

# **PRESIDENTIAL ADDRESS**

## **PUSHING BACK THE LIMITS TO SURVEYS**

---

ELEANOR SINGER

In the 50th anniversary issue of POQ, Philip Converse wrote that “there is nothing . . . that Blumer wished to take account of that is obscured” by the modern public opinion survey, and Jim Davis quipped, “I have been told for 35 years that surveys are about to be replaced by content analysis, observation, and experiments. Let’s see these wizards produce reliable, representative, and valid data within realistic budgets and deadlines.”

It is not only survey practitioners who are devotees of the art. If imitation is the sincerest form of flattery, then sample surveys are being flattered every day by look-alikes. Magazines tout haphazard questionings as “surveys,” and otherwise ethical organizations raise funds under the guise of doing a survey. The practice of imitation extends even to researchers who, on the one hand, insist that surveys cannot obtain valid data and, on the other, try to endow their findings with the respectability that comes from probability sampling and standardized methods of analysis.

And therein lies the paradox of survey research today. The success of the method seems assured. More surveys are being done than ever before. Yet, at the same time, the method arouses profound and pervasive dissatisfaction, ranging from complaints that it is superficial and imprecise to assertions that the information derived from it is inevitably biased and misleading, and perhaps even dangerous.

And so today, I would like to examine some of the limits of survey research as a way of gaining knowledge about social life. I do so with apologies to Turner, Martin, and their many coauthors (1984), who have said much of this before. Perhaps “limits” is not exactly the right word. “Limits” implies a barrier that cannot be breached, and at least some of the things I have in mind are not so intractable. Others, however, may be.

ELEANOR SINGER, a Senior Research Scholar at the Center for the Social Sciences, Columbia University, was president of AAPOR in 1987–88. This is a slightly revised version of her presidential address at Toronto, Canada, 21 May 1988. The author thanks Tom Cook, Theresa Rogers, Howard Schuman, Alan Singer, and Peter Strauss for helpful comments on an earlier draft.

This talk will take the following form. First, I will try to make clear what problems I will not address at all. Next, I will talk about some problems intrinsic to surveys as a research method and indicate some remedies that have been proposed for them. Neither the list of problems nor the list of remedies is exhaustive, but after a reasonable time I will move on to consider briefly some alternatives to the survey method. Finally, I ask where we might fruitfully go from here, and I suggest some steps that seem to me to be sensible. Inevitably, I will only touch on matters that deserve more thorough and more careful attention.

### **Problems I Will Not Address**

There are several kinds of problems that I will not address at all today, not because they are not serious, but because they are not intrinsic to the survey method as I have defined it for purposes of this talk. Chief among these are nonresponse and other coverage errors. Nor will I address a concern expressed by many, namely, the extent to which a survey should determine policy decisions or substitute for such decisions. Nor, finally, will I address such ethical issues as the need for informed consent or protection of respondents from potential harm. My inquiry is limited to problems peculiar to surveys as a means of questioning people in order to obtain valid information about their feelings, characteristics, and behavior. And the remedies I propose are likewise limited to this particular set of problems.

### **Problems Intrinsic to Surveys, and Some Proposed Solutions**

1. One objection that is often raised against surveys is that respondents won't answer sensitive questions or, if they do, will not answer truthfully, especially if a standardized question format is used.

The short answer to the first part of the objection is that respondents *are* willing to answer many such questions and have done so on a variety of surveys in the past. Furthermore, experiments with different kinds of confidentiality assurances, randomized response techniques, and other ways of guaranteeing either confidentiality or anonymity of response have increased people's willingness to answer such questions, though generally not by much (e.g., Singer, 1978; Bradburn, Sudman, et al., 1979; National Research Council, 1979; Zdep et al., 1979).

However, we know very little about how respondents actually per-

ceive such efforts to assure confidentiality of response. Are they reassured by them or, in fact, made more anxious? Or do they perhaps not care very much one way or the other? More research is needed in this area, and only some of it ought to be done by means of surveys.

So far as the second part of the objection is concerned—that respondents will not answer truthfully—we know very little. A few validation studies have been done with behaviors such as voting and even with more sensitive matters, such as drug use or arrests for drunk driving (e.g., Parry and Crossley, 1950; Locander, Sudman, and Bradburn, 1976; Katosh and Traugott, 1981; Miller and Groves, 1985). Such studies show considerable validity of response, but also some error—about 12%–15% in the case of voting, but much higher than that, about 35%, in the case of arrests for drunk driving. They also indicate that validity can be increased by using methods that on the face of it assure greater anonymity, and also by asking questions in certain ways rather than others. But since many of the sensitive behaviors we want to know about cannot be validated, and none of the sensitive attitudes—for example, racial hatred—can be validated in any simple way, we cannot answer the second part of the objection with great assurance. However, since no other research method permits validation of such attitudes or behaviors either, we are no worse off with surveys than with other methods of inquiry.

2. The objection that usually follows immediately upon the first one is that attitude surveys cannot tell us anything about what people really do, or *would* really do: in other words, they cannot shed much light on people's behavior. This area of attitudes versus actions has been extensively reviewed (e.g., Schuman and Johnson, 1976; Dawes and Smith, 1985:555 ff.), and I will make only two points here.

First, as Schuman (1972) has pointed out, the problem is often one of attitudes versus attitudes, not of attitudes versus actions. That is, if one wants to know about attitudes—for example, toward affirmative action—under different real-life conditions, one ought to ask about these different conditions in the survey. Responses are likely to vary from one situation to another, but not because respondents are lying or because they are “really” much more prejudiced than they appear to be. Rather, the variation may come about because different values, or different interests, are evoked by different situations. These conditions can be approximated by appropriate questioning.

The second point is that if it is behavior we are interested in, we ought to be doing more surveys of behavior. Real-life situations constrain response in ways that even sophisticated survey questioning cannot. For example, several colleagues and I replicated, on a small scale, a survey of discriminatory behavior in New York restaurants that had first been done 30 years earlier (Schuman et al., 1983). We

found that such behavior had been reduced but by no means eliminated. Similar surveys—which usually involve some elements of a field experiment—have been done to test compliance with fair housing laws. Such behavioral surveys cannot always be used to supplement questioning, but they can probably be used more often than they are if we once begin to entertain the possibility of doing so. It may even be possible, sometimes, to explore the reasons for discrepancies between expressed attitudes and observed behaviors by questioning respondents—as, for example, LaPiere (1934) might have done (but did not).

The more general principle that emerges from this example is that different research methods are appropriate to different research questions. Sometimes we attempt to elicit by means of questioning what might more accurately be learned from an analysis of documents, for example, or by observational methods. So one question that ought to be asked more often is, Is this survey really necessary?

3. A third objection sometimes raised is that surveys using standardized questions cannot capture the richness of respondents' experience. That is undoubtedly true, though we could do better in this regard than we ordinarily do. For example, we could include a series of interrelated branching questions on a topic, in order to explore a respondent's feelings or experience in depth. And we could supplement standardized instruments with much less directive kinds of conversations with respondents selected by some combination of random and purposive methods. For example, interviewers could be instructed to have such supplementary conversations with every  $n$ th respondent giving a particular answer to a certain question, after the conclusion of the standardized interview. These conversations could be tape-recorded and analyzed along with the survey responses.

4. A fourth objection is not that respondents *won't* tell us about their feelings or experience, as in objection number 1, but that they *cannot*: either because they cannot remember or because they cannot introspect.

Considerable research informs us that memory errors are large, that even for what one might think of as salient events, they begin to be substantial after a fairly short time, and that they are not random (e.g., Cannell, Fisher, and Bakker, 1965; Loftus, 1982). Indeed, unless research on ways of improving recall and reducing memory errors convinces us otherwise, survey researchers should probably forgo most retrospective questions.<sup>1</sup> If observations at two or more time points are

1. There are, of course, exceptions, especially those we might refer to as life-course events. The date of one's birth, the date of one's marriage, the birth dates of one's children—these are more likely to be recalled accurately, at least in part because some are commemorated every year.

important, then research designs should be modified—either by substituting longitudinal panels for cross-sectional surveys, or by having respondents keep diaries, or by some other aided recall procedure.

Nor should we rely on respondents to tell us *why* they are doing or feeling certain things. As Lazarsfeld (1935) pointed out more than fifty years ago in “The Art of Asking Why,” “we cannot leave it up to respondents to tell us whatever they are inclined . . . we must state precisely in which of the infinite number of determinants of an action we are interested.” Lazarsfeld’s advice on this point still repays careful study.

5. A fifth objection really consists of a series of related difficulties involving meaning: (a) The distribution of responses to a question may change as a result of apparently nonsubstantive changes in the wording of the question; (b) response distributions may change if the question is embedded in a different context; (c) some respondents will answer a question even if they have no opinion about it, the limiting case being an opinion about something which does not exist; (d) many of the quantifiers used in opinion surveys—words such as many, large, or often—have been shown to mean different things to different respondents; (e) respondents tend to stay within the frame of reference provided by the survey researcher’s response alternatives, so that these alternatives constrain the range of responses observed; (f) some respondents—Belson (1968) maintains at times as many as 80%—fail to understand the meaning of ordinary survey questions.

Taken together, this cluster of response effects is a large indictment of the survey method. In principle, we know some of the ways of defending against them. (a) Because we know that question wording may affect responses, we ought to ask questions in systematically different ways, preferably using split-ballot experiments. Such experiments may reveal that although marginal frequencies vary, relationships to other variables do not. Or the variations in marginal frequencies may provide important clues to the meaning of differently worded questions to respondents, and those differences in meaning can then be followed up directly.<sup>2</sup> In some cases, we ought to ask a series of questions about an attitude domain, instead of making one question do for the whole lot. (b) Similar recommendations have been made about ways of dealing with the effects of variations in context, but

2. The variations demonstrated by Tversky and Kahneman (1981) in people’s willingness to take risks, depending on whether the questions are framed in terms of potential gains or potential losses, are sometimes cited as indictments of the fallacy of questioning. But the point is that these questions do not *mean* identical things to respondents, even though the mathematical odds are the same. The problem is not the survey method, but the psychology of risk perception.

doing so is costly and time-consuming, and will not necessarily lead to unambiguous results.

(c) Belson's objection, that most survey questions are not understood by respondents, is perhaps an even more serious one to deal with. Again, in principle, defenses exist. For example, more extensive pretesting of survey questions can lead to questions that are better understood. Supplementing that, we could include measures of comprehension in the survey, so that we could, after the fact, separate the responses of those who understand more and those who understand less. (d) The same principle applies to nonattitudes; that is, we can allow respondents to exclude themselves from questions they have not previously thought about, and we can, if we wish, separate those who have thought a lot about an issue from those who have considered it only casually.

(e) Like other forms of communication, survey questions constitute frames of reference. They do so, for example, by limiting the response alternatives respondents consider, if the questions are of the closed-ended variety; and research by Schuman, Ludwig, and Krosnick (1986) has shown that even open-ended questions are not immune to such framing effects. Because marginals are most susceptible to such effects, some have argued that they should seldom, if ever, be reported without a comparative context. Such a context may be provided by changes over time, or by variations among subgroups, or by variations due to experimental alterations in the survey.

Constraint on responses is one way in which bias enters into survey research. Again, in principle, there is a solution to the problem, though it may not preclude all forms of bias. The solution, as Schuman pointed out in his presidential address two years ago, is to submit questions to people with divergent perspectives on an issue. Such a procedure can serve to identify response alternatives unimaginable to a researcher with a particular point of view and may point to biases in the wording of the question as well as in the response alternatives offered. It can even be used to suggest additional questions for inclusion.

In a paper titled "Postpositivist Critical Multiplism," Tom Cook (1985) has proposed a general strategy for dealing with biases due to limited perspectives. The general idea is to transcend these limits by bringing multiple perspectives to bear on research—in its design, in the formulation of questions, in the choice of data collection methods, and in the use of analytic techniques. The recommendation for soliciting the views of opposing groups when researching controversial issues is just one example of this strategy. Recommendations for seeking out widely divergent views in assembling the domain of statements to be used in constructing an attitude scale are based on similar logic, as is Egon Brunswik's (1947) recommendation for "representative research

design,” which was essentially a recommendation for systematically varying and testing the context in which a hypothesized relationship holds. The recommendation sounds simpler than it will prove to be in practice, but I believe it is the only research strategy capable of, as Cook puts it, “offering defensible interpretations of what is in the outside world.”

Before going on to more general considerations, I want to mention just one more vexing problem. Surveys have traditionally relied on standardized question wordings. Now some cognitive psychologists and anthropologists have raised the possibility that standardizing the wording of a question may not standardize its meaning for respondents; that, in fact, in order to standardize the meaning, which is what we really intend, we may have to *vary* the wording. Those researchers who have struggled to measure attitude change over time will recognize here a familiar dilemma: whether retaining the same meaning is best served by keeping question wording constant or by adapting the wording to presumed changes in meaning that have taken place over time. In resolving this dilemma, we will have to resort to analytic techniques that permit us to infer underlying constructs from overt measures.

### Are Other Research Strategies Better?

As I see it, the issue is not whether other research strategies should be used in *conjunction* with surveys, in order to improve the quality and the range of information obtained. The answer to that is clearly yes. The issue is, rather, whether an alternative method would, if substituted for surveys, yield better information about social life. The answer here, I think, is clearly no.

Consider some of the alternatives. One is to avoid the “demand characteristics” of surveys by using unobtrusive measures—for example, to look through household garbage for evidence of eating (or drinking) habits, instead of querying the occupants about them, in order to avoid the bias of social desirability.

If what we want is information about behavior, then, on research grounds, unobtrusive measures—for example, optical scanners in supermarkets—may well be preferable to questioning respondents. But if the aim is to link behavior to motive or cause, or if the object is not behavior at all but opinions or attitudes, then we are once again thrown back on methods involving more than simple observation.<sup>3</sup>

3. And we need only think about Dottie Renfrew, in Mary McCarthy’s *The Group*, carrying her pessary to a public park in a plain brown paper bag, and leaving it under a bench, to remind ourselves that respondents, like researchers, are aware that their trash can be used against them.

Another alternative to surveys sometimes proposed is ethnographic field work, which replaces the artificiality of the survey interview with prolonged observation and conversations in a natural setting. But while such methods can, at their best, avoid the sterility of some survey questioning, as well as its demand characteristics, they usually pay for it in representativeness. Morley's (1986) study of television viewing behavior is based on 18 British working-class families. Lane's (1962) study of working-class ideology was based on a sample of 15. Though the results of these studies are sometimes cited as if they applied to all members of their class, neither the size nor the method of choosing the sample will sustain such generalizations. Nor is it clear that field methods can—or do—avoid the errors and biases alleged against surveys. On the contrary, the potential for both seems greater, and issues of replicability and falsifiability are rarely addressed. Further, while it is at least possible to estimate the size of measurement error in the case of surveys, no available theory will allow us to do the same for ethnographic research methods.

### **Little Man, What Next?**

A little reflection will indicate that many of the survey problems cited earlier, though not all, could be remedied or at least ameliorated. They are not because, in the calculus of everyday research, doing so takes too long and costs too much. There are, it seems to me, remedies for this problem, though they too entail certain costs. They involve giving up the notion that anybody can, and should, do a new survey on any topic.

Pushing back the limits to surveys identified above requires expensive and extensive developmental work. No single survey organization is in a position to do that for every survey. But if more surveys were patterned after the General Social Survey or the National Election Survey, some of what is needed could be done.

What I have in mind is greater standardization of survey questions and procedures in order to improve the quality of individual surveys and increase the cumulative utility of survey results. Such standardization should apply first of all to how we ask about demographic characteristics, and even which demographics we ask about. But in addition, I would urge the development of standardized questions and even question sequences for attitudes and issues of recurrent concern.

There are obviously many forces working against such a development. My colleague Theresa Rogers and I tried in a modest way to encourage the various survey organizations regularly polling the public on AIDS to adopt common wordings for at least a few questions, and despite general agreement that that might be a good idea, we were

largely unsuccessful. Furthermore, it turns out that even when the wording is identical, the distributions obtained by different survey organizations often are not, making it impossible to merge the data points of different survey houses into a single trend line.

Considerable research would be required to establish the reasons for the differences, as well as the best questions for eliciting the information we really want. But such knowledge would be of obvious benefit to policymakers, and would also serve to increase public confidence in the validity of the survey method.

There are various ways of working toward the goal of developing more precise, valid survey instruments. One of the more practical, it seems to me, is through a fund for methodological research, established by a tax on survey activities, administered by a coalition of industry, academic, and government researchers—a standing committee of AAPOR, shall we say. Such a committee could be charged with developing an agenda for methodological research. Survey organizations would compete for funds to carry out the projects, and the results would be made available to all researchers, who would also receive incentives for using them.

Another possibility is for a number of survey organizations to share the responsibility (and the cost) of developing and testing questions in various areas. One, for example, might work on questions about racial attitudes, another on questions about AIDS. The benefits of the developmental work would then be shared among them all. Besides development, each organization would also assume responsibility for obtaining continual estimates of response error and response bias, and for developing new questions as an issue area expands. These activities are intended to supplement, not supplant, government-funded research on survey methods and on cognitive and other psychological principles underlying survey response. Nor are they intended to discourage experimentation by individual survey organizations.

Writing in *Surveying Subjective Phenomena*, Howard Schuman noted that the “basic approach to learning about large populations by means of posing questions to relatively small probability samples is a sound one. It becomes complicated not because of artifacts unique to surveys, but because human responses are inherently both subtle and complex” (Turner and Martin, 1984:148). Therein lie the limits to surveys, but with determination and effort, those limits could be pushed much further back.

## References

- Belson, William A. (1968)  
“Respondent understanding of survey questions.” *Polls* 3:1–13.

- Bradburn, Norman M., Seymour Sudman, et al. (1979)  
Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass.
- Brunswick, Egon (1947)  
Systematic and Representative Design of Psychological Experiments with Results in Physical and Social Perception. Berkeley and Los Angeles: University of California Press.
- Cannell, C. F., G. Fisher, and T. Bakker (1965)  
"Reporting of hospitalizations in the health interview survey." *Vital and Health Statistics*, ser. 2, no. 6, 1-71.
- Cook, Thomas D. (1985)  
"Postpositivist critical multiplism." Pp. 21-62 in R. L. Shotland and R. R. Mark (eds.), *Social Science and Social Policy*. Beverly Hills, CA: Sage.
- Dawes, Robyn M., and Tom L. Smith (1985)  
"Attitude and opinion measurement." Pp. 508-566 in G. Lindzey and E. Aronson (eds.), *Handbook of Social Psychology*, 3d ed., vol. 1. New York: Random House.
- Hippler, Hans-Juergen, Norbert Schwarz, and Seymour Sudman, eds. (1987)  
*Social Information Processing and Survey Methodology*. Berlin: Springer-Verlag.
- Katosh, J. P., and M. W. Traugott (1981)  
"The consequences of validated and self-reported voting measures." *Public Opinion Quarterly* 45:519-535.
- Lane, Robert E. (1962)  
*Political Ideology*. New York: Free Press.
- LaPiere, R. T. "Attitudes vs. actions." *Social Forces* 13:230-237.
- Lazarsfeld, Paul F. (1935)  
"The art of asking why." *National Marketing Review* 1:32-43.
- Locander, William, Seymour Sudman, and Norman M. Bradburn (1976)  
"An investigation of interview method, threat, and response distortion." *Journal of the American Statistical Association* 71(354):269-275.
- Loftus, Elizabeth (1982)  
"Memory and its distortions." In A. G. Kraut (ed.), *G. Stanley Hall Lectures*. Washington: American Psychological Association.
- Miller, Peter V., and Robert M. Groves (1985)  
"Matching survey responses to records: An exploration of validity in victimization reporting." *Public Opinion Quarterly* 49:366-380.
- Morley, David (1986)  
*Family Television: Cultural Power and Domestic Leisure*. London: Comedia.
- National Research Council (1979)  
*Privacy and Confidentiality as Factors in Survey Response*. Washington: National Academy of Science.
- Parry, Hugh J., and Helen M. Crossley (1950)  
"Validity of responses to survey questions." *Public Opinion Quarterly* 14:61-80.
- Schuman, Howard (1972)  
"Attitudes vs. actions vs. attitudes vs. attitudes." *Public Opinion Quarterly* 36:347-354.
- Schuman, Howard, and M. P. Johnson (1976)  
"Attitudes and behaviors." *Annual Review of Sociology* 2:161-207.
- Schuman, Howard, Jacob Ludwig, and Jon A. Krosnick (1986)  
"The perceived threat of nuclear war, salience, and open questions." *Public Opinion Quarterly* 50:519-536.
- Schuman, Howard, Eleanor Singer, Rebecca Donovan, and Claire Selltitz (1983)  
"Discriminatory behavior in New York restaurants: 1950 and 1981." *Social Indicators Research* 13:69-83.
- Singer, Eleanor (1978)  
"Informed consent: Effects on response rate and response quality in social surveys." *American Sociological Review* 43:144-162.
- Turner, Charles F., and Elizabeth Martin, eds. (1984)  
*Surveying Subjective Phenomena*. 2 vols. New York: Russell Sage Foundation.

Tversky, Amos, and Daniel Kahneman (1981)

“The framing of decisions and the psychology of choice.” *Science* 211:453–458.

Zdep, S. M., et al. (1979)

“The validity of the randomized response technique.” *Public Opinion Quarterly* 43:544–549.