KM 33(1990)
pag 137-159

SURVEY QUESTIONING; AN OVERVIEW OF RECENT METHODOLOGICAL STUDIES

N.J. MOLENAAR

Abstract
In this article an overview is presented of some 50 methodological studies on survey questioning, published since 1986 untill 1987.

INTRODUCTION

Survey interviewing with a standardized questionnaire is a very frequently used procedure to collect data in social science research. Along with its popularity, the procedure also receives much critical attention from methodologists. These methodological studies are all without exception concerned with the quality of the data obtained, mostly by investigating the features of the procedure that may unintentionally influence respondents' answers to the questions asked. In this article we will present an overview of recent methodological studies on survey questioning, published since 1986 untill 1987.

The studies to be considered are quite diverse. Some studies are chiefly theoretical: they propose general perspectives, theories, models or concepts relevant for understanding what is going on in survey questioning. Other studies, in fact most of them, are empirical: they present new research findings along with some limited theory. The empirical studies are divided into reliability studies, validity studies, and studies devoted to the problem of non-response.

Studies on the reliability of survey data are relatively scarce. The vast majority is devoted to validity problems (systematic errors). These studies are divided into four categories according to the main sources of errors in survey data. These are: the interviewer, the respondent, the questionnaire and the circumstances. It must be noted, however, that several studies deal with some combination of these categories, for example when interaction effects are studied.

The final section contains some general theoretical studies on errors in surveys. The studies are presented in some detail, but without any attempt to integrate the findings and limited theories into an overall "error"-theory.

N.J. Molenaar
Vakgroep Methoden en Technieken, Vrije Universiteit
Koningslaan 22-24
1075 AD Amsterdam
tel. 020-6647131

RELIABILITY

Nearly all methodological studies on the quality of social survey data deal with attitudinal data; factual data are commonly assumed to be comparatively less subject to biasing factors. Furthermore, studies on the reliability of survey data, whether attitudinal or factual, are relatively scarce.

Porst and Zeifang (1987) are an exception: they studied the reliability of factual data obtained in a German survey using face-to-face interviews. Respondents were asked three times, with one month between times, the same socio-demographic questions; the percentages of the answers that remained stable over the three waves were calculated. Most stable were the answers to questions on respondents' gender, marital status and age (in six categories), the stability percentages being about 98 percent. The questions on education and religion reached 89 percent stable answers. Employment status, current occupation, occupational training were minimally acceptable with 81, 73 and 72 percent stable answers, respectively. Own first occupation (44 percent), and father's occupation (52 percent) were below the level of acceptance. Monthly income (eleven categories) scored only 27 percent stable answers. The authors observed the tendency for stabilities to decrease as the number of answer categories increases. Although some part of the instabilities could have been due to real changes in the variables through time, the findings indicated that the reliability of factual data leaves much to be desired.

VALIDITY

The major part of recent methodological studies on survey interviewing deals with validity problems, with design features systematically related to the answers obtained from the respondents. These systematic relationships are generally called response effects. The studies to be considered in this section are organized according to the main sources of systematic errors: the interviewer, the respondent, the instrument, and the circumstances.

The interviewer

Interviewers can significantly affect respondents' answers to survey questions: by their appearance, age, gender, personality and other role-independent characteristics, or by their role-dependent characteristics such as their motivation and capability to interview as prescribed by the researcher, the manner in which they ask the questions, their probing behaviour, the manner in which they communicate their own opinions on the issues to the respondents, and so on. Three main types of research on interviewer effects can be distinguished: the correlation approach, the experimental approach, and the behavioural approach. On each approach some new evidence is available.

The correlation approach

In this approach interviewer effects are typically established by correlating characteristics of the interviewers (their opinions, race, gender and the like) with characteristics of the answers obtained from the respondents. In the extreme case, interviewer effects are established by determining only how much of the variance in respondents' answers can be accounted for by differences between interviewers, without specifying any

further interviewer characteristic ("interviewer variance").

The interviewer variance approach was used by Groves and Magilavy (1986) for estimating the interviewer effects on approximately 300 survey statistics from nine telephone surveys. The overall mean interviewer effect, as measured by intraclass correlation coëfficiënts, was estimated near to 0.01, an average somewhat smaller than those reported from personal interview surveys (0.02). Factual items did not appear to be subject to any different interviewer effects than attitudinal questions, nor did open vs. closed questions. Only in open questions like "What is the most important problem facing this country today?" the number of different entities mentioned and the type of the second mention were subject to interviewer differences, probably due to differences in probing behaviour. With respect to various interviewer characteristics and respondent characteristics no significant relations with interviewer effects could be observed. There was some evidence, however, that older respondents are more susceptible to interviewer effects than other age groups. Finally, small reductions in interviewer effects were obtained when interviewers were trained to limit their verbal interactions with the respondent to a specified set of probes and feedbacks, in addition to the survey questions.

As to the specific interviewer characteristics, new evidence is available on the voice, the ethnicity and the opinions of the interviewer.

Oksenberg et al. (1986) investigated the possible effect of interviewers' voice quality on refusals in telephone interviews. Telephone interviewers with lower refusal rates were reliably rated by judges as having better voice quality (higher pitched voices, greater variation in pitch, greater loudness, faster rates of speaking, more distinct pronunciation). On the basis of their voice characteristics they were also judged as having a more positive approach to the interview, and as being more competent. Although the associations between refusal rate and these vocal and inferred personal characteristics were moderately strong, vocal cues seem to play a significant role in telephone interviewing. Perhaps voice characteristics should be a criterion for selecting interviewers.

Reese et al. (1986) investigated the possible effects of interviewers' race on the responses in a telephone survey of Anglos and Mexican-Americans using Anglo and Hispanic interviewers and questions related to Mexican-American culture. Ethnicity-of-interviewer effects are currently explained by deference theory: respondents tend to avoid offending an interviewer, particularly in ethnicity-related questions. The data showed that Anglos were, indeed, less likely to disagree that they preferred Mexican-American culture when they were interviewed by an Hispanic than when they were interviewed by another Anglo. Similar effects were observed on three out of five other ethnicity-related items. On the other hand, those same items failed to show significant interviewer effects among the Mexican-Americans themselves. Other items which dealt specifically with ethnicity were examined; some items showed interviewer effects in the predicted direction, others were against the deference hypothesis. As predicted, in the non-cultural items no significant ethnicity effects were observed. The results seem to support the deference theory only partially; further study would be needed to explain why deference occurs in some items and not in others.

Another possible source of interviewer effects is the own opinion of the interviewer on the issues that are dealt with in the questions. In a rather complex study Bruinsma (1987) asked a sample of workers in a Dutch steel factory about their willingness to participate in union actions. Interviewers were social science students who varied in action-mindedness. The stronger the action-mindedness of the interviewer, the more likely

respondents' answers to the questions were affected by their perception of interviewers' opinion. Action-minded interviewers, however, did not seem to be better in transferring their own opinion than the other interviewers. Perhaps the action-minded interviewers showed a stronger involvement in the topic and may have emphasized the importance of the issue.

Interviewers' opinions can also have less straightforward effects on the research outcomes, as is demonstrated by Loosveldt (1986). He reported on a Belgian survey among married women, in which trained and untrained female interviewers were used. Fourteen pairs of questions on household tasks were asked, each consisting of a factual question (Who does task X in your household; man/woman/both) and, in that order, a wish-question (Who do you wish to do task X in your household; man/woman/both). The opinions of the interviewers on the wish-questions appeared to be correlated with respondents' answers to these questions, both among the trained and the untrained interviewers. Further analyses, however, revealed that the trained interviewers saw more discrepancies between wish and fact than the untrained did, and their respondents also made the distinctions better. This last finding indicates that the trained interviewers were better in communicating the difference between wish and fact to their respondents than the untrained, which made question order effects less likely to occur.

The experimental approach

In the experimental approach the researcher tries to assess experimentally the effects of interviewer behaviours that are assumed to lead respondents to give complete, relevant and accurate information; for example, by training the interviewers in clarifying the goal of the research, explaining the task of the respondents, and giving adequate feedback on respondents' answers. Will the quality of the data be better if the interviewers are trained in these aspects? One measure of data quality is the number of item non-responses obtained. Loosveldt (1986) found in a nine-question Belgian survey that the trained interviewers obtained 33 percent item non-response, and the untrained 55 percent. Another measure of data quality is the amount of information obtained (it is commonly assumed in this context that more information is better information). Loosveldt (1986) reported on a Belgian survey in which thirteen factual questions were asked, mostly dealing with health-related behaviours. In seven questions a training effect could be observed: in the trained condition the reported frequencies were significantly higher than in the untrained condition. The effects were not so much on mentioning or not mentioning an entity, but on the frequencies reported. These findings suggest that the quality of survey data can be improved significantly by training the interviewers.

The behavioural approach

The behavioural approach deals with how interviewers actually behave during the interview interaction, adequately or inadequately. The behaviour is typically studied by analysing verbatim transcriptions of tape-recorded question-answer sequences. Do interviewers behave as they are instructed to do?

Loosveldt (1986) analyzed more than 2,700 question-answer sequences, obtained from survey interviews in which trained and untrained interviewers were used. In 60 percent of the sequences, reading the question was immediately followed by an adequate answer of the respondent,

mostly in the trained condition. In 20 percent of the sequences respondents only gave a partial answer to the question. How do interviewers act in these situations? In the trained condition 12 percent of the problematic sequences were closed by the interviewer providing an answer, against 27 percent of the sequences in the untrained condition. The trained interviewers were also better in subsequent probing (less suggestive probing) and in giving adequate feedback than the untrained. In a quarter of all cases the partial answer ended up in a final adequate answer to the question, in the trained more than in the untrained condition.

It may be concluded from these data that training can indeed improve interviewer behaviour. Similar studies on actual interviewer behaviour can be found in Brenner (1985) and Dijkstra et al. (1985).

All studies on interviewer behaviour demonstrated that even trained interviewers make mistakes. Some insight into the conditions in which these inadequate behaviours are most likely to occur, was provided by Van der Zouwen et al. (1986) who analysed more than 2,000 question-answer sequences, obtained from interviews in which trained interviewers were used. Some interviews were held in a businesslike, formal style, others were held in a warm, socio-emotional style. Four main types of inadequate interviewer behaviour were identified, most frequent of which were "choosing" and "hinting".

"Choosing" is: if the respondent does not select one of the answer-categories presented, the interviewer himself chooses an answer. It occurred in 16 percent of the sequences. It seemed to be unrelated to interview style nor did it seem to affect the research outcomes seriously.

"Hinting" is a special form of asking leading questions. If the respondent hesitates in giving an answer, the interviewer gives a hint for an answer ("So I think you mean..?"). It occurred in 19 percent of the sequences, especially in the socio-emotional style of interviewing, with open questions and among female interviewers; it also appeared to affect the research outcomes.

From their cybernetic frame of reference, the authors concluded that these types of inadequate interviewer behaviour occur more frequently in socio-emotional interviewing style and with open questions, just because in these kinds of situation the control tasks of the interviewers are relatively most complex; asking suggestive or leading questions seems to be an efficient way to manage all control-tasks simultaneously.

These and other research findings are summarized by Dijkstra (1987) in an overall model. The socio-emotional style of interviewing was said to be more motivating than the formal style, which, other things being equal, enhances the likelihood that respondents' answers are adequate and valid. On the other hand, the socio-emotional style enhances the likelihood that the interviewer will pose suggestive questions, which in turn enhances the likelihood that the answers obtained are adequate indeed, but invalid. This style effect will especially occur in unstructured situations, e.g., in open probing questions.

A similar but somewhat more elaborated model was presented by Dijkstra and Van der Zouwen (1987). This model describes the possible effects of interviewing style on the quality of the responses, together with their mediating processes and the conditions in which the style-effects are most likely to occur. It is argued, for example, that the socio-emotional style of interviewing offers more opportunities for interviewers to reinforce the respondents' previous answers, which may lead to biased responses. The style will also positively affect respondents' mood, which in turn can lead to positively biased reportings. On the other hand, the style is said to be better in motivating respondents, enhancing the likelihood that

adequate and accurate answers will be given. Whether these style-effects will occur, however, depends on the interview topic. The biasing style-effects will only occur if the topic is evaluative, that is if the response alternatives differ in their social desirability. If the topic is non-evaluative, the socio-emotional style of interviewing is to be preferred, because the greater motivating force of that style leads to a better respondent performance. If the topic is evaluative and highly threatening, the mood of respondent will be negatively affected, enhancing the likelihood of negatively biased answers; this possible effect can then be compensated for by using the socio-emotional style of interviewing. If, however, the topic is not threatening but highly salient, the formal style is to be preferred.

Although the foregoing models are hypothetical, they may help the researcher to decide what kind of social relationship should be preferred in which circumstances in order to enhance the likelihood that adequate and valid information will be obtained from the respondents.

THE RESPONDENT

Characteristics of respondents may unintentionally affect the research outcomes, particularly if these characteristics are related to the tasks to be performed or other features of the survey interview procedure. Respondent effects are mostly investigated in interaction studies, to find out among which categories of respondents certain response effects are most likely to occur. Two recent studies about the effects of respondent characteristics on the research outcomes will be reported in this section.

Silver et al. (1986 b) estimated that 20 to 30 percent of those who did not vote in the past elections, falsely reported to survey organizations that they had voted. Analysis of vote validation data from three surveys revealed that college graduates and middle aged people were most likely to overreport voting. The tendency to overreport appeared to be also positively related to respondents' general political attitudes. The strongest predictor for voting overreport, however, was whether the respondent stated in a pre-election survey that he or she planned to vote. Among those who said they planned to vote and actually did not, 51 percent said that they voted when interviewed after the election; in contrast, among those who said that they did not plan to vote and who actually did not vote, only 2 percent overreported. From further analysis it was observed that vote expectation is a powerful filter for the effects of general political attitudes on vote overreporting: the effects were most pronounced in the "expect to vote" condition.

The second study on respondent effects was more general and concerned the question whether differences in survey research results between older and younger respondents might be attributable, in part, to differences in the quality of the survey measures obtained from these groups. In a multitrait-multimethod design, Andrews and Herzog (1986) collected a heterogeneous set of measures from six surveys: 106 measures that tapped twenty-six different concepts (traits) and fourteen different response scale types (methods). On the basis of an appropriate measurement model, a set of quality estimates was constructed for each measure; these quality estimates were then related to the age of the respondents. The construct validity of the measurements appeared to decline gradually as age increased, indicating that the measures obtained from older respondents were likely to be somewhat less valid indicators of the concepts they were intended to reflect. Moreover, the correlations between the concepts tended to rise with age, indicating that older respondents tended to

perceive their worlds in more global or less carefully differentiated ways than the younger. The authors also observed a rising trend with age in the estimates of response-scale-variance (method-variance); older persons might be more concerned about the possible failures in what seemed to them a testing situation and thus might have overly restricted their reactions. Finally, older people's responses seemed to be more subject to random error.

THE QUESTIONNAIRE

Responses to survey questions measuring beliefs and attitudes may be significantly altered by apparently trivial changes in the wording of the questions, the form of the questions, and the context in which the questions are asked. In addition, respondents' answers to the questions may also be affected by characteristics that are related to the content or topic of the questions. Recent studies on these four aspects of the instrument will be presented successively in this section.

Wording of the questions

Different words designed to tap the same object or feeling state can actually serve as different stimuli and influence the responses. Smith (1987) observed from several studies a large and persistent difference in support for spending priorities between "welfare" and "assistance (or caring) for the poor"; on average, favouring more "assistance for the poor" is 39 percent higher than favouring more "welfare". By analyzing the associations with other variables, it was shown that the two terms tapped slightly different dimensions. "Poor" seemed to be better for measuring people's support for programmes to equalize conditions and provide for the care of people; "welfare" on the other hand triggered more concerns about the costs of public assistance and perhaps more of a concern with waste, fraud and programme abuse.

A second study on wording effects concerns the famous "forbid-allow" effect. For an example of such an effect we cite an experiment from 1976 (Schuman and Presser, 1981, p.277). On the question "Do you think the United States should forbid public speeches against democracy?" 21 percent said Yes (forbid) and 79 percent said No (not forbid). On the alternative question "Do you think the United States should allow public speeches against democracy?" 52 percent said Yes (allow) and 48 percent said No (not allow). In this case the forbid-allow effect was 27 percent.

Hippler and Schwartz (1986) suggested that the asymmetry is due to the response behaviour of respondents who are indifferent to the issue. They hypothesized that the terms "not forbid" and "not allow" both connote weaker positions with respect to the issue than their respective affirmative counterparts. People indifferent to the issue are therefore more likely to say "No" to both question forms, to express the opinion that they neither support nor oppose the issue. In two experiments the authors found a forbid-allow effect of 23 percent and 14 percent, respectively. From ratings of the statements involved it appeared that "not allow" connoted, indeed, a weaker opposition to the issue than "forbid", and that "not forbid" connoted a weaker favourable position than "allow", confirming the first part of the hypothesis. In another experiment it was shown that the observed forbid-allow effect (16 percent) could, indeed, be attributed to the response behaviour of the respondents who were earlier identified as "indifferent" to the issue: they said "No" on both the allow- and the forbid-statement. These findings suggested that

the forbid-allow asymmetry would vanish if indifferent people were explicitly provided the possibility for expressing their indifference, e.g., "don't care" or "leave things as they are".

Formal characteristics of the questions

Responses to survey questions may be affected not only by the wording, but also by certain formal characteristics of the questions. The studies to be considered concern the effects of using an open vs. a closed question, offering vs. omitting a middle alternative, changing the number and the range of the answer categories, reversing the order of the response options, explicitly offering vs. omitting a "don't know" option, and the effects of varying the manner in which the information is presented to the respondents to base their opinions upon.

According to many scientific and political leaders, the threat of nuclear war is the most serious problem facing the United States. Yet, "threat of nuclear war" is seldom mentioned by respondents to the open question "What do you think is the most important problem facing this country today?". Schuman and al. (1986) hypothesized that the infrequent mention of nuclear war may be due to the question form; the standard open question would only elicit issues that are momentarily salient in respondents' immediate environment but not issues that are deeply important to the respondents. The authors designed a closed version of the open "most important problem"-question, offering the three issues that were chosen most frequently in the earlier open versions, along with "nuclear war". If the hypothesis is true, then the nuclear war choice should show a greater open to closed form increment than do the other categories. The hypothesis could not be confirmed: low mentioning of nuclear war could not be accounted for by the question form.

Another question form effect observed was that any given opinion was less likely to be volunteered in the open-response format than to be endorsed in the closed-response format, which confirmed the findings of earlier experiments with open vs. closed questions. Also the aggregate rankings of the four category issues were identical for the open and closed question forms. And finally, the less frequently an issue was mentioned spontaneously to the open question, the greater was its increment in choice on the closed question.

Closed-ended questions, at least some of them, can be formulated either with or without a middle alternative together with the other alternatives. What difference does it make? Bishop (1987 a) conducted several experiments to study the possible effects of offering vs. omitting a middle alternative in opinion survey questions. He also varied the placement of the middle alternative in the question. As expected, people were much more likely to select a middle response alternative when it was offered to them as part of the question than when it was omitted. The differences were about 20 percent, sometimes even more. This result confirmed the findings of Schuman and Presser (1981). It was also observed that merely mentioning the middle alternative in the preface of the question made it more likely that respondents would select it, the difference being about 15 percent.

These findings demonstrated that respondents were remarkably constrained by the form of the question. The data also suggested that the harder it was to choose between two opposite alternatives on an issue (indicated by how evenly opinions are divided in the "omitted" form), the more likely a person was to select the middle alternative when it is offered. When the middle alternative responses were excluded from the

analysis, a significant difference between the two question forms (i.e. offered and omitted middle alternative) in the percentage of opinions on the polar sides of the issue was observed in half of the twelve comparisons made. The differences, however, were modest (between 7 and 10 percent). This finding did not agree with the earlier findings of Schuman and Presser (1981).

As to the answer categories added to closed-ended questions, there are three other formal characteristics that may influence the research outcomes: the number of categories, their range, and the order in which the response options are presented to the respondents.

Studies using aggregate longitudinal data show that prosperity enhances the electoral chances of incumbents. Parallel investigations at the individual level, however, reveal that changes in citizens' personal economic well-being seem to have little or no bearing on their political preferences. In explaining this striking anomaly, Rosenstone et al. (1986) hypothesized that the question for measuring personal change in economic well-being was too poor a measurement ("Would you say that you are better off or worse off than you were a year ago? Or the same?"). The authors developed a new measurement instrument, among others, by extending the traditional three category format to a five category format. The data revealed that the three category format of the question was as valid as the new five category format (construct validity). However, the predictive validity (the effectiveness in predicting political evaluations) of the five category format appeared to be very much larger than the traditional three category format. It was concluded that the traditional question caused the political effect of changing personal economic conditions to be underestimated substantially. Clearly, items that produce greater variance in the responses are to be preferred to items that produce less variance.

Schwartz and Hippler (1987) investigated the effects of varying the range of the answer categories in questions that asked respondents to indicate the frequency of behaviours (such as hours of TV watched, church attendance). They hypothesized that respondents, when answering such closed frequency questions, would not try to recall and count all instances of the target behaviour. Instead they would make an estimation, using the range of the frequency categories offered as the frame of reference. It was found, indeed, that respondents who were offered a low frequency scale reported substantially less hours of TV watching than those who were offered a high frequency scale. The range of the frequency categories offered may also be used by respondents to infer the typical behaviour in the real world, which may then serve as a standard of comparison in making subsequent related judgements. Respondents in the low frequency scale condition would perceive themselves to watch more TV than is typical (as inferred from their low frequency scale); as a consequence they would judge their own TV consumption to be more important in their life than people in the high frequency scale condition. Data confirmed the hypothesis.

Similar effects of frequency scales on subsequent related judgements were observed in questions dealing with other topics.

Finally, the range of frequency categories offered may also determine how respondents will define the target behaviour, when it is ambiguous as to what experiences should be reported, only the serious or extreme experiences (which are less frequent) or the less extreme experiences (which are more frequent). People in the low frequency scale condition reported more concrete and more serious instances of annoying experiences than people in the high frequency scale condition. So, the frequency scale seemed to determine the specific referent of the question.

The third experiment on formal aspects of the answer-category system in

closed survey questions concerns the order in which the response options are presented to the respondent. The response option order may affect the selections the respondents make from the list, resulting in either a primacy or a recency effect. Krosnick and Alwin (1987) derived from psychological evidence the hypothesis that for visually presented options (on show cards) a primacy effect is to be expected: early items are likely to be subjected to more carefull cognitive processing than later items. The second hypothesis was, that cognitively sophisticated people are likely to be motivated to seek optimal answers; as a consequence they would be little susceptible to the order of the response options. Less sophisticated people, on the other hand, are more likely to settle for an answer that is only satisfactory; they will tend to choose the first acceptable alternative among the offered choices. In the experiment respondents were asked to indicate the most important child qualities from a list of thirteen qualities. In the alternative version the order of the items was reversed. The authors found clear evidence of a primacy effect; for some items the order effect was as large as 15 percent. The data also supported the "satisficing"-hypothesis: less sophisticated respondents were more influenced by the response-order manipulation than those who are more sophisticated.

It may be derived from the foregoing experiments that people, when answering survey questions, are constrained by the form of the questions submitted to them. Further clear evidence as to this phenomenon is provided by Bishop et al. (1986), who investigated why people are quite willing to express opinions on fictitious issues, objects and events, such as "The 1975 Public Affairs Act". The authors assumed that people would give opinions on fictitious issues because of the demand characteristics of the question or the interview. Another hypothesis was that people would give opinions on such matters because of a disposition which make them either more or less willing to admit their ignorance. Both hypotheses were tested in a repeated experiment in which three fictitious public affairs issues were inserted in the questionnaire. The questions were asked in one of three forms:

A. with an explicit "don't know"-filter, which allowed respondents to indicate that they hadn't thought much about the issue;

B. without such a filter; a "don't know"-answer was accepted without further probing;

C. without an explicit filter, but if "don't know" was volunteered the interviewers were instructed to pressure the respondent into selecting one substantial alternative.

To tap the disposition to admit ignorance one item from the Marlowe-Crowne Social Desirability Scale was included. The dispositional hypothesis could not be confirmed, possibly because the measurement of the disposition was too poor. Form A yielded 4 to 14 percent substantive answers, form B 22 to 39 percent and form C 32 to 56 percent. These findings indicate that people give opinions on fictitious issues because of the pressure to answer which is created by the form of the question and by the manner in which the "don't know" reponses are handled by the interviewer. The authors recommend giving people an explicit opportunity to say that they do not have an opinion on the subject (DK-filter). That, however, will cause a loss of 20 to 25 percent of the cases for analysis. If that is too great a cost, it may be better to stop probing altogether.

The last investigation to be mentioned in this subsection concerns the manner in which the necessary information is presented to respondents. Sometimes it is necessary to offer respondents information on the topic of the question, if it can be assumed that they are not sufficiently knowledgeable about the subject matter on which they are requested to give

their opinion. Particularly this will be the case in nationwide referenda that are organized by representative bodies to consult the nation. Because the subject matter of a referendum is mainly quite complex, it may be seriously doubted whether people have sufficient knowledge of the subject at hand for their opinions to be valuable for the decision makers.

Guided by decision theory Neijens (1987) developed the Choice Questionnaire, which claims to elicit not just people's opinions, but their informed opinions. The instrument was tested for its claims in a nationwide sample of Dutch citizens. The issue in the study was what sources of energy should be used in the future for the generation of electricity. People were delivered charts on which the alternative options were mentioned along with their possible consequences (as to the outcomes, the costs, the milieu, the likelihood and seriousness of accidents, etc.) as they were seen by experts in the field. Respondents had to evaluate successively each of the consequences of each option; finally they were asked to choose their most preferred option. By varying the manner of presenting the information the author was able to evaluate whether the Choice Questionnaire worked as intended. When no information was provided at all, that is in a situation comparable with a simple public opinion poll, 38 percent of the respondents made a final choice of good quality, that is a choice that is consistent with their evaluations of the consequences of the various options. Adding information in a narrative form did not seem to work very well in questionnaires: the number of consistent choices increased to only 48 percent. In the Choice Questionnaire 68 percent of the respondents made a consistent choice. Furthermore, between 24 and 36 percent of the respondents were aided in their decision making by the Choice Questionnaire; 32 percent did not really need the Choice Questionnaire, because they made a consistent choice without the aid of the Choice Questionnaire; 27 to 38 percent failed to make a consistent choice even with the Questionnaire aid. The main reason for the Choice Questionnaire to elicit better opinions was that respondents had to perform judgement tasks when going through the information and so were stimulated to process the presented information actively.

Context of the questions

Responses to survey questions may be affected not only by characteristics of the questions themselves, but also by the content of previously answered questions. These effects are called context effects or question order effects.

Green and Waxman (1987) investigated whether drawing attention to a threatening group would lower respondents' subsequent willingness to support the civil liberties of another, unrelated group. Tolerance-responses for Communists were established in either of two conditions: in the treatment condition the tolerance questions for Communists immediately followed similar questions for racists (a possible threatening group); in the control condition the Communist questions immediately followed similar questions for socialists (not threatening group). In addition it is hypothesized, that the treatment effect for whites would be close to zero, because "racists" are not a threatening group for whites. The racist stimulus had, indeed, no effect on whites. Blacks gave noticeably more intolerant responses on the Communist questions after having been confronted with similar questions about racists. The sequence effect varied with education: blacks with some college training were all but unaffected by the threatening stimulus, whereas blacks with less education were profoundly influenced, showing an

average of 9 percent lower tolerant responses in the treatment condition. The most likely explanation of this interaction was, that the well educated were better equipped to recognize the difference between racists and Communists. As a consequence the well educated were less likely to allow their feelings about racists to spill over into their responses about Communists, which reduced the question order effects.

To explain how question order effects occur in a particular case, Bishop (1987 b) derived from information processing theory the hypothesis that people, when answering a particular question, search in their memory only for recently stored (primed) information judged to be relevant for answering the question. From that information they will infer what their answer to the target question should be. The inference hypothesis was tested in a series of experiments, in which three questions about respondents' political knowledge were involved; one of these was easy to answer (name of the governor) whereas the two other questions were rather difficult (record of respondents' Congressman). The target question concerned respondents' interest in government and public affairs. People appeared to be less likely to say that they were interested in politics if they were asked about it after the hard knowledge questions than if they were asked about it first. This clear order effect was not reduced by increasing the number of the intervening buffer items, nor by giving people the easy knowledge question to answer between the hard questions and the target question. Only respondents who were unable to answer the difficult knowledge questions displayed less political interest on the subsequent interest-question (effect about 12 percent): from their being unable to answer the knowledge questions they inferred that they were (appearently) not so much interested in politics. Being unable to answer the easy question, however, did not have such an effect, presumably because that question was considered too easy to imply that one was not interested in politics.

In short: question order effects may occur when answers to previous questions have direct implications for an answer to a target question, regardless of when or where the context question was asked in the interview.

Content-related characteristics of the questions

Respondents' answers to survey questions may also be affected by characteristics related to the content of the questions concerned. More specifically, it is commonly assumed that people are reluctant to give true answers to questions dealing with sensitive topics, and will tend to bias their answers in the direction they perceive to be most favourable, possibly the socially desirable direction.

One such a sensitive topic is respondents' own past delinquent behaviour. Junger (1987) investigated the validity of selfreports of delinquent behaviour obtained in face-to-face interviews. In a Dutch survey 188 young boys who had trouble with the police in the past were asked to indicate on each of 19 delicts "ever done" and "done last year"; similar questions were asked with respect to their contacts with the police. The answers were compared with official record data. The validity of the delict reportings on the "ever" questions was moderate: 30 percent of all boys gave invalid answers (underreporting), the Moroccan and Turkish boys significantly more (40 percent) than the Surinam and Dutch boys (13 percent). A similar pattern was observed in reportings on the "last year" questions.

The validity of the reportings of police contacts, however, was far below the level of acceptance: 70 percent of all selfreports on the "ever"

questions were invalid, 87 percent among the Turkish boys and 60 percent among the other groups. As to the "last year" question 77 percent of all boys did falsely not report a police contact, with no significant differences between the ethnic groups. Further analysis revealed that ethnicity did not affect the validity directly, but via group attachment. That is: boys who accepted values that condemned delinquent behaviour and felt attached to the traditional values of their group were more likely to suppress their offences and police contacts. Interestingly, these were the same factors assumed to influence delinquent behaviour itself.

The validity of self-reportings on sensitive questions was more generally investigated by Marquis et al. (1986), who summarized the results of 52 studies in which survey responses to factual questions on sensitive topics were compared with external criterion data, such as official records or biological test results. The studies were divided into six sensitive topic areas: receiving welfare (six studies), income (six studies), drug use (eight studies), alcohol use (seven studies), criminal history (twelve studies), embarassing medical conditions (thirteen studies). For each study the direction and size of the response bias was estimated, as well as the response reliability (ratio of the record values variance to the survey values variance). It was observed that for most topic areas the distribution of the bias estimates centered near zero, or no net response bias. Taken as a whole, the data suggested that, on the average, respondents did not deny or underreport undesirable attributes or sensitive personal information as is commonly assumed. The response problem in sensitive topic areas seemed to be rather a reliability problem: except for income, the reliability estimates for the topics fell in the problem range, defined as less than 0.70 reliability. Unreliabilities of that magnitude could seriously distort correlations, regressions and other relationships.

Responses to factual survey questions may also be in error for quite another reason: the questions are too difficult for the average respondent to answer correctly. Hodson and Seitz (1987) compared the self-reports of workers about the industry of their employing company with the published reports of company's product lines. Using fourteen broad categories 6 percent mismatches were observed; when 81 and 213 categories were used 18 percent and 21 percent mismatches were found, respectively. When the self-reports were compared with the company's first reported industry the mismatches were respectively 19, 40 and 46 percent. These results indicate that there are serious problems in using respondents' self-reports for measuring organizational or technological characteristics of their employing companies. In general, the mismatches did not appear to be related with individual characteristics of the workers (gender, annual earnings, education, status), but rather with structural aspects of their employment relations (size of the company, plant size, vertical integration, prevalence of subcontracting, conglomerate interlocks).

In the foregoing overview, the attention was directed to recently published new evidence on response effects stemming from aspects of the questionnaire used in surveys. To be complete some other recent studies in this field will be mentioned shortly.

In a summarizing study Hippler and Schwartz (1987) reviewed a lot of past experiments on the wording, form and context of survey questions within the theoretical framework of cognitive psychology.

Grémy (1987) summarized the past French experiments in this field, experiments that remain mostly unmentioned in overviews.

Molenaar (1986) explored the possibilities of a non-experimental design for studying question wording and question form effects to overcome the problem that the results of single experiments can not be validly

generalized over other questions.

Converse and Presser (1986) published a small practical guide for formulating questions and constructing questionnaires.

CIRCUMSTANCES

Respondents' answers to survey questions may, in part, be affected by the circumstances in which they are requested to answer the questions or to fill in the questionnaire. On this particular topic only a few investigations have been published recently. The first study deals with the possible effects of the presence of others during the interview, the second is about the effects of the mode of administration, and the last study concerns the possible effects of the pressure that is exerted on respondents to cooperate with the survey.

Silver et al. (1986 a) estimated that between 20 and 30 percent of the respondents in American post-election surveys claim to have voted whereas they actually did not vote. If overreporting of voting is caused by the tendency of respondents to exaggerate their performance of socially desirable behaviour, as is commonly assumed, then it may be expected that the tendency to overreport voting is affected by whether third parties are present during the interview. From election studies it is observed that third parties, mostly respondents' spouse or a child, were present in about 50 percent of the interviews, in spite of the instruction to interview the respondent alone. Analysis of the data, however, revealed that overreporting was not affected by the presence of others during the interview.

Quite another circumstance that may influence the answer behaviour of the respondents is the manner in which respondents' answers to the survey questions are administered. The mode of administration is obviously different for face-to-face interviews, telephone interviews, paper-and-pencil questionnaires, and computer steered interviewing procedures.

De Pijper and Saris (1986) developed a new general computer program for automatic tele-interviewing. Besides the features that are available in other interview programs, their computer program also contains facilities for using psychophysical scaling techniques, handling multiple response questions, semi-automatic coding of responses to open questions, and others. Respondents can easily handle the computer procedures for themselves at home. Interviewers are not needed anymore: the questionnaire and the data can be send back and forth to the respondent through telecommunication facilities. Elsewhere the authors discuss their experiences with the program, the possibilities for their application and the various settings in which the program can be used (Saris and De Pijper, 1986).

Data with respect to the quality of the answers obtained by computer steered interviewing procedures are relatively scarce by now. Kiesler and Sproull (1986) recently compared the responses obtained in an electronic survey with those obtained in an equivalent paper survey. The experiment was conducted in a relatively computer-intensive environment and on a not representative sample. The questions, most of which were closed-ended, dealt with health and personal characteristics. The response rate in the paper survey appeared to be higher than in the electronic survey, but the electronic survey was returned more quickly. The electronic survey data contained fewer item completion mistakes in the closed ended items, fewer refusals to answer questions and fewer socially desirable answers. The modes did not differ in the length of answers to open questions, the

number of personal pronouns and trait descriptions; also no differences were observed in agreeing tendencies in responses to Likert items, nor in response-extremity. In conclusion it seemed that, although the responses were largely similar, the modes were not interchangeable.

Dillman (1978) developed a set of procedures, called the Total Design Method (TDM), to improve the response rate in mail surveys. Apart from its being effective indeed, the pressure exerted by TDM on potential respondents to co-operate may enhance the influence of factors assumed to threaten the validity of the data: social desirability, respondents' attitude to the study, perceived question threat, item nonresponse and response patterns (scalability of items in scales). De Leeuw and Hox (1987 a) compared the quality threatening factors (measured by appropriate scales) in a survey in which two versions of a loneliness scale were included: a neutral version, designed to measure common feelings of loneliness, and an extreme version, designed to measure feelings of extreme isolation. Respondents were grouped as to whether their questionnaire was returned after the first, the second and the third (certified) mailing. Respondents in the first mailing group appeared to have more positive attitudes towards the study and perceived less threat for the questionnaire than respondents in the second and the third group. It was argued that these differences reflected respondents' reaction on the increasing pressure exerted by TDM. No differences were found with respect to the percentage missing data and the scalability of the responses on both loneliness scales; but correlations were observed between the loneliness scale values and response biasing factors (attitudes towards the study, perceived question threat, missing values), particularly in the extreme version. It was concluded that the use of follow-ups may increase the risk of response bias.

NON-RESPONSE

The success of a survey depends for a major part on the rate of response obtained from among the eligible people. So it is important to study how trends in non-response are going through time, to find out the causes of non-response, in particular the reasons to refuse any co-operation with the survey, and to investigate what measures could be taken to increase the response rate, or to keep the response rate at the present level at least. Two sets of studies relevant in this problem area can be distinguished. The first set concerns the general social conditions that may be assumed to affect the non-response rate (and, the quality of the responses obtained); the second set deals with problems and aspects of non-response itself.

Social conditions that may influence future non-response

Survey researchers are strongly dependent on the willingness of the public to co-operate with the survey. It is reasonable to assume that this willingness is largely dependent on the level of respondents' past participation, their attitudes toward the opinion research industry, and attitudes toward being interviewed. Some recent surveys on surveys present evidence about these problem areas.

Roper (1986) found in a nationwide American sample that more than a half of the respondents reported that they had been interviewed once or more times in the past. Schleifer (1986) found a similar participation level; among the past year participants (23 percent of the sample), 56 percent had been interviewed two or more times, and 18 percent even four

or more times. Particularly young better educated people with higher income, and females are comparatively oversurveyed. Goyder (1986), who found similar high participation levels in a Canadian local survey, presented strong evidence that the greater the number of times a person was asked to participate in a survey, the more unfavourable was his or her attitude toward surveys.

As to the attitudes toward the polling industry, Roper (1986) concluded from his study that the American public was skeptical about the value of polls, but that few people are entirely negative. Schleifer (1986) reached a similar conclusion: the public did not object to surveys and polls as such (they were serving a useful purpose), but many people remained suspicious of the industry's methods and motives. Goyder (1986) showed that most opinions about the survey industry were unstable surface opinions; the more general opinions, however, seemed to reflect a stable attitude.

As to the attitudes toward being interviewed: Roper (1986) found that 51 percent rated polls an enjoyable and satisfactory experience; only 5 percent judged polls an annoying and unsatisfactory experience, mainly because of their length (take too much of your time) and because they interrupt you at inconvenient times. Other data revealed that the percentage of those critical of being interviewed has grown over the last ten years. Schleifer (1986) observed a downward trend in those past year participants who rated their last interview as a pleasant experience. Increasing numbers said that the interview was too long and contained confusing and overly personal questions. The most critical factor in shaping respondents' attitude toward being interviewed seemed to be the personality of the interviewer; questionnaire design, topic and inconvenient time were also among the critical factors.

These data suggest that interviews should be kept short, that questions should be clearly worded, and that unnecessary personal and sensitive questions should be eliminated from the questionnaires.

Non-response

In all surveys, some part of the eligible people can not be reached in spite of all reasonable efforts. Secondly, part of those who are contacted refuses to co-operate at all. Thirdly, part of those who agree to co-operate refuse to answer particular questions (item non-response). All these non-responses may seriously distort the research outcomes.

Using data from five annually repeated surveys held by the central statistical agency in the Netherlands, Mokken (1987) clearly demonstrated that non-response has increased steadily through time, from about 20 percent in 1972 to 35 percent in 1984. The number of refusals (number of contacted persons who refuse any co-operation) also increased, from about 15 percent in 1972 to 25 percent in 1984. Direct data as to the causes of this alarming phenomenon are not available, but it is hypothesized that the trend is caused by the growing inaccessibility of the people (undomestic), the growing privacy-awareness among the public, and oversurveying. Mokken (1987) also mentioned a recent Dutch census-like survey on the existing needs for houses, in which it was observed that non-response was related to age: the younger were more willing to participate but less accessible, whereas the older were more accessible but less willing to co-operate. Non-response appeared to be lower among the married (appearently because they are more at home) and higher among members of one- or two-person-households. In rural areas there was more willingness to co-operate than in urban areas.

That neither total non-response (refusals to co-operate at all) nor

item non-response can be assumed to be random, was demonstrated by Hermann and String (1986) in a German mail survey among a rather specific public, namely criminal judges and public prosecutors. They specifically tested the hypothesis that the causes of total non-response were largely the same as those of item non-response. The questions dealt with the significance of criminology in making penal decisions and with the willingness to reform. After seven weeks a short questionnaire was send containing only questions that are relevant for establishing the possible causes for answering or not answering the former questionnaire. Path analysis revealed that the strong relationship between answering/not answering the former and the second questionnaire is caused by a stable action orientation factor. Item non-response also appeared not to be random, but dependent on the content of the questions. Thirdly, the path coefficients in the models to explain total non-response and item non-response respectively, appeared to be very similar, indicating that the causes of total non-response and item non-response are largely the same. The stable action orientation factor consisted, in both cases, of social desirability and interest in the topic of the questions.

What can be done to improve the response rate?

Along with some sustaining evidence, Mokken (1987) suggested that the response rate in face-to-face surveys could be improved by raising the number of visits: going from three to four visits, the response rate is estimated to increase by 2 percent. The effect of sending an introduction letter in advance was estimated to be a 4 percent higher response rate. In mail surveys, however, sending an advance letter can have a rather different effect. De Leeuw and Hox (1987 b) found that sending an advance letter one week before the actual mailing significantly improved the initial response rate (by 9 percent), but not the final response rate; it seemed to have only the effect of speeding up the data collection. They also found that inserting a reminder by telephone after four weeks instead of a certified mail after six weeks did not significantly increase the final response rate. They argued that because of the high saliency of the questionnaire, the initial response rate was already maximum before the telephone reminder was introduced.

Another possible measure to improve the response rate in surveys is payments. Payments are effective in improving the response rate in surveys, as has been demonstrated in past experiments. It may be assumed that the size of payment is a cue for respondents to judge a survey's importance. The mode of payment may offer an additional cue: payment in advance may be expected to lead to higher response rates than payment only upon completion of the questionnaire, which is a rather bold and unusual gesture. Berry and Kanouse (1987) sent a questionnaire to more than 2000 physicians, half with pre-payment and half with post-payment. The response rate, defined as the number of completed questionnaires divided by the number of eligible cases, turned out to be substantially higher in the pre-payment condition than in the post-payment condition. The number of contacted cases was also substantially higher, whereas the refusal rate among the contacted was lower. So, pre-payment should be considered for populations that are difficult to contact, such as physicians.

Another factor that is held to improve the response rate and the quality of the responses, particularly the quality of the responses to sensitive questions, is the explicit assurance that the information will be kept confidential. Survey interviews generally include a guarantee of confidentiality in an introductory message, but such an introductory assurance does not seem to have much impact on the response rate or the

data quality. In a telephone survey, Frey (1986) inserted a confidentiality reminder just before the demographic (sensitive personal) questions were going to be asked. Instead of lowering the non-response on the subsequent questions, the reminder appeared to yield higher item non-response rates. Maybe an inserted reminder undoes the effect of the introduction and of the rapport established between interviewer and respondent; it seems to alert respondents to be careful in answering the subsequent questions.

THEORETICAL STUDIES

Most of the studies considered in the foregoing pages present some limited theory from which hypotheses and predictions are derived about the possible effects of the particular variable under study. General theoretical frameworks for the conceptualization of the processes underlying responding survey questions are relatively scarce. Such general frameworks, however, are highly valuable, because they can explain how particular kinds of response effects are related to each other by their stemming from the same underlying processes. In this last section we will draw attention to some recently published theoretical studies.

Most general is the theoretical framework developed by Esser (1986). Esser discusses extensively the concepts "true value", "bias", "situational effects" and other related concepts from the perspective of social action theory, starting from the idea that the basic mechanism in any social process, including the interview, is peoples' striving for optimal personal satisfaction (utility) from the contact. Respondents are assumed to evaluate the possible alternative answers for their expected consequences and to choose the answer that has comparatively the highest utility. In this process personal dispositions of the respondents are involved as well as the situation as they define it. In this view respondents behave like rational actors. This rational behaviour can be specified by decision theory, particularly the SEU-model which describes how people, by weighting the profits and losses of alternative actions, come to choose the action with the highest "Subjective Expected Utility" (SEU-value).

Esser describes in detail what the model looks like when applied to the behaviour of respondents in interviews. For the sake of simplicity the SEU-model is collapsed by distinguishing only two basic answer tendencies of respondents: the tendency to disclose the true answer, called the valid answer tendency, and the tendency to base the answer on situational cues. Both tendencies are then dichotomized giving rise to four types.

Type I: both tendencies are weak (because the "true" attitude is non-existent, weak or insignificant, the question is ambiguous, the situation-aspects are too unclear); respondents' answer will be random, resulting in unreliable data.
Type II: the valid answer tendency is strong (the attitude is existent, question is clear) but the other tendency is weak (the situation is ambiguous). In this case, the respondent will choose the true answer as the "best" answer, resulting in reliable and valid data.
Type III: the valid answer tendency is weak but the situation related answer tendency is strong. Although the data will be highly reliable and formally valid in this case, one does not measure what is intended to be measured. The data will be biased; systematic situational effects can be established.
Type IV: both tendencies are equally strong; in this

"inconsistency"-case the answers are unpredictable; the data may be reliable or unreliable, valid or biased.

The ideas of Esser offer an interesting device for understanding the nature of "true value", "bias" and situational effects. Whether the ideas can be tested empirically, however, remains to be seen. It seems to be not at all easy to measure whether respondents follow the rules of the SEU-model indeed when giving their answers.

The theoretical study of Tourangeau (1987) is more specific. To understand the mechanisms that produce response effects in attitudinal questions, Tourangeau describes the process of answering attitudinal questions from a cognitive perspective. An attitude is represented cognitively as an organized package of beliefs, feelings and knowledge about classes of situations and things, called a "script". Such a high-level cognitive scheme guides the interpretation of the question, organizes the retrieval of relevant beliefs, feelings and knowledge, shapes the process of rendering a judgement and the process of selecting a response. Which script will be activated by a particular issue depends on the relative overall "strength" of the script (the frequency of its use, the importance in life), and on the situational availability of alternative scripts, determined by their being recently used and by one's momentary mood state. The situational availability of particular scripts may be affected by the wording of a question (feeling tone) and by the answers given to previous questions (context). Survey researchers, however, do not intend to measure these short term available scripts: from their perspective they represent unintended response effects. For abstract, multifacetted items a larger number of scripts is likely to be relevant than for concrete items. Therefore, response effects are most likely to occur with abstract items.

Strack and Martin (1987) conceptualize context effects in answering survey questions from the perspective of cognitive information processing. They distinguish four major stages in answering survey questions: (1) understanding the question, (2) generating an opinion on the issue, (3) fitting the opinion into the provided format, and (4) editing the overt answer. In each of the stages context effects may occur.

(1) To understand the semantic meaning of an ambiguous target question, respondents will rely on concepts that are activated by the preceding questions. The intended meaning of the question will be inferred from the content of the question, from the response scale (Schwartz and Hippler, 1987), and from the answers given to preceding questions. These (answers to) preceding questions inform the respondent about what the questioner already knows. According to the normal conversation rules, he assumes that the interviewer asks for new information in the target question, that is information that is not yet provided in the answers to preceding questions; this explains the wellknown "part-whole" or "specific-general" context effects.

(2) When generating an opinion, the respondent may recall an opinion that is already stored in memory, as will be the case with important issues. But if he has not previously evaluated the issue on the dimension of the survey question, he must access the relevant information from memory. The accessibility of information is determined by the frequency and the recency of its prior activation. Previous questions may activate a normative standard or a mood state. They may also activate respondents' own specific knowledge or past behaviour, which provides information to infer his own attitudes (Bishop, 1987). Furthermore, previous questions may direct the attention to specific domains or aspects which may subsequently determine the global evaluation in the target question.

Finally, previous questions may activate a particular exemplar of the category in question. Subsequent evaluation of the target may then be based on features that the target has in common with the invoked exemplar (assimilation or consistency effect) which will specifically occur when the items appear to be interrelated, or on features in which they differ (contrast effect) which is likely to occur when the items do not appear to be interrelated.

(3) When formatting and choosing the response, the range of the response categories provided may be used to infer the average response and to derive one's relative position in a frequency distribution (Schwartz and Hippler, 1987).

(4) Finally, when editing the response, social desirability effects are likely to occur; negative self-reports are considered socially undesirable which will lead to positively biased reportings.

To end this section, two other rather specific theoretical explorations will be mentioned briefly. Ostrom (1987) contrasts two alternative conceptual approaches toward understanding how people respond to bipolar subjective survey questions: the traditional dimensional approach and the information processing approach. Strube (1987) discusses the role of memory in answering survey questions, specifically questions that ask for frequency judgements about past events or behaviours and questions that ask for the date of particular events; he proposes some practical measures to improve such judgements.

CONCLUDING REMARKS

To conclude we will make a general remark, particularly about the practical relevance of experiments on aspects of the questionnaire. Because in attitudinal questions "true" values are unknown, experimental findings on question wording, form and context can not unequivocally reveal which question wording, form or context is the best. The most general conclusion that can be drawn from the experimental evidence is, that the form, the wording and the context of the questions will largely determine the frame of reference for respondents to stay within.

As soon as the respondent consents to be interviewed, he agrees to play the game according to the rules. The experimental findings may help the researcher to identify which constraints are likely to be introduced by which characteristics of the questions, the questionnaire, and, to be sure, by other features of the procedure as well. The experimental findings also serve as a warning to be careful in interpreting the responses obtained.

LITERATURE

Andrews, F.M. and A.R. Herzog (1986). "The quality of survey data as related to age of respondents", Journal of the American Statistical Association 81: 381-389.

Berry, S.H. and D.E. Kanouse (1987). "Physician response to a mailed survey; an experiment in timing of payment", Public Opinion Quarterly 51: 102-114.

Bishop, G.F. (1987 a). "Experiments with the middle response alternative in survey questions", Public Opinion Quarterly 51: 220-232.

Bishop, G.F. (1987 b). "Context effects on self-perceptions of interest in

government and public affairs" in: H-J. Hippler, N. Schwartz and S. Sudman (eds.) Social Information Processing and Survey Methodology. New York: Springer Verlag, pp. 179-199.
Bishop, G.F., A.J. Tuchfarber and R.W. Oldendick (1986). "Opinions on fictitious issues: the pressure to answer survey questions", Public Opinion Quarterly 50: 240-250.
Brenner, M. (1985). "Survey interviewing", in: M. Brenner, J. Brown and D. Canter (eds.) The Research Interview: Uses and Approaches. London: Academic Press, pp. 9-36.
Bruinsma, C. (1987). Systematic Errors in Survey-Interviews; a Methodological Study on the Effect of Significant Others and Other Sources of Bias on Evaluations of Job Qualities and Union Actions. Amsterdam: VU-Uitgeverij. Dissertation.
Converse, J.M. and S. Presser (1987). Survey Questions; Handcrafting the Standardized Questionnaire. Beverly Hills: Sage; Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-063.
Dijkstra, W. (1987). "Interviewergedrag bij informatie-inwinnende gesprekken", in: J. de Jong-Gierveld and J. van der Zouwen (eds.) De vragenlijst in het sociaal onderzoek; een confrontatie van onderzoekspraktijk en -methodiek. Deventer: Van Loghum Slaterus, pp. 131-149.
Dijkstra, W., L. van der Veen and J. van der Zouwen (1985). "A field experiment on interviewer-respondent interaction", in: M. Brenner, J. Brown and D. Canter (eds.) The Research Interview: Uses and Approaches. London: Academic Press, pp. 37-63.
Dijkstra, W. and J. van der Zouwen (1987). "Styles of interviewing and the social context of the survey-interview", in: H-J. Hippler, N. Schwartz and S. Sudman (eds.) Social Information Processing and Survey Methodology. New York: Springer-Verlag, pp. 200-211.
Dillman, D.A. (1978). Mail and Telephone Surveys; the Total Design Method. New York: Wiley.
Esser, H. (1986). "Können Befragte lügen? Zum Konzept des "wahren Wertes" im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung", Kölner Zeitschrift für Soziologie und Sozialpsychologie 38: 314-336.
Frey, J.H. (1986). "An experiment with a confidentiality reminder in a telephone survey", Public Opinion Quarterly 50: 267-269.
Goyder, J. (1986). "Survey on surveys: limitations and potentialities", Public Opinion Quarterly 50: 27-41.
Green, D.P. and L.M. Waxman (1987). "Direct threat and political tolerance; an experimental analysis of the tolerance of blacks toward racists", Public Opinion Quarterly 51: 149-165.
Grémy, J-P. (1987). "Les expériences françaises sur la formulation des questions d'enquête; résultats d'un premier inventaire", Revue française de sociologie 28: 567-599.
Groves, R.M. and L.J. Magilavy (1986). "Measuring and explaining interviewer effects in centralized telephone surveys", Public Opinion Quarterly 50: 251-266.
Hermann, D. and F. Streng (1986). "Das Dunkelfeld der Befragung; unit-nonresponse und item-nonresponse bei einer schriftlichen Befragung von Richtern und Staatsanwälten", Kölner Zeitschrift für Soziologie und Sozialpsychologie 38: 337-351.
Hippler, H-J. and N. Schwartz (1986). "Not forbidding isn't allowing: the cognitive basis of the forbid-allow asymmetry", Public Opinion Quarterly 50: 87-96.
Hippler, H-J. and N. Schwartz (1987). "Response effects in surveys" in:

H-J. Hippler, N. Schwartz and S. Sudman (eds.) Social Information Processing and Survey Methodology. New York: Springer-Verlag, pp. 102-122.

Hodson, R. and P. Seitz (1987). "Identifying workers' industries: survey responses versus corporate product lines", Social Science Research 16: 154-168.

Junger, M. (1987). "Validiteit van zelfrapportages van delinquentie bij jongeren uit ethnische minderheden", in: J. de Jong-Gierveld and J. van der Zouwen (eds.) De vragenlijst in het sociaal onderzoek; een confrontatie van onderzoekspraktijk en -methodiek. Deventer: Van Loghum Slaterus, pp. 153-174.

Kiesler, S. and L.S. Sproull (1986). "Response effects in the electronic survey", Public Opinion Quarterly 50: 402-413.

Krosnick, J.A. and D.F. Alwin (1987). "An evaluation of a cognitive theory of response-order effects in survey measurement", Public Opinion Quarterly 51: 201-219.

Leeuw, E.D. and J.J. Hox (1987 a). "Artifacts in mail surveys: the influence of Dillman's Total Design Method on the quality of the responses", in: W.E. Saris and I.N. Gallhofer (eds.) Sociometric Research, Vol. I. London: MacMillan, pp. 61-73.

Leeuw, E.D. and J.J. Hox (1987 b). "The effect of an advance letter, a reminder by telephone, and the enclosure of a second questionnaire on the response to mail surveys", Kwantitatieve Methoden 25: 127-139.

Loosveldt, G. (1986). "Kwaliteitscontrole van survey-data; effecten van het interviewergedrag", Economisch en Sociaal Tijdschrift 40: 445-460.

Marquis, K.H., M.S. Marquis and J.M. Polich (1986). "Response bias and reliability in sensitive topic surveys", Journal of the American Statistical Association 81: 381-389.

Mokken, R.J. (1987). "De vragenlijst in processie", in: J. de Jong-Gierveld and J. van der Zouwen (eds.) De vragenlijst in het sociaal onderzoek; een confrontatie van onderzoekspraktijk en -methodiek. Deventer: Van Loghum Slaterus, pp. 25-47.

Molenaar, N.J. (1986). Formuleringseffecten in survey-interviews; een non-experimenteel onderzoek. Amsterdam: VU-Uitgeverij. Dissertation.

Neijens, P.C. (1987). The Choice Questionnaire; Design and Evaluation of an Instrument for Collecting Informed Opinions of a Population. Amsterdam: Free University Press. Dissertation.

Oksenberg, L., L. Coleman and C.F. Cannell (1986). "Interviewers' voices and refusel rates in telephone surveys", Public Opinion Quarterly 50: 97-111.

Ostrom, T.M. (1987). "Bipolar survey items: an information processing perspective" in: H-J. Hippler, N. Schwartz, S. Sudman (eds.) Social Information Processing and Survey Methodology. New York: Springer-Verlag, pp. 71-85.

Pijper, W.M. de and W.E. Saris (1986). "The formulation of interviews using the program INTERV". Amsterdam: Sociometric Research Foundation.

Porst, R. and K. Zeifang (1987). "A description of the german general social survey test-retest study and a report on the stabilities of the sociodemographic variables", Sociological Methods & Research 15: 177-218.

Reese, S.D., W.A. Danielson, P.J. Shoemaker, T-K. Chang and H-L. Hsu (1986). "Ethnicity-of-interviewer effects among Mexican-Americans and Anglos", Public Opinion Quarterly 50: 563-572.

Roper, B.W. (1986). "Evaluating polls with poll data", Public Opinion Quarterly 50: 10-16.

Rosenstone, S.J., L.M. Hansen and D.R. Kinder (1986). "Measuring change in personal economic well-being", Public Opinion Quarterly 50: 176-192.

Saris, W.E. and W.M de Pijper (1986). "Computer assisted interviewing using home computers", European Research 14: 144-150.
Schleifer, S. (1986). "Trends in attitudes toward and participation in survey research", Public Opinion Quarterly 50: 17-26.
Schuman, H., J. Ludwig and J.A. Krosnick (1986). "The perceived threat of nuclear war, salience, and open questions", Public Opinion Quarterly 50: 519-536.
Schuman, H. and S. Presser (1981). Questions and Answers in Attitude Surveys. London: Academic Press.
Schwartz, N. and H-J. Hippler (1987). "What response scales may tell your respondents: informative functions of response alternatives" in: H-J. Hippler, N. Schwartz and S. Sudman (eds.) Social Information Processing and Survey Methodology. New York: Springer-Verlag, pp. 163-178.
Silver, B.D., P.R. Abramson and B.A. Anderson (1986 a). "The presence of others and overreporting of voting in american national elections", Public Opinion Quarterly 50: 228-239.
Silver, B.D., B.A. Anderson and P.R. Abramson (1986 b). "Who overreports voting?", American Political Science Review 80: 613-624.
Smith, T.W. (1987). "That which we call welfare by any other name would smell sweeter; an analysis of the impact of question wording on response patterns", Public Opinion Quarterly 51: 75-83.
Strack, F. and L.L. Martin (1987). "Thinking, judging, and communicating: a process account of context effects in attitude surveys", in: H-J. Hippler, N. Schwartz, S. Sudman (eds.) Social Information Processing and Survey Methodology. New York: Springer-Verlag, pp. 123-148.
Strube, G. (1987). "Answering survey questions: the role of memory" in: H-J. Hippler, N. Schwartz, S. Sudman (eds.) Social Information Processing and Survey Methodology. New York: Springer-Verlag, pp. 86-101.
Tourangeau, R. (1987). "Attitude measurement: a cognitive perspective", in: H-J. Hippler, N. Schwartz and S. Sudman (eds.) Social Information Processing and Survey Methodology. New York: Springer-Verlag, pp. 149-162.
Zouwen, J. van der, W. Dijkstra and J. van de Bovenkamp (1986). "The control of interaction processes in survey interviews", in: F. Geyer and J. van der Zouwen (eds.) Sociocybernetic Paradoxes; Observation, Control and Evolution of Self-steering Systems. London: Sage, pp. 55-63.

Ontvangen: 26-04-1989
Geaccepteerd: 26-10-1989