

DISCUSSION OF THE EFFECT OF THE QUESTION ON SURVEY RESPONSE

Charles F. Turner, National Research Council(1)

Kalton and Schuman's paper prompts many reactions, indeed many more than could run their full course in my allotted time. This is, in itself, high praise for the authors' work. In my comments I will dwell upon three points rather than attempting to traverse the full range of the author's arguments and insights.

1. IMPORTANCE: At the outset, there is one general commentary that needs to be spoken. It is that work of the sort reviewed by these authors is extraordinarily important. The importance of this work devolves from the reliance which is placed upon survey research measurements in public affairs that many people take to be of national importance.

Americans witnessed last year the spectacle of a sitting president retreating to Camp David in order to diagnose our "national malaise" based largely upon data supplied by his pollster (cf. Zeisel, 1980). The League of Women Voters will decide which presidential candidates take part in its televised debates based upon the candidates standing in the public opinion polls. Last year, the House of Representatives almost enacted legislation to hire a survey research firm to poll the public on major national issues (cf. Washington Post, 1979). And during the last decade, various federal agencies began to issue social indicator reports incorporating, among other things, time-series data on public opinions toward government, science, national problems, and so forth.

The proliferating use of such survey data make particularly important and urgent the sort of research discussed by Kalton and Schuman. To make proper use of such data, one must have a reliable basis for estimating the magnitude of both the sampling and non-sampling variability which affects these measurements. The effects of the wording and context of our survey questions are an important and inadequately understood source of variability.

Should anyone remain unconvinced of the importance of such work, I would offer as a final piece of evidence the front page of the Washington Post for Sunday, April 20th, 1980.

The lead story on that day was reported with a banner headline across the top of the first page; it read: "55% favor use of force in Iran." The story reported the results of a Washington Post poll purporting to show that:

"a strong majority of Americans -- almost 2 to 1 -- now supports military action even if it imperils the lives of the hostages."

This banner-headlined story was accompanied on the first page by a column of news analysis headlined "Logic points to military move by U.S." [For those who are not good with dates, I should remind you that April 20th was the Sunday immediately preceding the American raid on Tabas, Iran.]

The text of this dramatic poll story is of particular interest in two respects. First, the poll was reported in a way which implied that the President must attend to these sentiments or court doom at election time. Thus, the second sentence of the story read: "The hostage crisis,

in addition, appears by far the dominant voting issue of the 1980 presidential campaign." Secondly, the results reported in the headline were derived from a poll question which forced respondents to choose between two alternative policies:

- 1: "The United States should set a deadline for the return of the hostages and take military action if they are not returned by then."
- 2: "No matter how long the hostages are held in Iran, the United States should take no action that could threaten their lives."

The forcing of a choice between alternatives is not an unknown practice in survey research. However, it is reasonable to suspect that many respondents were not in complete agreement with either of these positions. Indeed, it appears that 15% of the sample had the temerity to refuse to make the choice demanded by the poll. At a minimum one suspects that a different formulation of this question might have produced rather different results.

Clearly, improving both our knowledge and the public's understanding of the potential weakness and strengths of such survey measurements are urgently important tasks.

2. IMPLICATIONS OF IGNORANCE. I do sometimes find it instructive to reflect upon the lexicon used to discuss errors in sample survey data. In particular, I note the common dichotomy between sampling vs. non-sampling error (see, for example, Andersen et al., 1979, p. 6). All of what Kalton and Schuman discuss fall within the category of non-sampling error.

What is curious is that our lexicon defines this area as a residual category, i.e., all errors other than sampling. Indeed, to distort things a bit for the sake of argument, it often seems as if there exists one category of survey errors, i.e., sampling errors, for which there is a well-developed theory and a general consensus as to the characteristics of a well-designed sample.² In contrast, errors due to non-sampling factors appear to be diverse in origin and theoretically unintegrated. They are related only in the fact that they have nothing to do with sampling. In this area one finds many empirical demonstrations, but there is no general theory (or set of specific theories) that will allow unambiguous inferences about the magnitude of potential non-sampling errors in our survey data. (See, for example, the recent statement in the layman's primer on survey research produced by the ASA section on survey research methods; Ferber et al., 1980, p. 18-19).

The underdeveloped state of theory in this area, thus, coexists with the previously noted and important consequences which can flow from a failure to realistically anticipate the magnitude of the non-sampling errors which affect our measurements.

An example: How can we be subverted by our ignorance? Let us take an example from an area

that is of concern to all of us -- public support of scientific research.

Bi-annually since 1972 the National Science Board has published a series entitled Science Indicators (NSB, 1972-). Each of the first three volumes contained a concluding chapter reporting survey data on public perceptions of science.

The Board's survey asked, among other things, for respondents to report the problem areas in which they would most like to see tax money for science and technology spent. The results from the first two surveys (in 1972 and 1974) were perplexing. They indicated that the public wanted to pour money into scientific research on problems which scientists had little hope of solving, such as "reducing crime."

In an attempt to understand these results, the Board included questions in its 1976 survey asking respondents in which areas science and technology would be most and least useful. These questions immediately preceded the question asking about support for spending.

The resultant measurements suggested an apparently massive decline between 1974 and 1976 in public endorsement for spending in all areas of science and technology.³ These declines ranged up to 27 percentage points! Chastened by this result and alert to the possibility that the change in the content of the questionnaire may have caused these declines, the Board did not follow its previous practice of presenting the new results beside those obtained in previous years. Instead, the earlier data were omitted and the authors observed in a footnote that because of the change in the content of the questionnaire the earlier "results are not strictly comparable to the 1976 results" (NSB 1976, p. 180).

In short, the authors posited the existence of a large non-sampling component of variance in these estimates (deriving from changes in questionnaire context) to justify suppression of some very disagreeable data.

Similar time-series measurements made by other surveys over the same period suggest that the Board's strategy was, in hindsight, quite reasonable (cf. Turner, 1981a). Nonetheless, I would ask you to contemplate the potential mischief which might have been done if these data had been interpreted by an unsympathetic party--say a Senator Proxmire.

Kalton and Schuman rightly point to the ubiquitous potential for such context effects in time-series measurements--while properly observing that the frequency of such effects is unknown. I would echo the conclusions of Drs. Kalton and Schuman that more research and surer knowledge are required if we are to avoid such potential disasters. Indeed, I would go slightly further than the authors and suggest that survey researchers consider the routine use of experimental designs in their surveys to provide information on the robustness of key measurements over specified variations in questionnaire context and question wording. While such demonstrations cannot incorporate all possible variations, they would still provide useful information on the robustness of important measurements.

3. FUTURE RESEARCH: Before considering the authors' proposals for future research, I would

like to offer a general suggestion for consideration.

Are there "true values" for subjective measurements? To sustain logically the concept of "error" or "bias" in a measurement one must posit a forthright notion of true value for the measurement. When one enters the subjective domain, this notion becomes murky. In particular, it becomes difficult to imagine any simple test which would (even in theory) lead one to conclude that a particular question wording (or context) produced measurements that were closer to the "true value" than another question wording (or context). For example, there is evidence suggesting that responses to general questions on happiness are highly labile; one seems to obtain different responses when such questions are asked after inquiries about the happiness of respondents' marriages vs. their financial situation. How would one determine which of these measurements is closer to the "true value"?

The notion of true value is not well founded in the subjective domain. Hence, one encounters considerable difficulty with the subsidiary notions of bias and error. In this regard, I recall that Stanley Payne (1954, p. 72) in his lovely little book on question wording noted in discussing question bias that: "A 'good' question, among other things, is one which does not itself affect the answers."

The inane of that statement speaks mightily about the fruitfulness of these notions when applied to non-sampling effects on subjective measurements. I would, thus, suggest that, when it is theoretically impossible to conceive of the direct measurement of a variable (i.e., without relying upon respondent's self-report), we might consider disciplining ourselves to speak only of the variability due to question wording, questionnaire context, etc. This might preclude sterile discussions of error and bias which presume the impossible notion that there exists a true value for these subjective measurements which is independent of and unconditioned by the act of measurement.

Research Topics. The Kalton and Schuman review concentrates upon the formal attributes of a question. That is, whether the question is balanced or not, whether it offers a middle position, whether the question is open-ended or whether we force respondents to choose between specific response categories, and so forth. Work in this vein is clearly basic to our understanding of the vagaries of survey measurement, and it provides a most reasonable place to begin.

However, I would hope that this research would be broadened. Questions vary not only in their form but also in their content; there are both the formal aspects of syntax and the essential questions of meaning -- albeit these are always intermixed. A comprehensive treatment of question effects will need ultimately to deal with issues of communication and meaning.

At a practical level it would be useful to consider whether it is possible to delimit areas in which the meaning of questions and the consequent stability of measurements are more (and less) robust. Consider the public and factual characteristics of respondents, e.g.,

their gender or age. It is hard to imagine changing a question wording or context so as to produce a 15 percentage point discrepancy in aggregated categories of the (self-reported) age or sex distribution. In contrast, a series of three recent pilot experiments conducted by SRC, NORC and the Washington Post poll indicates that common subjective measurements of "happiness" can show such discrepancies when the questionnaire context is uncontrolled (cf. Turner, 1981b).

These experimental studies manipulated the order of presentation of two questions: one on general happiness⁴ and one on the more specific question of respondents' happiness with their marriages.⁵ It was found that the distribution of responses to the more specific marital happiness item was unaffected by the measurement context, while the response distribution for the more amorphous question on general happiness evidenced large fluctuations when the measurement context was altered.

Kalton and Schuman note that two questions on abortion show a parallel pattern of context effects. Deviations were found in the response distribution for a general abortion item when the questionnaire context was experimentally varied, but no variation was found for a specific item (abortion when birth defects are expected).

One hypothesis to be entertained in reviewing such results is that the context in which questions are asked is sometimes part of their meaning. The context can suggest a frame of reference or prime a particular perspective. It is thus reasonable to expect that the importance of a context in supplying meaning should be most pronounced when the stimulus question is least specific. The generality of these findings is an appropriate topic for future research. I would also suggest that more explicit research on the meaning of survey questions is required to complement past studies on the form of questions. Among other things, we need to know if the conclusions that have been drawn from past research on question form will generalize to all topics. Is it safe to assume, for example, that the effects of offering a middle alternative are the same for all subject areas? Are there questions for which it is impossible to induce context artifacts or which are particularly vulnerable to such artifacts? etc.

At a more general level, I think it is important to recognize that what is at issue in the survey interview and in its interpretation are meanings. The complexities of the interpersonal exchange between the interviewer and respondent and the hidden exchange that takes place between analysts and their data are all struggles over the communication of meaning. A respondent must decode our overt and covert meaning when we ask whether s/he trusts the government, expects to have any more children, will vote in the next election or wishes to go to war with Iran. We, in turn, are blessed with data that has been filtered through the subjectivities of question writers, interviewers, respondents and coders. The data embody all of the complexities of a rich series of complex symbolic interchanges; this complexity must be respected and the processes better understood if such survey measurements are to be of enduring use to us.

NOTES

¹The views expressed in this article should not be attributed to the National Research Council.

²We also have a basis for comparison when one wishes to discuss the impact of deviations from this ideal (see, for example, Stephenson, 1979).

³Average decline was 13 percentage points for 12 areas (see, Turner, 1981a).

⁴Question wording: "Taken all together, how would you say things are these days--would you say that you are very happy, pretty happy, or not too happy?"

⁵Question wording: "Taking things all together, how would you describe your marriage? Would you say that your marriage is very happy, pretty happy, or not too happy?"

REFERENCES

- Anderson, R., Kasper, J., Frankel, M., and Associates. Total Survey Error. San Francisco; Jossey Bass, 1979.
- Ferber, R., Sheatsley, P., Turner, A. and Waksberg, J., What is a survey? Washington, D.C.: American Statistical Association, 1980.
- Payne, S. The Art of Asking Questions. Princeton: Princeton University Press, 1954.
- National Science Board, Science Indicators. Washington, D.C., U.S. Government Printing Office, biennial from 1972.
- Schuman, H. and Presser, S. Questions and Answers. New York: Academic Press, 1981, in press.
- Stephenson, C.B., Probability sampling with quotas: an experiment. Public Opinion Quarterly, 43, 1979, 477-496.
- Sudman, S. & Bradburn, N., Response effects in surveys. Chicago: Aldine, 1974.
- Turner, C.F., Surveys of Subjective Phenomena: a working paper. In D. Johnston (Ed.), Measuring Subjective Phenomena. Washington, D.C.: U.S. Government Printing Office, 1981a, in press.
- Turner, C.F., Patterns of Disagreement: a reply to Angus Campbell. In D. Johnston (Ed.), Measuring Subjective Phenomena. Washington: U.S. Government Printing Office, 1981b, in press.
- Washington Post, "Plan to hire polling firm runs into snag in House," December 17, 1979, p. A-3.
- Washington Post, "55% Favor use of force in Iran", April 20, 1980, p. A-1.
- Zeisel, H., Lawmaking and public opinion research: the president and Patrick Caddell. American Bar Foundation Research Journal, 133, 1980, 133-139.