *acceptance of consequences*, and extensiveness of the *information base*.

CONCLUSIONS AND FUTURE DIRECTIONS

Now that we have had occasion to explore these efforts to gather measures of informed public opinion, what are we to conclude about their quality? At this point, these ambitious and intriguing new attempts to modify polling techniques leave us with far more questions than answers. Taken together, they force us to take more seriously the connection between survey practice and the democratic functioning of society. Attacks on political surveys and polls, and complaints about the roles they play in political affairs, are none too rare. What is unique about these techniques is that they represent active efforts to translate those concerns into alternative research and measurement strategies. In altering standard ways of doing business, they present a useful challenge. They expand considerably the ways in which we can imagine assessing public opinion and incorporating it into collective decision making.

Certainly these techniques are—to varying degrees—different from conventional opinion polls. But are they better? Do they succeed in their efforts to measure informed public opinion? We do have some evidence at hand to support that claim for what we have here called ‘surveys of informed public opinion’. The strongest empirical case for increased quality can be assembled at this point for the Choice Questionnaire. This is so mainly because the researchers involved were able to articulate some fairly limited objectives—measuring popular selections among a set of competing policy alternatives with the understanding of necessary trade offs between potential consequences—and because they undertook an experimental evaluation of the effort enabling them to bring opinion–quality data to bear upon their stated objectives. The results of that evaluation, some of which we have synopsized here (see also Neijens 1987) are sufficiently promising that we might rightly feel encouraged about the prospects for gathering more ‘informed’ public opinion via such techniques. Of course, one may argue with the ‘correctness’ of the stated objectives themselves (e.g. with the focus on the consistency of choices with evaluations of various consequences, drawn from decision theory); but then the argument is not so much over the scientific validity of the data—over what they properly represent—but over whether the qualities achieved are in fact desirable.

What of deliberative polls? Do they offer readings of higher-quality public opinion? On the basis of their experience conducting 20 planning cells across Germany, Renn *et al.* (1984, p. 43–5) conclude that the program offered four major benefits: (1) elicitation of ordinary citizens’ preferences concerning very

complex technological decisions; (2) an opportunity to explore underlying reasons for citizens' attitudes and beliefs; (3) measurement of projected popularity of alternative government proposals; and (4) the prediction of opposition to specific objects or government policies. At the same time, they acknowledge as many drawbacks: (1) the inability of ordinary citizens to provide any new solutions; (2) the difficulty of eliciting careful trade offs between different advantages and disadvantages of complex alternatives (cf. the Choice Questionnaire); (3) problems of aggregating individual weights for different dimensions of a problem because of response sets, political group loyalties, or other distortions; and (4) limited political legitimacy due to the fact that such citizen jury outcomes 'are interpreted as manipulated as soon as they deviate from one's own opinion' (Renn *et al.* 1984, p. 45). More generally, they noted some discomfort of participants over the possible 'dual' purpose of the exercise: Was it to elicit their opinions as a representative jury, or was it to examine, for political advantage, their underlying motives or reasons? Renn and colleagues conclude that the planning cell 'cannot substitute for political decision making. Rather, it should be regarded as a decision aid to form and shape political judgments according to the latent and overt value structure of the concerned public' (pp. 45–6). They are, in balance, less sanguine about the reform potential of the Planning Cell than is Fishkin (1995) about the Deliberative Poll.

The unsatisfying answer we must accept at the moment is that we cannot be certain of the proper interpretation of data from such deliberative exercises. This is so in part because of the complexity of the quality objectives involved. These methods aim at a broad range of *different* quality criteria that apply to *different* phases and/or participants in the democratic decision making process. Attempts to maximize several distinct qualities at once make it rather difficult to identify specific objectives for assessing success. Second, as we have noted several times during our review, there has been a tendency to offer deliberative or informative packages that bundle together a number of distinct variables—provision of certain information, structured in a particular way, coupled with group discussions, supplemented by persuasive appeals, and so on—without incorporation of experimental designs allowing disaggregation of these variables or their effects. Third, actual measures of opinion quality generally have not been collected, leaving us without much empirical leverage at all on judgments of validity.

MAPPING THE ROAD AHEAD

In sum, then, we can identify three necessary ingredients for assessing quality claims in connection with efforts to gather readings of higher-quality public opinion: (1) clearly identified quality objectives; (2) experimental decomposition

of complex research manipulations into simpler constituent methods; and (3) measurements of relevant dimensions of opinion-quality along with any 'bottom-line' opinion outcomes. Although assembling these ingredients is by no means a simple task, there is much to be gained in searching for each.

Increased *specification of quality objectives* is a necessary first step in the further development of these new survey techniques. Exactly how is a given technique envisioned as contributing to the quality of collective decision making, and exactly how is it to improve public opinion? For example, it is conceivable that deliberative polls might better assist decision making by eliciting values than serve as a method for developing options (see Renn *et al.* 1984) or registering the mass evaluation of proposals. The ATIF survey program may be a successful method of floating 'trial balloon' proposals during the developmental phase and garnering preliminary indications of mass response, even if they do not register especially well-informed opinion or force the public to grapple with complicated policy trade offs. We suspect that each of the techniques we have examined has *some* important and valid applications, though they may be found, in the light of experience, far more limited than their advocates desire.

A second necessary step is *focused empirical attention on the separate methodological components* comprising each of these techniques. The various methodological factors we have identified and discussed above may as easily produce *undesired* as desired outcomes. Our present lack of knowledge about how method influences individual and collective opinion outcomes renders several of these complex, 'bundled' techniques quite problematic. As a research community, we are confronted not so much by four alternative techniques, but by a much larger number of different methodological innovations that have been variously employed in each of the four programs reviewed. The agenda for fruitful experimentation is thus extensive. Comparing the ATIF survey format to the Choice Questionnaire, we are left wondering what difference it makes to elicit serial evaluations of different alternative policies (as does ATIF) or to present competing options for a single selection (as in the Choice Questionnaire). What difference does it make to provide information in the form of political arguments (ATIF) or as neutral information about consequences (Choice Questionnaire)? Likewise, comparing the Deliberative Poll format to the Planning Cell, we are left wondering what difference it makes to gather one large group on a single site for debate (as in the Deliberative Poll) or to replicate standardized-format discussion groups across many sites (as in the Planning Cell)? How does it alter citizen behavior to place them 'on stage' in a media event (Deliberative Poll) or to have them work as citizen consultants outside the glare of publicity (Planning Cell)? Just as these are complex methodologies, they pose many important questions for experimental exploration.

A third necessary step, perhaps the most important, is *empirical specification*

of quality outcomes. Assessing the best methodological choices with respect to any of the variables just enumerated will require that researchers carefully specify and measure relevant dimensions of opinion quality (whether it be outcome-oriented or process-oriented, individual or collective). It is clearly not sufficient simply to document changes in the 'bottom line,' that is, individual or aggregate shifts of opinion, in order to lay claim to greater quality. Thus, the various changes of opinion recently reported in connection with participation in the Deliberative Poll, coming along as they do with a large dose of researcher manipulation and 'structuration,' are ambiguous in terms of what they can be taken to represent—both scientifically and politically. They were produced by a complex combination of methods, none of which has yet been coupled with any clearly articulated or measured dimensions of opinion quality.

The scientific challenge before us is thus to understand the link between each of the methodological procedures identified above and specific opinion-quality criteria. Only then will public opinion researchers be in a stronger position to modify and improve the mix of techniques, and to argue for a particular use of their data in collective decision making. Only then will we know whether they can contribute to improvement of our democratic ways, as we hope they can.

REFERENCES

- Alwin, D. F. and Krosnick, J. A. (1985): 'The measurement of values in surveys: A comparison of ratings and rankings', *Public Opinion Quarterly*, 49, 535–52.
- Beckwith, N. E., and Lehmann, D. R. (1975): 'The importance of halo effects in multi-attribute attitude models', *Journal of Marketing Research*, 12, 265–75.
- Bettman, J. R. and Kakkar, P. (1977): 'Effects of information presentation format on consumer information acquisition strategies', *Journal of Consumer Research*, 3, 233–240.
- Bishop, G. F., Oldendick, R. W., Tuchfarber, A. J. and Bennett, S. E. (1980): 'Pseudo-opinions on public affairs', *Public Opinion Quarterly*, 44, 198–209.
- Blumer, H. (1946): 'Collective behavior'. In A. M. Lee (ed.) *New Outlines of the Principles of Sociology*, New York, Barnes and Noble, pp. 167–222.
- Bryce, J. (1888): *The American Commonwealth* (Vol. 3), London, Macmillan.
- Burnstein, E. and Vinokur, A. (1977): 'What a person thinks upon learning he has chosen differently from others', *Journal of Experimental and Social Psychology*, 11, 412–26.
- Converse, P. E. (1987): 'Changing conceptions of public opinion in the political process', *Public Opinion Quarterly*, 51, S12–S24.
- Converse, J. M. and Presser, S. (1986): *Survey Questions: Handcrafting the Standardized Questionnaire*. Sage University Papers series on Quantitative Applications in the Social Sciences, series no. 07–063. Newbury Park, CA, Sage.
- Deutsch, M. and Gerard, H. B. (1955): 'A study of normative and informational

- influences upon individual judgment', *Journal of Abnormal and Social Psychology*, 51, 629–36.
- Dienel, P. C. (1978): *Die Planungszelle: Eine Alternative zur Establishment-Demokratie. Der Bürger plant seine Umwelt*, Opladen, Westdeutscher Verlag.
- Dienel, P. C. (1989): 'Contributing to social decision methodology: Citizen reports on technological problems'. In C. Vlek and G. Cvetkovich (eds.) *Social Decision Making for Technological Problems*, Dordrecht, Kluwer Academic Publishers, pp. 133–51.
- Edwards, W. (1977): 'Use of multiattribute utility measurement for social decision making'. In D. E. Bell, R. L. Keeney and H. Raiffa (eds.) *Conflicting Objectives in Decisions*. New York, Wiley.
- Edwards, W. (1983): 'Human cognitive capabilities, representativeness, and ground rules for research'. In P. Humphreys, O. Svenson and A. Vari (eds.) *Analysing and Aiding Decision Processes*, Amsterdam, North-Holland.
- Estlund, D. M. (1994): 'Opinion leaders, independence, and Condorcet's Jury Theorem', *Theory and Decision*, 36, 131–62.
- Fischhoff, B. (1984): 'Informed consent for transient nuclear workers'. In R. Kasperson and R. W. Kates (eds.) *Equity Issues in Nuclear Waste Disposal*. Cambridge, MA, Oelgeschlager, Gunn and Hain.
- Fischhoff, B., Slovic, P. and Lichtenstein, S. (1980): 'Labile values: a challenge for risk assessment'. In J. Conrad (ed.) *Society, Technology and Risk Assessment*, London, Academic Press.
- Fishkin, J. S. (1991): *Democracy and Deliberation: New Directions for Democratic Reform*, New Haven: Yale University Press.
- Fishkin, J. S. (1994): 'Britain experiments with the deliberative poll', *The Public Perspective*, July/August, 27–9.
- Fishkin, J. S. (1995): *The Voice of the People: Public opinion and Democracy*, New Haven, Yale University Press.
- Habermas, J. (1989): *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (T. Burger, Trans.): Cambridge, MA, MIT Press. (Original work published 1962).
- Iyengar, S. (1991): *Is Anyone Responsible? How Television Frames Political Issues*, Chicago, University of Chicago Press.
- Kahneman, D., P. Slovic and A. Tversky (eds.) (1982): *Judgment under Uncertainty: Heuristics and Biases*, London, Cambridge University Press.
- Kay, A. F., Henderson, H., Steeper, F. and Lake, C. (1994a): *Interviews with the Public Guide Us ... On the Road to Consensus*, Americans Talk Issues Foundation, St. Augustine, Florida.
- Kay, A. F., Henderson, H., Steeper, F. and Lake, C., Greenberg, S. B., Blunt, C. (1994b): *Steps for Democracy: The Many versus the Few*, Americans Talk Issues Foundation, St. Augustine, Florida.
- Keeney, R. L. and Raiffa, H. (1976): *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York, Wiley.
- Key, V. O., Jr. (1961): *Public Opinion and American Democracy*, New York, Knopf.
- Krosnick, J. A. and Alwin, D. F. (1988): 'A test of the form-resistant correlation

- hypothesis: Ratings, rankings, and the measurement of values', *Public Opinion Quarterly*, 52, 526–38.
- Merkle, D. M. (1997): 'Polls-review: The National Issues Convention Deliberative Poll', *Public Opinion Quarterly*, 60, 588–619.
- Molenaar, N. J. (1982): 'Response effects of "formal" characteristics of questions'. In W. Dijkstra and J. van der Zouwen (eds.) *Response Behaviour in the Survey Interview*, London, Academic Press.
- Moscovici, S. (1985): 'Social influence and conformity'. In G. Lindzey and E. Aronson, (eds.) *The Handbook of Social Psychology* (Vol. 2, 3rd edn.), New York, Random House, pp. 347–412.
- Neijens, P. (1987): *The Choice Questionnaire. Design and Evaluation of an Instrument for Collecting Informed Opinions of a Population*, Amsterdam, Free University Press.
- Neijens, P., de Ridder, J. A. and Saris, W. E. (1988): 'Variation in response functions and prescription for information integration'. In W. E. Saris (ed.) *Variation in Response Behavior: A Source of Measurement Error in Attitude Research*, Amsterdam, Sociometric Research Foundation.
- Neijens, P., Ridder, J. A. de, and Saris, W. E. (1992): 'An instrument for collecting informed opinions', *Quality and Quantity*, 26, 245–58.
- Neuman, W. R. (1986): *The Paradox of Mass Politics: Knowledge and Opinion in the American Electorate*, Cambridge, MA, Harvard University Press.
- Petty, R. E. and Cacioppo, J. T. (1981): *Attitudes and Persuasion: Classic and Contemporary Approaches*, Dubuque, IA, W. C. Brown Co.
- Petty, R. E., Rennie, G. A., and Cacioppo, J. T. (1987): 'Assertion versus interrogation format in opinion surveys. Questions enhance thoughtful responding', *Public Opinion Quarterly*, 51, 481–94.
- Price, V., and Neijens, P. (1997): 'Opinion quality in public opinion research', *International Journal of Public Opinion Research*, 9, 336–60.
- Price, V. and Tewksbury, D. (1997): 'News values and public opinion: A theoretical account of media priming and framing'. In G. Barnett and F. J. Boster (eds.) *Progress in the Communication Sciences*, New York, Ablex, pp. 173–212.
- Renn, O., Stegelmann, H. U., Albrecht, G., Kotte, U. and Peters, H. P. (1984): 'An empirical investigation of citizens' preferences among four energy alternatives', *Technological Forecasting and Social Change*, 26, 11–46.
- Rohrman, B. (1986): 'Evaluating the usefulness of decision aids: A methodological perspective'. In B. Brehmer, H. Jungermann, P. Lourens and G. Sevón (eds.) *New Directions in Research on Decision Making*, Amsterdam, North-Holland, pp. 363–81.
- Schooler, J. W. and Wilson, T. D. (1991): 'Thinking too much: Introspection can reduce the quality of preferences and decisions', *Journal of Personality and Social Psychology*, 60, 181–92.
- Schuman, H. and Presser, S. (1981): *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*, New York, Academic Press.
- Slovic, P. and Lichtenstein, S. (1971): 'Comparison of Bayesian and regression approaches to the study of information processing in judgment', *Organizational Behavior and Human Performance*, 6, 649–744.

- Slovic, P., Fischhoff, B. and Lichtenstein, S. (1980): 'Informing the public about the risks of ionizing radiation'. Paper prepared for the public meeting 'A Proposed Federal Radiation Research Agenda'.
- Slovic, P., Fischhoff, B. and Lichtenstein, S. (1982): 'Response mode, framing, and information-processing effects in risk assessment'. In R. M. Hogarth (ed.) *New Directions for Methodology of Social and Behavioral Science*, San Francisco, Jossey-Bass.
- Tanford, S. and Penrod, S. (1986): 'Jury deliberations: Discussion content and influence processes in jury decision making', *Journal of Applied Social Psychology*, 16, 322-47.
- Tindale, R. S., Davis, J. H., Vollrath, D. A. and Nagao, D. H. (1990): 'Asymmetrical social influence in freely interacting groups: A test of three models', *Journal of Personality and Social Psychology*, 58, 438-49.
- Traugott, M. W. and Tucker, C. (1984): 'Strategies for predicting whether a citizen will vote and estimation of electoral outcomes', *Public Opinion Quarterly*, 48, 330-43.
- Turner, J. C. (1982): 'Towards a cognitive redefinition of the social group'. In H. Tajfel (ed.) *Social Identity and Intergroup Relations*, Cambridge, Cambridge University Press, pp. 15-40.
- Tversky, A. and Kahneman, D. (1982): 'Judgments under uncertainty: Heuristics and biases'. In D. Kahneman, P. Slovic and A. Tversky (eds.) *Judgment under Uncertainty: Heuristics and Biases*, New York, Cambridge University Press, pp. 3-20.
- Van der Salm, C. A., van Knippenberg, D., and Daamen, D. D. L. (1997): 'A critical test of the Choice Questionnaire for collecting informed public opinions', *Quality and Quantity*, 31, 193-7.
- Van Knippenberg, D. and Daamen, D. (1996): 'Providing information on public opinion surveys: Motivation and ability effects in the information-and-choice questionnaire', *International Journal of Public Opinion Research*, 8, 70-82.
- Wilson, T. D., Lisle, D. J., and Schooler, J. W. (1993): 'Introspecting about reasons can reduce post-choice satisfaction', *Personality and Social Psychology Bulletin*, 19, 331-9.
- Worcester, R. M. and Harrop, M. (1982): *Political Communications: The General Election Campaign of 1979*, London, George Allen & Unwin, pp. 78-87.
- Yankelovich, D. (1991): *Coming to Public Judgment: Making Democracy Work in a Complex World*, Syracuse, NY, Syracuse University Press.

BIOGRAPHICAL NOTES

Vincent Price is associate professor at the Annenberg School for Communication at the University of Pennsylvania. He is author of *Public Opinion* (Sage Publications, 1992) and other articles and essays on mass communication, public opinion, and persuasion.

Peter Neijens is associate professor at the Amsterdam School of Communications Research (University of Amsterdam). He is also director of the Foundation for Fundamental Research on Commercial Communication and deputy director of the Dutch Press Foundation. Both foundations are affiliated with the Department of Communication at the University of Amsterdam.

Correspondence should be addressed to Vincent Price, Annenberg School for Communication, University of Pennsylvania, 3620 Walnut Street, Philadelphia, Pennsylvania, 19104-6220, USA. Tel: (1) (215) 898-7041; and Peter Neijens, Department of Communication, University of Michigan and the Amsterdam School of Communications Research, University of Amsterdam, Oude Hoogstraat 24, 1012 CE 1012 CE Amsterdam, The Netherlands. Tel: (31) 20 5253998; e-mail: neijens@pscw.uva.nl