

After the Initial Interview

NORC interviewers were trained in recruiting respondents to be delegates to the Convention. They were given briefing materials that describe the event, what is being offered to the respondents, and what they are expected to do in Austin. We prepared a set of questions commonly asked and the answers to them. We also provided them an 800 number to call if they had any further questions. The interviewers also have a letter of invitation from the President of the University of Texas, a folder with descriptive materials about the Convention and the city of Austin, reprints of newspaper articles on the event, and a reprint of a *Parade* magazine article which appeared during the middle of the field period. (The interviewers had reprints to use in advance of its publication). In addition, PBS had prepared a short video tape in which Jim Lehrer described the event and stressed the importance of the respondent's participation.

Respondents were asked to come to Austin from Thursday, January 18 to Sunday, January 21 to participate in discussions on major issues facing the nation. Transportation and all expenses for the trip are being paid for by the sponsors, and each participant receives \$300 to cover other expenses. Delegates may bring a companion, but will have pay for the companion's transportation, at a reduced fare provided by American Airlines. (The hotel will be free assuming that the companion shares the room with the delegate.) Since an event like this has never been attempted in the US, we have no idea what propor-

tion of the respondents will come. That is the challenge for the survey organization.

After respondents agree to become delegates, they are assigned to one of 30 small groups within which they will carry out their deliberations. Delegates are randomly assigned to groups so that each group will be a small random sample of the whole. Discussion materials prepared by the National Issues Forum and the Public Agenda Foundation will be sent to the delegates shortly before they leave for Austin.

Credibility Depends on Representativeness

The credibility of the event depends heavily on the representativeness of those who come to Austin. Even if we attain a very high completion rate in the initial survey, the sample in Austin may be biased if there is selective attrition in the sample. Conventional wisdom says that those who are wealthier, younger, better educated, more politically interested, more likely to vote, more used to traveling, etc. will be more likely to come. Evidence from the British experience, however, suggests that conventional wisdom might be wrong. At this point we do not know.

Because this point is so important, we are making every effort to minimize non-participation. As we do with non-respondents to the interview, we will do extensive follow-ups to persuade reluctant respondents to come. We have set up procedures to find out what problems need to be overcome for respondents to

feel comfortable about attending. For example, we have set up a special fund to pay for extra costs in cases of need. Most often these needs involve health concerns—respondents may have special requirements such that they must have a companion, or special facilities in Austin. If their employer does not want to give them time off, we are prepared to contact the employer and make the case for the importance of their attending. We have also set in place a network of persons charged with keeping in contact with the delegates between the time they have agreed to come and their arrival in Austin. We expect that some proportion of those who say they will come will later have doubts and want to back out. We will make every effort to keep them motivated to participate.

We will not get everyone to come. But we are in a different position here than when we assess the nature of bias from non-response in a cross-sectional interview. We have interviewed everyone and can compare the responses of those who come and those who do not come. If there turns out to be biases in the sample, we will have considerable data about the nature of the biases to use in analyzing the effects of participation.

The challenges in carrying out a deliberative poll are enormous. We look forward to reporting how well we have met them.



Norman Bradburn is senior vice president, Research, the National Opinion Research Center, University of Chicago

Fishkin's National Issues Convention Has Real Scientific Merit

By Philip E. Converse

I signed on with the National Issues Convention project about ten months ago because, simply, I was fascinated by what the results might be. I have no doubt that there is considerable scientific merit in this enterprise, and I am grateful to Everett Ladd for inviting me to express my opinion on it.

Although I had long been aware of Fishkin's work, my first close view of it came

from a videotape of his "deliberative poll" in Britain in 1994. As I became engrossed in the proceedings, I naively asked a better informed colleague how the citizens I was watching in group discussion had been assembled. I was delighted to learn that they were part of a national probability sample; this meant to me a kind of generalizability that the event otherwise would utterly lack. On my private, intuitive "10-point scale of information value," the video suddenly

The Deliberative Poll

shot from a two or three up to seven. (On this scale, the elucidation of public opinion by a politician defines the zero-point for information value, assuming negative numbers are not allowed.) My next question was, naturally, what fraction of the parent sample had come to the Manchester weekend. Had the answer been "all" or nearly so, my rating for the video would have climbed to nine-plus. Instead, the answer was that somewhat more than a third, though much less than half, had come. This wasn't reassuring, and my score for the event was put on hold at seven pending more information about the apparent bias of the people in Manchester, relative to the sample as a whole.

What I subsequently learned about the size and expense of the British undertaking led me to feel strongly that great care should be taken to maximize the scientific value of any other such project, if it were to be done at all. Last February I was asked to serve on a "Technical Review Committee" to consider seriously this maximization problem for a 1996 US version of the project. I agreed. I hasten to say that I do not think I have much to contribute in this regard, because things were already in very good hands. I have found little to quarrel with in the ongoing plans.

Some of the adverse commentary I have read about the project seems to imply that a project of this design can have little or no scientific merit, since there is no way to generalize the information collected beyond the purely phenotypic convention event itself. I find this contention bemusing, and hope that I have misunderstood it, and may have, since it seems to be more of an undercurrent than a crisp assertion. At the very least I should explain how I see the matter.

Fishkin's Project is a Field Experiment

From my wife Jean's work on the history of survey research comes the vignette that in perhaps the 1940s a psychologist, depressed at the apparent ignorance and casual responses of true samples of the electorate, suggested

sending out questionnaires in advance, to give respondents a chance to "study up," thereby curbing the incidence of non-attitudes and improving the quality of their answers. This suggestion seems to have found no resonance in the polling community—with good reason: The main point of polls is to assess the citizenry as it stands "in nature;" nothing could be more obviously self-defeating and "unscientific" than artificially contaminating it with special baths of information. But scientific inquiry is not limited to any single way of probing reality. Ideal, probably, is the experiment with random assignment to treatment and control groups. There are, however, many settings and many types of query that cannot be tackled by controlled experiment, and must be approached by naturalistic observations

“*Some of the adverse commentary I have read about the project seems to imply that a project of this design can have little or no scientific merit, since there is no way to generalize the information collected beyond the purely phenotypic convention event itself. I find this contention bemusing, and hope that I have misunderstood it, and may have, since it seems to be more of an undercurrent than a crisp assertion.*”

and mere statistical controls. Even here, there are many types of reality probes that can be leveled at a given target. The Fishkin project fits, in the most obvious and comfortable way, with what has long been called the "field experiment," where the setting is a natural one but where the investigator intrudes with interventions, often quite complex, to see what happens on variables crucial to his or her theory.

The complementary strengths and weaknesses of these diverse modes of probing reality are well known and need

no rehearsal here. But the relationship of each with reality differs in some particulars. For example, the controlled experiment is often aimed more at what *can* happen than what *does* happen routinely, as in physical probes of boundary conditions not known to occur in pure form in nature at all. Of course, Fishkin also wants to look at what *can* happen under quite *unreal* conditions. It is hard for me to see that his work has thus necessarily lost scientific merit.

Physical science analogies are often unwelcome in social science, where many "findings" replicate poorly if at all. But, this fact merely leads on to a tougher point: The dominant remedy for non-robust findings—in fact, the only constructive one I know—is a multi-method approach, which considers what patterns seem to survive the several angles of vision of different methods of probing reality. With this in mind, it's hard to see how scientific merit could possibly be served by canceling the one ambitious probe of the voter's issue psyche from an unusual methodology, to save resources to add a few hundred more conventional campaign polls to the thousands that are taking place in any event. I say this knowing full well that scientific merit may for many participants in this debate be a tertiary consideration, even irrelevant. But it is my assigned topic in a broader division of labor, and the generalizability of the information the "deliberative poll" will yield has been called into question.

With all this said, serious attention should be given to the potential weaknesses of this particular type of design. All of them can't be covered here, but two forms of weakness seem especially important to me.

A Mere N of One Event?

One concern is that whereas we may have a large "N" of respondents, the nature of the interactions between our guests at the convention, such as tone set by dominant personalities in one direction or another, may mean that what we really have is a mere N of one event. In a certain sense this is surely true. When my colleagues and I at

Michigan set out to learn about the American voter in the 1950s, we found ourselves in a neatly analogous predicament. We had Ns of 1,700 potential voters, which was a lot to work with; but sooner or later the humbling reality set in that we only had an N of one presidential election, with all of its local peculiarities of candidate pairings (one a victorious general!) and unique conjunctures of dominant events. Of all the many patterns in the data that were significant, which ones might display themselves in just about all presidential elections, as opposed to being mere products of unique features of this particular contest?

Exercising common sense isn't a panacea for this problem, but it surely helps. We discovered that citizens more interested in politics were more likely to turn out to vote than those less interested, and it wasn't hard to imagine that most other presidential elections would display this feature. On the other hand, had we discovered peculiar patterns in the vote choice of World War II G.I.s between Ike and Adlai, we might have felt that they would not survive generalization to other elections with different protagonists. I use polar examples here that are easily sorted; many patterns would be less easily assigned to one side or the other. But I do believe that had we sorted interesting patterns into two such piles on common-sense grounds with our N of 1, time and replication would have proved us right much more often than not.

Actually, in writing the report of the 1956 election, we were on somewhat firmer ground than just implied, in that a small study had been done on the 1948 election and a full-dress study in 1952. So our N of elections was more like 2 1/2, and we used this broader base quite consciously in sorting the probably-idiosyncratic from the probably-general. Analysts of the 1996 US deliberative poll will be in a similar position, since two other deliberative polls have been monitored as well. Patterns that are peculiar to only one or two of these assessments, and absent or reversed in the other, will be treated differently than those that are clear in all three instances—

despite change of country, change of convention format, *dramatis personae*, and the like.

Will Group Dynamics Ruin Generalizability?

Moreover, in other important aspects of the design, the effective N will be more than the single event. Worry has been expressed that "group dynamics" in the small-scale discussion groups may ruin generalizability. As Norman Bradburn has pointed out in his accompanying article, delegates to the convention will be randomly assigned to one or another of 30 discussion groups, each made up of 20 persons. This gives an N of 30 where group-dynamic variability is concerned. This will be a very strong hand relative to the old cottage industry of small-group experiments designed to see if group discussion produces opinion convergence. A long run of such single experiments suggested that probably on balance convergence occurred, but results were various, including some with no effect and others with actual polarizations. The 30-group design, with pre- and post-discussion measures in hand for all participants, can be examined in ways that might help explain the somewhat variable outcomes of this kind of discussion.

From Baseline Survey to Actual Attendees

The final concern has to do with the discrepancy between the sample interviewed in the baseline survey in November and those who attend the convention. There is obviously the potential for major bias through attrition, compounded by the fact that the initial response rate in any pre-convention baseline or "before" study will be well below 100%. The initial response rate in the personal interviews being conducted by NORC is expected to be at least 75%, a level which is almost certainly well above average for the general run of campaign polls, most of which are telephone interviews and many of which must go in and out of the field at such a rapid pace that extensive callbacks to represent less accessible telephone owners aren't feasible.

The initial response will be good, but the fraction of the sample willing to come to Austin for a three-day convention is as yet unknown. The one saving grace is that unlike those who escape the initial interview, we will know a lot from the baseline study about how those who come to the convention differ from those who cannot. Experience from the two British events shows no significant differences between these two groups. Put another way, no variable of the many dozens in the baseline survey predicts who will and will not attend the convention; the reasons surrounding the deci-

“*Worry has been expressed that “group dynamics” in the small-scale discussion groups may ruin generalizability. As Norman Bradburn has pointed out in his accompanying article, delegates to the convention will be randomly assigned to one or another of 30 discussion groups, each made up of 20 persons. This gives an N of 30 where group-dynamic variability is concerned.*

sion to visit appear orthogonal not only to the critical test variables of the study, but to any variables there at all. Given the very substantial attrition that is likely between the baseline survey and the Austin visit, there is a frightening amount of room for gaping biases to distinguish attendees from their parent sample—so it is fortunate indeed that none was found in the British experiments.

Forty years ago I would have flatly rejected such a null finding from the British experience as well-nigh incredible on its face. I am mildly surprised at it now. Still, it jibes very well with other things I have learned in the interim, so I am much more acceptant that it may be true—though I have not had a chance to examine the British data firsthand. My intervening learning comes from a number of painstaking analyses I have car-