CONSTRUCTING QUESTIONS FOR INTERVIEWS AND QUESTIONNAIRES

THEORY AND PRACTICE IN SOCIAL RESEARCH

William Foddy

Department of Anthropology and Sociology.
Monash University

CAMBRIDGE UNIVERSITY PRESS
1993
A major issue that is addressed is the question of the degree to which the use of answers respondents are required to give must be specified if successful communication is to occur between interviewers and respondents. This issue has been at the centre of many past methodological discussions and I have corded it central significance here. Indeed, the principal thesis advanced in this book is that most of the problems associated with the construction of statements are either avoided or lessened in importance by a clear specification of the kind of answers that respondents should give. I hasten to add that by the phrase 'kind of answers' I do not want to suggest that respondents could be pressured into giving this or that substantive response. The concept of 'kind of answer' is explained at length in chapters 3–6.

Finally, I would like to stress that, because this book is an attempt to deal with fundamental aspects of question-answer behaviour, it is directed at all those who use questions in social research. Although many of the examples I have used to illustrate particular issues have been taken from the survey research literature, the matters that they raise have a wider relevance. Certainly ad in mind social science in general, and not just social surveys, as I worked on the chapters. My hope is that sociologists, anthropologists, political scientists, psychologists, social psychologists, social policy researchers, social workers, pollsters and market researchers will all find the contents of this text useful.

William Foddie

Chapter 1

AN INITIAL STATEMENT OF THE PROBLEM

There is no doubt that the use of verbal data has come to dominate the social sciences. Asking questions is widely accepted as a cost-efficient (and sometimes the only) way, of gathering information about past behaviour and experiences, private actions and motives, and beliefs, values and attitudes (i.e. subjective variables that cannot be measured directly). A review of practices adopted in the 1960s revealed that:

the sociologist today limits himself rather generally to the construction and conduct of questionnaires and interviews. In 1940–41, 25 per cent of the 85 empirical studies depended on interviews and questionnaires for their data; in 1965–66, 48 per cent did. However, if we consider studies based on secondary data that, in turn, derived from interviews, then 64 per cent of the 156 research papers in the latter two years were based on such verbal reports.

Increasingly, these verbal reports are limited to expressions of attitudes, feelings, and opinions rather than to factual accounts of past behaviour and interactions. In 1940–41, 8 of the 22 studies using questionnaires and interviews obtained statements about opinions or feelings, 6 focused on actual behaviour and 8 gathered information about both past behaviour and attitudes. In 1965–66, 49 of 66 studies in which interviews were collected dealt only with opinions and sentiment, 6 with behaviour and 8 with both behaviour and attitudes. It would seem that our colleagues tend to ignore actual behavioural patterns and also fail to come to grips with the fundamental problem of the relation of attitudes and sentiment to behaviour. To an even greater extent, sociology is becoming the study of verbally expressed sentiments and feelings, rather than an analysis of human performance. (Brown and Gilmartin, 1969:288)
This was the situation in the sixties but there is no reason to believe that the pattern would be any different today — if anything, it is likely to be stronger.

Such an entrenched interest in the use of verbal data would not, in itself, be a bad thing if it could be shown that it always, or even usually, leads to valid conclusions in social research. Unfortunately, it must be admitted that our ability to construct questions which produce data that are reliable and lead to valid conclusions has not been very impressive to date. What is more, recognition of the prevailing situation is far from new. In a report of a survey of expert opinions, Hovde (1936) noted that 74 per cent of the experts who responded mentioned improperly worded questions as a principal defect in commercial social research. The next most frequently mentioned complaint concerned the related issue of faulty interpretations (58 per cent). These figures contrast with 52 per cent mentioning improper statistical methods. The lack of progress since Hovde reported his findings is indicated by Belson’s (1986:56) conclusions that the principal causes of error in the gathering of data through survey procedures are:
(a) respondents’ failure to understand questions as intended;
(b) a lack of effort, or interest, on the part of respondents;
(c) respondents’ unwillingness to admit to certain attitudes or behaviours;
(d) the failure of respondents’ memory or comprehension processes in the stressed conditions of the interview; and,
(e) interviewer failures of various kinds (e.g. the tendency to change wording, failures in presentation procedures and the adoption of faulty recording procedures).

Examples that illustrate the inadequacy of many of the questions that have been used in social research in the past

It is not difficult to find examples to reinforce the claim that there is a great deal of scope for improving the quality of the data we collect for social research. All ten problems discussed in the next few pages demonstrate this.

Factual questions sometimes elicit invalid answers

Palmer (cited by Deming, 1944) found that, when respondents in a Philadelphia study were re-interviewed eight to ten days after an initial interview, 10 per cent of the reported ages differed by one or more years between the interviews. Likewise, Parry and Crossley (1950) reported that objective checks revealed that 5–17 per cent of a random sample of over 900 Denver residents gave incorrect answers to a series of factual questions. The questions included whether or not respondents had registered and voted in various elections, had contributed to the community chest; and possessed library cards and driving licences, as well as details of car ownership.

If questions concerning such simple and apparently objective matters as 'age' elicit inaccurate data, one must wonder about the validity problems that might be associated with more threatening, more complex, or less well known issues. A summary of the findings of a number of health surveys published by Marquis (1970) indicates how serious the problem might be. Typically 12–17 per cent of known hospital episodes, 23–26 per cent of recorded visits to physicians, and at least 50 per cent of the chronic and acute conditions listed in medical records were not disclosed by respondents.

The relationship between what respondents say they do and what they actually do is not always very strong

Social scientists have long been tempted to assume that respondents’ behaviour is congruent with their attitudes. Yet the evidence for this link has never been very strong. This issue was first given salience by LaPiere (1934/35) in a paper which has subsequently been discussed by a number of writers (e.g. Deutscher, 1966 and 1975; Phillips, 1971; Schuman and Johnson, 1976). LaPiere spent some weeks during the 1930s travelling around the United States with a Chinese couple. He kept a record of the way in which they were treated at sixty-six hotels and motels in which they had wanted to stay and 184 restaurants and cafes in which they had wanted to eat — only one establishment had refused them service. Six months later, LaPiere wrote to the places in which they had either been given accommodation or meals, asking the proprietors of each establishment if they would accept members of the Chinese race as guests. Fifty per cent replied to his letters and, of these, 90 per cent said ‘No!’. This finding focused the spotlight on the apparent fact that respondents do not always do what they say they do.

Findings like those reported by LaPiere have led a number of methodologists (e.g. Cicourel, 1964, 1982; Deutscher, 1966, 1973; Phillips, 1971; Douglas, 1985; Briggs, 1986) to argue that social and social psychological factors which operate in the interview situation invalidate most, if not all, attempts to predict behaviour on the basis of verbally expressed attitudes. The general argument is perhaps extreme; nevertheless, the evidence for a relationship between attitudes and behaviours has always been weak. While this observation might reflect the true nature of the relationship between the concepts, it is also possible that it reflects either a lack of clear conceptualisation of what is being measured and an inadequate theoretical explication of the assumed link between the concepts (see e.g. Weigel and Newman, 1976), or the use of
inadequate questions to test hypotheses — this last possibility is discussed in chapter 11.

Respondents’ attitudes, beliefs, opinions, habits, interests often seem to be extraordinarily unstable

Twentieth-century social science has predominantly been directed by behaviourist, reductionist and naturalist premises — namely that the objects of inquiry (i.e. the stuff out there’) has two primary properties: stability and accessibility. And yet available evidence suggests that many sorts of respondents’ answers are strikingly variable over time. Converse (1964), for example, reports very low correlations between attitudes expressed by the same respondents over a two-year period. In another study, Bishop et al. (1984) found that respondents were less likely to claim that they follow what is going on in public affairs if they had first responded to a set of difficult questions about a congressman’s record than if they had to answer the questions about the congressman’s record after they had reported their own interest in public affairs. In a third study, Gritting (1986) asked respondents the same question (designed to measure their attitudes toward the setting up of a gambling casino in their community) at the start and at the end of an interview schedule. Gritting reports that 17.6 per cent of the respondents changed their position during the course of the interview.

The results of the studies that have just been reviewed leave us in the position of not knowing whether the observed variation is due to: true variability in the respondents’ memory processes (a topic that we will return to in chapter 7), inadequacies or instability in the interpretation of the questions themselves, the impact of cognitions that have been stimulated by earlier questions, the impact of social-psychological variables (e.g. interviewer-respondent status differences — discussed in chapter 9); or the tendency for respondents to spontaneously ask interviewers for their views during interviews (discussed by Oakley, 1981).

Small changes in wording sometimes produce major changes in the distribution of responses

This problem is nicely illustrated by results from Butler and Kitzinger’s study of the response distributions for different questions that were formulated by National Opinion Polls to gauge the British people’s attitudes toward entering the European Common Market in 1975 (Butler and Kitzinger, 1976:60). On the one hand, the difference between the percentages of ‘pro’ and ‘anti’ responses for the question: ‘Do you accept the government’s recommendation that the United Kingdom should come out of the Common Market?’, was 0.2 per cent in favour of the ‘pro’ Market position. On the other hand, for the question: ‘Do you accept the government’s recommendation that the United Kingdom should stay in the Common Market?’ the difference between the percentages of ‘pro’ and ‘anti’ responses was 18.2 per cent in favour of the ‘pro’ Market position.

A second example that demonstrates the impact that supposedly innocent changes in question wording can have is provided by Bishop et al. (1978), who conclude that apparent ‘trends’ in attitudes over the last thirty years in the United States are likely to be due to changes in the format of the questions that have been used in surveys. Before 1964, the Michigan Survey Research Center asked respondents if they had an opinion about an issue before asking them to rate the issues on a 7-point scale (‘No opinion — Agree strongly, Agree but not very strongly, Not sure . . . it depends, Disagree but not very strongly, Disagree strongly, — Don’t know’). Between 1964 and 1972, a dichotomised format was used: ‘Some people feel X while others feel Y; have you been interested enough to side with X or Y?’ After 1972, respondents were told that, ‘Some people feel X while others think Y and of course some people have opinions somewhere between’, before being asked where they would place themselves on a 7-point numeric scale (with alternative X being paired with 1 and alternative Y being paired with 7). The post-1972 scale does not include either a ‘No opinion’ or a ‘Don’t know’ response option. Whereas the format used prior to 1964 eliminated one-third of the respondents and the format used from 1964 to 1972 eliminated less that one-sixth of the respondents, the format used after 1972 hardly eliminated any respondents. In other words, after 1972, all respondents were forced to give answers to the questions whether or not they had preformed opinions to give, or had been interested in the topic in the past. Bishop et al. suggest that differences between the filtering power of each of the formats used (i.e. their ability to eliminate respondents for whom the topic is not relevant) are great enough to account for the apparent changes in public opinion in the United States since the fifties.

A third example that illustrates the effects of what, on the surface, appear to be harmless differences in wording comes from an Australian newspaper (the Melbourne Age 6/7/85:5). It was reported that when respondents in a national survey had been asked to rate the performance of the two main federal parliamentary parties on a scale of ‘Very good’ to ‘Very bad’, 39 per cent thought the prime minister was doing a good job and 27 per cent thought the same of the leader of the opposition. It was also noted in the same report, however, that two other national polls which had been conducted about the same time had found that 47 per cent approved of the prime minister while 48 per cent approved of his opponent; and 56 per cent saw the incumbent as making the better prime minister, compared with 27 per cent thinking the same for the leader of the opposition. Since the differences between the percentages are too large to be ascribed to sampling differences, it has to be concluded that ‘doing a good job’,
being 'approved of' and being seen as 'making the better prime minister' must be quite different matters — even though there would have been little reason to suspect this before the polls were conducted.

Last, lest it be thought that only opinion or attitude questions are susceptible to the effects of small differences in wording, it is worth noting that even questions about supposedly simple matters of fact can be vulnerable. Peterson (1984) has reported the results of a study that was designed to compare the non-response rates associated with different ways of asking respondents for their age. Whereas only 3.2 per cent of a random sample of registered voters failed to answer the question 'What is your age?', the question 'How old are you?', put to a similar sample, produced a 9.7 per cent non-response rate. Again, a researcher would have had no reason to anticipate this outcome before the experiment was carried out.

**Respondents commonly misinterpret questions**

A number of writers have discussed this problem at length (e.g. Cantril and Fried, 1944; Nuckols, 1953; Cannell, 1977; Belson, 1981, 1986; Hunt *et al.*, 1982). Nuckols (1953) discusses the results of an experiment in which nine questions that had been used by a national polling organisation were re-presented to a sample of respondents who were asked to repeat in their own words their interpretations of them. In all, 17 per cent of the interpretations given were judged to be either partially or wholly incorrect. Belson (1981) investigated respondents' interpretation of a number of common words including 'usually', 'generally', 'people', 'children' and 'weekday', and concludes that even these words elicit a wide range of different interpretations. Cannell (1977:44) presents similar data. It would appear that variation in the way respondents interpret everyday words is a common feature of questions used in social research.

**Answers to earlier questions can affect respondents' answers to later questions**

Available evidence suggests that 'contrast' and 'consistency' effects can be generated by prior questions. Rugg and Cantril (1944:28) found that the proportions of respondents agreeing with the idea that Americans should be allowed to enlist in: (a) the French army and (b) the German army, were affected by the order in which the questions concerning the two armies were presented. And Noelle-Neumann (1970) found that, when German respondents were asked to rate various foods in terms of how 'German' they were, potatoes were seen as being particularly 'German' by more respondents if that item followed rather than preceded rice.

No doubt it is because of the perceived likelihood of the occurrence of order effects that most methodologists advocate asking general questions about a topic before going on to more specific questions about it. Certainly, answers to prior specific questions often seem to influence answers to later, more general questions, although the nature of this influence is not always the same. Sometimes a consistency effect seems to operate and sometimes a redundancy effect seems to operate. For example, a consistency effect occurs when respondents are asked to indicate how often they pray before being asked to indicate how religious they think they are — that is, respondents who have just reported that they pray a lot are more likely to say that they are very religious (McFarland, 1981). A redundancy effect appears to occur, however, when respondents are asked about how happy they are with their work before being asked about how happy they are in general — that is, respondents seem to exclude consideration of work when answering the second question (see Strack and Martin, 1987; McClendon and O'Brien, 1988; note: we will return to these issues on pages 61–66 of chapter 5).

**Changing the order in which response options are presented sometimes affects respondents' answers**

The order in which response options are presented sometimes affects the probabilities of respondents selecting particular options. More specifically, it appears that respondents are more likely to endorse the options that they see first when they are able to read the items for themselves, and more likely to endorse items that they hear last when the items are read out aloud to them (e.g. Krosnick and Alwin, 1987). Notwithstanding these findings, it also appears to be the case that the first response option in a list has a greater impact on evaluative judgements than the last. Thus respondents who are asked to indicate which income category they fall into tend, if the categories are arranged in ascending order, to endorse lower categories than they would if they were arranged in descending order (Locander and Burton, 1976).

**Respondents' answers are sometimes affected by the question format per se**

Open ended questions (i.e. questions that allow respondents to supply their own answers) often produce quite different results from closed ended versions of the same questions (i.e. questions that force respondents to select a response
from a pre-set list of alternatives). Two issues seem to differentiate the two formats. In the first place, respondents are more likely to endorse a particular option if it has been explicitly listed than if they have to spontaneously think of it for themselves. In the second place, it has been found that respondents often give very different types of answers to open questions than they do to congruent closed questions.

The following two examples demonstrate the fact that respondents are more likely to endorse an answer if it has been explicitly listed for them than if it has not. The first is provided by Belson and Duncan (1962:160), who found that a checklist question yielded higher periodical and newspaper readership rates than an open question — for one periodical (the Radio Times) the percentage of respondents who checked it on the checklist was over five times the percentage who mentioned it when answering the open version of the question (38%: 7%). The second example comes from Schuman and Presser (1981:36). They report that, whereas 22 per cent of a sample of respondents who had been presented with an open question about what they thought was the most important problem facing the country mentioned the energy shortage, less than 1 per cent of a sample presented with a comparable closed question, which did not specifically include this issue as an option, mentioned it.

That respondents often give quite different kinds of answers to equivalent questions is illustrated by results reported by Schuman and Presser (1981). Some of the respondents, who gave the answer 'the pay' to an open question about what they would most prefer in a job, were subsequently found to have meant 'high pay' while others had meant 'steady pay'.

**Respondents often answer questions even when it appears that they know very little about the topic**

Respondents frequently answer questions that appear to be marginally relevant to them or about which they have thought little.

It has been found, for instance, that up to 25 per cent of respondents will check substantive options when a 'Don't know' is not offered but check a 'Don't know' option when it is offered (Schuman and Presser, 1981:186). And Ferber (1956) notes that in a random sample of 600 residents of Champaign Urbana, between 14 and 62 per cent of the 50 per cent of respondents who did not know about particular items in an array of topics, still volunteered opinions about them. Similarly, Gallup (1978:1176, cited by Smith, 1984a:221) found that, while 96 per cent of a national sample offered opinions about the importance of a balanced Federal budget, 25 per cent did not know whether the budget was balanced or not, 8 per cent wrongly thought that the budget was balanced, 40 per cent thought that it was unbalanced but did not know how much, and 25 per cent thought that it was not balanced but either underestimated or overestimated the amount it was out of balance by more than 15 per cent. In all, Gallup estimates that a mere 3 per cent of the respondents offered an answer that was based on accurate information. Other evidence that relates to respondents' willingness to offer answers that are not well grounded in knowledge comes from studies in which respondents have been asked about either fictitious or extremely obscure topics. Typically, these studies have disclosed that up to 30 per cent of the respondents have been prepared to answer the questions as if they dealt with topics that were real and familiar to them. (See Smith (1984a:223) for a review of the literature.)

**The cultural context in which a question is presented often has an impact on the way respondents interpret and answer questions**

An example that illustrates the importance of the context in which a question is asked is provided by Briggs (1986), who discusses an ethnic survey that was designed to assess the way a neighbourhood facility could best meet the demands of the residents in the area. He notes that it almost failed because a key question was interpreted quite differently by Navajo respondents, as compared to Zuni, Mexican-American and Anglo-American respondents. The level of demand for services indicated by the Navajo was much lower than that for each of the other groups. While this could have been interpreted as meaning that the Navajo were much less interested than the others in using the services, Briggs realised that the Navajo had not answered the question in the same manner as the other ethnic groups because of their cultural beliefs. Apparently the Navajo deem it highly inappropriate to speculate about the beliefs held by others. In their culture, such talk is seen as a usurpation of others' decision-making powers.

Speculating on the preferences of one's spouse and children would accordingly be deemed extremely rude. Rather than doing so, Navajo respondents would estimate which services they themselves would be likely to use. The use of a probe to obtain data on the family members generally yielded statements such as 'no, I don't think so...'. (Briggs, 1986:97)

**Critiques of past practices**

Perhaps because of the sorts of problems discussed above, several trenchant critiques of the use of questionnaires and interviews in social research have been published over the last two decades (see, e.g., Deutscher, 1966, 1973; Phillips,
interact with the cognitions and motivations of the interviewers and respondents, or to the properties of the relationship between interviewers and respondents. We must come to grips with the idea that all of these elements somehow constitute a dynamic, interrelated set of elements.

Summary

The use of verbal data has been made the keystone of contemporary social science and there is no sign of this situation changing. There is much evidence, nevertheless, to support the conclusion that the verbal data we collect are very often of dubious validity and reliability. It is also clear that we have lacked a proper understanding of the causes of these shortcomings. This is not to say that other kinds of data used in the social sciences suffer from less severe validity and reliability problems. Rather, attention has been directed at the problems associated with verbal data simply because the use of verbal data plays such a major part in contemporary social research.

Up to now the most influential attempts to provide a basis for improving interview or question-answer methodology have been based on efforts to summarise a huge number of diverse, ad hoc research findings (see, e.g., Sudman and Bradburn, 1974; Van der Zouwen and Dijkstra, 1982). Efforts to provide a more sophisticated basis for the formulation of question wording have been hampered by the lack of an encompassing theoretical framework, although a number of methodologists have made suggestions about the possible directions that might be taken. Phillips (1971), for instance, has stressed the social interactional nature of question-answer situations. Cicourel (1982) has stressed the relevance of cognitive and linguistic processes and the idea that question-answer behaviour should be treated as a form of communication. And Douglas (1985) and Briggs (1986) have stressed the impact of contextual variables on the way respondents interpret questions. The suggestions made by these writers should be taken on board, if for no other reason than the fact that no great improvements have been made over the last fifty years to our techniques for collecting verbal data.
assumptions and purposes to help them interpret the researcher's acts and
decide what information they should give. It was further suggested that, when
each respondent is allowed to do this in his or her own way, each will, in essence,
be answering a different question. Finally, it was noted that the problems
associated with the respondents' freedom to arrive at different interpretations
and then give different answers are exacerbated by the fact that topics are multi-
dimensional, the fact that observations can be made in individual or collective
terms, the fact that there are many different kinds of explanations, and the fact
that evaluations can be made in terms of different standards of comparison.

Subsequent chapters will take up the task of assessing, within the symbolic
interactionist framework, each of the assumptions that has been identified as
underlying survey research, so that the conditions which must be satisfied
before each can be seen to be valid can be specified. The overall aim is to
understand the conditions that must be met if we are to increase our confidence
that the questions we formulate for use in questionnaires and interviews will
work as we intend them to work.

Chapter 3

DEFINING TOPICS PROPERLY

This chapter focuses upon the first three assumptions that were identified as
underpinning survey research at the start of the last chapter. They were: that
the researcher has clearly defined the required information; that the respond-
dents have the required information; and that the respondents can access the
required information under the conditions of the research situation. All three
relate to the need for questions to be directed at properly defined, accessible
information. And all three relate to the first step in a question-answer cycle
(figure 3.1, page 26).

The researcher has clearly defined the required
information

As with many assumptions that appear at first sight to be straightforward, this
assumption is more complex than it first appears to be. To begin with, it implies
that the researcher must begin with a clear definition of the topic to be inves-
tigated. It also implies that the researcher has a clear understanding of the kind
of information about the topic that will satisfy the theoretical or practical
reasons for carrying out the research. It is worth looking more closely at each of
these implications in turn.
The researcher has clearly defined the topic

Some years ago the author acted as a methodological consultant for a community self-survey project. A small town had received a grant to carry out a survey of the residents' views about how their town should be developed. A committee of community leaders had been formed to plan the project and develop a questionnaire. One of the committee members wanted to include the question: 'Do you think . . . [our town] — has a drug problem?' Mindful of what the media might make of the finding that a number of the residents of a well known tourists' centre thought that their town had a drug problem, the author advised against the inclusion of the question. The committee member, however, was adamant that it should be included. When asked what he thought they would find out if the question were included, he said that it would be valuable to find out what the residents thought should be done for student transients who camped on the shores of the lake beside the town each summer. Should the town build toilet facilities around the lake? Should a hostel be built in the town? It is not hard to see that the question that had been suggested would not elicit answers to these questions.

Figure 3.1 The focus of chapter 3

<table>
<thead>
<tr>
<th>I</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher/Interviewer</td>
<td>Respondent</td>
</tr>
<tr>
<td>Encodes question.</td>
<td>Decodes question.</td>
</tr>
<tr>
<td>The need to clearly specify the information that is required.</td>
<td>The assumption that respondents have the required information. The assumption that respondents can access this information.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher/Interviewer</td>
<td>Respondent</td>
</tr>
<tr>
<td>Decodes answer.</td>
<td>Encodes answer.</td>
</tr>
</tbody>
</table>

But lay people are not the only ones who run into problems when formulating questions because they have not defined the topic properly. It is not difficult to find examples in the professional research literature that suffer from the same shortcoming. For instance, almost every social researcher regularly collects data concerning respondents' occupations even though occupational status is a notoriously difficult topic to handle for reasons that the following discussion highlights.

The question sequence used by the National Opinion Research Center and the Michigan Survey Research Center to collect information about respondents' occupations (see Sudman and Bradburn, 1982:189-194) begins with a closed question designed to establish respondents' levels of education. Respondents are then asked whether or not they are presently employed, unemployed, retired, a student, a housewife, or what? Respondents who are in employment are asked to classify themselves as working full-time or part-time, and those who indicate that they are working part-time may be asked how many hours per week they actually work. Respondents who say that they are unemployed are asked whether or not they are looking for paid work. People who work 15+ hours as unpaid workers in a family business or a family farm are regarded as 'working'; volunteer charity workers are classified as 'not working'.

Employed workers are then asked to provide a job description of their present occupation:

a: What kind of work do you do? ........................................
What is your main occupation called? ..............................

b: Tell me a little about what you actually do in that job? ....
What are some of your main duties? ................................

(c) What kind of business or industry is that in? ................
What do they do or make at the place where you work? ........
[If respondents have changed jobs, they are asked to focus on the jobs that they have done for the longest time (i.e. their 'usual' job)]

Finally, respondents are asked:

d: Are you (were you) an hourly wage worker, salaried, on commission, self employed, or what? [Farmers who cannot be considered hourly wage workers are considered 'self-employed'.]

Despite the amount of information collected and the level of detail involved, Sudman and Bradburn (1982:193) still make the comment that respondents' answers about their occupations are among the most difficult to code. And they add that, if this information is needed, the responses that are collected should contain sufficient detail to allow them to be coded properly. They then cite examples and instructions that the Survey Research Center gives its interviewers to try to
ensure that they manage to get the required details — for example, interviewers are told:

— The name of the place the respondent works at is insufficient (e.g. 'Bank' — respondent may be a manager, teller or janitor).
— And vague job titles (e.g. engineer, teacher ...) can cover a variety of occupations.

The sequence of questions that has just been presented was designed to make sure that respondents give an unambiguous description of what they do. Does it achieve this?

Occupations, like any other phenomenon, are multidimensional. They can be full-time/part-time, paid/unpaid, permanent/temporary, casual/contract/seasonal, legal/illegal, in the primary/secondary/tertiary sectors, government/private enterprise and so on. They can isolate incumbents from the products of their labour, involve being supervised by others or having to supervise others, involve being responsible for one's own work schedules or not, be skilled/semiskilled, be considered 'clean' or 'dirty', involve day-time or night-time shifts, allow one to spend a little or a lot of time with one's family, and so on. The work can be voluntary or forced. Respondents can have more than one job at the same time, which gives rise to the concept of 'usual' or 'main' jobs versus 'fill-in jobs', 'second jobs' and 'moonlighting jobs'. Respondents can nominally have jobs but not be working because they are on sick leave, maternity leave, holiday or strike. And respondents can be paid by piece-work, commission, wage rate, salary or honorarium.

The researcher must decide which of the many different dimensions should be focused upon. But this cannot be done without taking into account the reason for wanting the information in the first place. If, for instance, the researcher wants to estimate the portion of the labour force that is involved in the provision of social services, the dimensions focused upon might include: the number of hours worked, whether the work is permanent or temporary, paid or unpaid, private or public; and whether or not respondents are nominally employed, even if they are not actually working at the moment. If, on the other hand, the researcher wants to code respondents' answers according to level of occupational prestige, the emphasis will be placed on such dimensions as: the legality of the occupation; the required level of education and training; whether it involves clean or dirty work; whether respondents have control over their own work schedules; whether or not they are supervised; how they are paid; whether the work is temporary/permanent, or casual/seasonal/contract; and whether respondents are employed or self-employed.

Here is another example. Imagine that a soft drink manufacturer wants to find out whether or not respondents think that they would be likely to buy a drink based on a new recipe that the company has developed. But what is meant by the word 'buy'?

Consider an innocent appearing question such as, 'what brand of soft drink do you usually buy?' If the question is taken seriously, the respondent must first make a mental decision on the time period involved. A second decision is then necessary on which conditions to include. Are purchases at work, in restaurants, at sporting events, and at movies to be included; or are only purchases for home use to be counted? The respondent must decide on the meaning of the word 'you'. Does it refer only to the respondent or to the household of which the respondent is a member? How are purchases by one household member for other household members to be treated? A final question to be resolved mentally is the definition of a soft drink. Are lemonade, iced tea, fruit punch, and mineral water to be included or not? (Sudman and Bradburn, 1982:59)

The occupation and soft drink examples that have just been discussed serve to underscore the fact that all topics are multidimensional. That is, it is possible to think of any topic either in global terms or in terms of a specific number of dimensions. This problem bears on question-answer behaviour in at least three ways:

(a) It has to be appreciated that respondents' answers in terms of specific dimensions will not always be congruent with their global judgements.

Roslow et al. (1940) cite results from two surveys that demonstrate this fact. In the first study, respondents were asked: 'On the whole, do you approve or disapprove of Roosevelt's international policy?' (August 1939). In the second, respondents were asked: 'Do you approve of Roosevelt's policies with regard to the European situation up to now? (September 1939)?' Whereas 69.2 per cent approved of Roosevelt's policies in regard to the European situation, only 40.5 per cent approved when asked to react globally to Roosevelt's 'international policy'.

Another example that illustrates the same problem is provided by Cantril and Fried (1944:9–10) who report that whereas only four of forty respondents answered 'no' to the question: 'Are you in favour of Labour Unions?', the level of support as indicated by answers to six specific questions was more problematic. For instance, eighteen of the forty respondents believed that the government should have more control over the unions and twenty-five felt that the unions should be prohibited if they got into 'radical' hands.

A moment's reflection upon the examples that we have just looked at forces one to the view that the researcher must not only decide how respondents should orient to the topic — that is, whether they should orient to it in global terms or in terms of a number of specified dimensions — but also instruct respondents accordingly so that they will all orient to the topic in the same way. If the researcher does not do these things, different respondents can either answer in terms of different dimensions from one another, or answer in terms of different dimensions to those
which the researcher has in mind. If either of these things happens, the researcher is likely to misunderstand the replies given by respondents and make invalid comparisons between the replies given by different respondents. The following anecdote indicates the sorts of things that can go wrong:

The visiting teacher from the country town was giving her first demonstration to a class of nine-year-olds in an upland farm area... She had chosen as her subject: 'Wool!', and started off by showing a large coloured picture of a sheep and remarking brightly: 'Now I am sure you all know what this is'. Much to her surprise, there was no response to her implied question. When she put it more pointedly, one sturdy boy asked if she could see the picture nearer at hand. On permission being given with much wonderment, he surveyed it carefully and hazarded: 'It's a two-year-old Border-Leicester, isn't it?' (Laing, 1957:24)

(b) It should also be appreciated that when respondents are required to answer in terms of specific dimensions, the dimensions that they are to focus on must be properly defined.

The next example indicates the kind of difficulties that arise when this is not done. In a study of the social changes and reforms that residents of the United States wanted to see once the Second World War had ended, Crutchfield and Gordon (1947) found that 41 per cent of the males and 54 per cent of the females who were interviewed wanted things to remain the same. Further analysis of the respondents' answers, however, revealed that while 80 per cent of the males had domestic changes or reforms in mind, only 53 per cent of the females had thought in terms of this dimension. Moreover, when the answers for the men and women who had domestic changes in mind were analysed separately, the percentages for the men and women who said that they would like things to remain the same were practically identical (60% and 59% respectively). It should also be noted that although most of respondents in the total sample interpreted the question in terms of the domestic changes or reforms dimension, a sizeable minority (43%) had a variety of other dimensions in mind — for example, social, economic, political or technical changes.

c) The fact that any topic can be defined in terms of different dimensions has the important corollary: how a topic is defined can dramatically affect the way the responses are distributed.

An example of the effect of increasing the specificity is seen in questions from a Gallup Poll of May-June 1945:

Do you think the government should give money to workers who are unemployed for a length of time until they can find another job? (yes: 65 per cent; no: 32 per cent; don’t know: 5 per cent).

It has been proposed that unemployed workers with dependants be given up to $25 per week by the government for as many as twenty-six weeks during one year while they are out of work and looking for a job. Do you favour or oppose this plan? (yes: 46 per cent; no: 42 per cent; don’t know: 12 per cent).

Would you be willing to pay higher taxes to give unemployed persons up to $25 a week for twenty-six weeks if they fail to find satisfactory jobs? (yes: 34 per cent; no: 54 per cent; don’t know: 12 per cent). (Bradburn, 1982:70).

Note that: Link, 1943:267; Cantril and Fried, 1944:9–10; Payne, 1951:196–197; and Smith, 1984a discuss similar examples.

To recapitulate, researchers must decide whether global or specific responses are required, and then clearly indicate this to respondents. If specific dimensions are to be focused upon, they must be properly defined for respondents.

There may be times when a researcher wants to measure global responses: for example, questions about overall party preferences may be used as a general indicator of how respondents will vote in an election. All the same, it must be appreciated that, when respondents are left to define a topic in global terms for themselves, the researcher cannot be sure that all will do it in the same way. In addition, if respondents are free to select their own dimensions of focus, there is nothing to stop them from changing these over time as different dimensions occur to them. If global responses are required, the dimensions that respondents should focus upon should be clearly specified. Freed is undoubtedly correct in his view that:

A common error is for a question to be so generally structured that neither the writer nor the reader understands what is being sought. This 'general fallacy' also invites a variety of interpretations so that the respondent must decide for himself what the researcher intended; thus, the researcher is often thwarted in his endeavour because the respondent misunderstood the intent of the question. General and vague questions are usually indicative of a lack of clarity on the part of the writer, bewilder the reader, and produce unreliable results. Don’t let questions that are too general be a subterfuge for inadequate preparation. A clear and pointed question elicits an effective and relevant response. (Freed, 1964:188)

More generally, it is difficult to reject the view that, unless all respondents focus upon the same topic and respond to it in terms of the same dimensions, the answers that they give cannot be meaningfully compared with one another.

The researcher has a clear idea of the kind of information that is required about the topic

Besides demonstrating the sort of problem that arises when the researcher fails to define the topic properly, the example of the local businessman wanting to ask residents whether or not they thought their town had a drug problem
(alluded to at the start of the last section) illustrates what can happen when the researcher does not have a clear idea of the information that is required about the topic. Although it is obviously the case that defining the topic properly is a major step toward having different respondents give the same kind of answers, it should not be forgotten that the information collected should satisfy the reason for conducting the research in the first place. Three rules advocated by Sudman and Bradburn (1982:13) are pertinent to this issue. They are:

(a) Do not formulate specific questions until you have thought through the research question.

(b) Write the research question down and keep it handy while formulating specific questions.

(c) Keep asking, 'Why do I want to know this?' ('It would be interesting to know' is not an acceptable answer.)

Adherence to these rules will go a long way to ensuring that the questions which are asked are relevant to the researcher's goals. Yet adherence to them is not enough to guarantee that the questions asked will cover all relevant aspects of the topic, so it is worth adding a fourth rule, which is advocated by both Kidder (1981) and Deming (1944):

An excellent test ... is the preparation of 'dummy tables' showing the relationships that are anticipated. By drawing up such tables in advance, the investigators force themselves to be definite about what data are required and how they will be used. They can even enter figures representing different possible findings in order to visualize the bearing each would have on alternative hypotheses and to see what new hypotheses they suggest. (Kidder, 1981:60)

Without some pretty good idea of what the analysis is going to show it will be impossible to design the questionnaire so that any useful proportion of [its] ... aims will be accomplished. An understanding of the subject is accordingly demanded as one of the qualifications for planning a survey. Likewise the tabulation program (which really should precede the questionnaire but usually does not) demands enough knowledge of the problem to see what tables are going to be needed and which ones will be significant. (Deming, 1944:364)

When researchers are not clear in their minds about the nature of the information they require, they cannot hope to ask questions that are relevant to their topics of interest. While this basic requirement might appear to be so obvious that it scarcely needs to be stressed, it is all too easily forgotten. Researchers are all too prone either to asking questions without bothering to clarify the meaning of key concepts or to letting initial formulations of the questions cause them to lose sight of their original goals. As Payne puts it:

Many of the problems of wording result from our going off half cocked. Once the issue is posed in even the vaguest terms, we start trying to put it in words that are understandable to the public ...

If we did but realize it, the first half of the battle consists of putting the issue in a form that we can understand ourselves. We need to first and foremost define the issue precisely, regardless of the general understandability of the words. The news reporter has the stock questions to ask himself about each issue: Who? Why? When? Where? How? ... We can well ask the same questions of ourselves for each issue we intend to pose ... (Payne, 1951:26) [my emphasis]

Lamentably, although clearly sensible, the advice offered by Sudman and Bradburn, Payne, Kidder and Deming, is seldom followed by the experts. And it is difficult to make novice researchers follow it. The author, for example, has found that it always requires concerted pressure to motivate students to systematically write the avowed purpose beside each question in the first drafts of questionnaires or interview schedules (i.e. to write beside each question what information they think it will elicit and what they will do with that information). It is as if the reason for wanting to ask a question is so obvious before it is explicated that researchers see little merit in writing it down. Yet, typically, when these working procedures are adopted, the adequacy of most questions becomes far less certain. Explication of the purpose of each question and how resulting data will be used has an almost fiendish ability to force implicit assumptions out into the open and highlight any mismatches between the researcher's actual interests and the proposed questions.

Respondents have the required information

An assumption survey researchers commonly hold is that the respondents have the information they want:

[The] structure and administration of the interview reflects an expectation of 'opinionatedness' characteristic of only a few levels of the social structure. (Manning, 1966:7:306)

The typical survey question incorporates assumptions not only about the nature of what is to be measured, but also about its very existence. (Converse and Presser, 1986:35)

Most methodologists would agree that there is very little point in asking respondents hypothetical questions. In Converse and Presser's words:

If we ask a hypothetical question, will we get a hypothetical answer — as some lighthearted critics have warned? Perhaps not, but the counsel of experience and research suggests that asking most people to imagine what if — what might have happened in their lives if things had been otherwise, or what they might do if — confronts them with a special task that is likely to be difficult ... Hypothetical questions provide no guarantee that respondents will feel the full force of political or economic realities ... (Converse and Presser, 1986:23)
Converse and Presser go on to suggest that if hypothetical questions are asked, at least one question pertaining to the respondents’ actual experience should be included.

In spite of the general suspicion regarding the worth of hypothetical questions, questions are commonly presented to respondents without first checking their relevance to respondents (see Converse and Schuman, 1984, for an analysis of the types of questions used by polling organisations). This must be coupled with the unfortunate fact that respondents’ willingness to answer questions cannot be taken as a sign of their relevance. Smith (1984a:215) reports, for example, that studies of the American electorate during the 1950s developed a picture that was startlingly at odds with the assumption of a rational citizenry. In extreme cases, it was found that up to 80 per cent of respondents answered attitude questions even though they had no real position on an issue. They appeared to randomly choose responses in order to come up with answers to the questions. In a similar vein, Belson (1981:371), after an intensive analysis of respondents’ interpretations of a series of market survey questions, concludes that when respondents find it difficult to answer a question they are likely to modify it so that they can answer it more easily. They might eliminate difficult elements, overlook qualifying phrases and clauses, limit the question to things that they know about, and interpret the scope of the question more or less broadly than the researcher intended.

Such findings have led many survey methodologists (e.g. Smith, 1984a:245; Converse and Presser, 1986:55–58) to argue for the use of filter questions that gauge the level of respondents’ knowledge about the topic, its salience to them, and the intensity of any opinions or beliefs that they may hold about it.

Cicourel (1982) has also suggested that it should not be simply assumed that respondents will hold opinions, attitudes and beliefs. He advances the view that ignorance should be treated as a variable:

Why not assume that the actor’s thoughts about social objects are loosely structured but are perceived as concrete until we begin to probe them with specific questions that put him on the spot about matters which he takes for granted and to which he seldom gives much time? Survey research procedures do not assign variable status to ignorance, much less acknowledge it as a critical factor in the structure of social action... (Cicourel, 1964:115)

Although the views put forward by methodologists (such as Smith, Converse and Presser, and Cicourel) touch on obviously important matters, it is interesting — if depressing — that similar views were expressed more than four decades ago. In the late forties, Riezman and Glazer (1948/49) proposed that an alternative model to the conventional survey researcher’s model might divide a population into:

(a) Those who have grounds for their opinions and those who do not (see e.g. Katz, 1945/6, who criticises a study of Whites’ perceptions about the way Blacks were being treated on the grounds that respondents were not asked whether they knew any Blacks).

(b) Those who take their opinions seriously and feel that they are of consequence and those who do not. ‘For most people in modern society, there is no... direct relation between responsibility for having an opinion and responsibility for action.’ (Riezman and Glazer, 1948/49:635)

Further, it would seem reasonable to assume that those respondents with most responsibility for formulating decisions and planning courses of action will be most motivated to gather information and spend time thinking about the relevant issues. And it would seem reasonable to predict that such respondents will be most likely to have associated information, beliefs and opinions.

Notwithstanding the apparent reasonableness of the suggestions that have been made over the years, it has to be admitted that they have yet to have much impact on general practice.

Still relevant to the issue of whether or not respondents have the information required by the researcher, it must be kept in mind that, even if respondents have the information at one point in time, it is possible that they will forget it with the passing of time. It is also possible that individuals will not always be aware, or conscious, of stimuli that impinge upon them and influence their behaviour. These considerations simply mean that, even when respondents have in the past been exposed to experiences which are relevant to the topic being investigated, it cannot be taken for granted that they will have retained any information about them. (Note: these issues are dealt with more fully in chapter 7.)

The conclusion that one is forced to accept, then, is that it should not just be assumed that respondents have the information that is sought. In the case of each respondent, the relevance of a particular question must be independently established. Ways in which this can be done are discussed in chapter 8.

**Respondents are capable of verbalising the information the researcher wants under the conditions of the research situation**

There are a number of factors that relate to the assumption that respondents can access the information they do have. Training respondents in research techniques is enough to convince anyone that the single most serious failing of novice interviewers is the tendency to go too fast — not to give respondents time to answer fully. In everyday situations, interactants typically ‘uhm’ and ‘ahh’, scratch their heads, look into the distance, and so on. In other words, they take their time to give answers to questions. Standardised interviews, however, pressure respondents to work at the interviewer’s pace and interviewers often proceed too quickly (see e.g. Cannell et al., 1981).
Fowler and Mangione (1990) add to this problem another twist that stems from the fact that interactants constantly interpret each other’s behaviour.

Fowler (1966) found that the pace at which the interviewer conducts the interview is a significant correlate of the respondent’s perception of the task. If interviewers go too fast, respondents conclude that accurate and complete answers are less important than quick answers. To our knowledge, a direct link between pace and data quality has not been documented. However, we think it is likely that having interviewers speak slowly is one important and practical way in which researchers can increase the standardization across interviews . . . (Fowler and Mangione, 1990:71) (my emphasis)

In addition to failures associated with interviewer speed, questions can fail because they fail to stimulate respondent recall. There is some evidence that questions must link up with cues associated with material in the respondent’s memory if that material is to be successfully recalled (see Thompson and Tulving, 1970; Tulving, 1974; Martin, 1986). They can fail, too, because question tasks may be too difficult for respondents — especially when they are pressured by interviewers who are impatiently waiting for answers. A question may be so complex that respondents have trouble comprehending the individual words let alone the overall sense that is intended. And questions can be so long that respondents are not able to concentrate long enough to absorb them.

More practically, if questions require respondents to remember information, they must be worded in a way that helps respondents accurately retrieve the information from their memories. They must be short enough to fit within each respondent’s attention span and simple enough to be fully comprehended. Techniques that can be used to achieve these goals are discussed in chapter 7.

**Summary**

A necessary precursor to a successful question-answer cycle is that both researcher and respondent have a shared understanding of the topic under investigation. Since all topics are multidimensional, respondents can orient to a topic in either a global or a more narrowly defined fashion. If different respondents define a topic in different ways and the researcher is not aware of this fact, the respondents’ answers will neither be properly interpreted nor meaningfully comparable with one another.

The researcher can take one of two tactics to meet this problem of different respondents giving different kinds of answers. The researcher can try to identify the dimension(s) each respondent has focused upon. Alternatively, the researcher can specify the dimension(s) upon which all respondents should focus. The first of these solutions runs into a number of difficulties, such as the fact that respondents may not be aware of the dimensions that influence their answers when they define the topic for themselves and the fact that their definitions of a topic are likely to vary over time. While the second solution runs into other problems, such as the danger of imposing inappropriate response categories upon the respondents, it does ensure both that the topic is clearly defined for each respondent and that each answer will be relevant to it.

To maximise the likelihood that questions are relevant to the topic as well as to the reason for wanting to collect information about it, a researcher should continually ask him- or herself: “Why do I need to know this?” and “What am I going to do with the answers respondents give?” In addition to indicating how the topic should be defined and what kind of information is required, the researcher must establish the basis upon which it can be assumed that respondents are likely to actually have the information that they are asked to provide. The researcher should try to avoid putting respondents in the position of having to entirely invent or fabricate answers on the basis of little information. And, last but not least, it is important that respondents be given enough time to formulate appropriate answers.
The last chapter dealt with the need for the researcher to begin with a clear definition of the topic and a clear idea of the information that is required about the topic. In this chapter, the focus will still be on the first step in the question-answer communication cycle, as we consider some of the problems the researcher has to confront when trying to formulate a request for the required information which will be understood by the respondent (see figure 4.1, p. 59).

In the past, social researchers have typically assumed that their questions would be understood by respondents as they are intended to be understood. But, just as it is often doubtful that researchers always have a clear idea of the information that should be collected, it is doubtful that respondents always understand what information the researcher wants.

If respondents are to understand a question as intended by the researcher, they must minimally give the same meaning to individual words and attribute the same overall meaning to the structural elements (i.e. component phrases, clauses, etc.) that make up the question. Both of these requirements involve a host of complexities that need to be discussed at length.

The meanings of individual words

Before respondents can interpret a question as the researcher wants, they must understand each word in the same way as the researcher understands it. Unfortunately, one cannot safely assume that the meanings of ordinary words — let alone exotic or uncommon words — are necessarily shared even by respondents who have been socialised in the same general culture as the researcher.

As men have known throughout the ages and as modern semantics has pointed out in detail, the meaning of even the simplest word can be slippery. When we add to the ordinary problem of verbal communication the additional problem of presenting a meaning to groups of people widely separated in background, experience, estimations and terminologies peculiar to interest or occupational groups, the difficulty confronting a public opinion investigator becomes clear. (Cantril and Fried, 1944:3)

The failure of respondents to give the same meaning to individual words is generally explained by reference to the following four factors: (a) the evolution of context-specific nuances of meaning; (b) their relative difficulty (defined either in terms of number of syllables or the frequency of occurrence in everyday life); (c) lack of clear empirical referents; and (d) the operation of related nuances of apparently similar words.

Figure 4.1 The focus of chapter 4

<table>
<thead>
<tr>
<th>I</th>
<th>Researcher/Interviewer</th>
<th>Encodes question. Formulating an intelligible request for information. Defining key terms for respondents and the virtues of simplicity and brevity.</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>Respondent</td>
<td>Decodes question.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV</th>
<th>Researcher/Interviewer</th>
<th>Decodes answer.</th>
</tr>
</thead>
<tbody>
<tr>
<td>III</td>
<td>Respondent</td>
<td>Encodes answer.</td>
</tr>
</tbody>
</table>
The evolution of context specific meanings

In chapter 2, attention was drawn to the symbolic interactionist hypothesis that human beings constantly negotiate the meanings of, and labels for, the social acts and events that affect them. This explains why acts and events, and the symbols that are used to represent them, can have different meanings at different times and in different places. For example, whereas the British use the word 'tea' to refer to an evening meal, Americans use this word to refer to a drink. Whereas Australians use the word 'backyard' to refer to the garden area behind a house, the British use it to refer to the small, concrete area that is typically found at the back of a low-cost house in a working class area — the area behind a middle-class home is called a 'garden'. Whereas younger generations use the words 'obscene' or 'gross' to mean disgusting or degrading, contemporary Australian teenagers use these words as marks of approval, especially after a joke has been told. Whereas people in Australia use the word 'outside' to refer to the outdoors, residents in the Canadian North West Territories use it to mean 'out of the territories' (i.e. 'down South'). And, whereas businessmen in Australia interpret 'yes' in answer to the question: 'Do we have a deal?' to mean that the deal has been finalised, Japanese businessmen take it to mean that they will go away and think about it.

Even when words seem to carry shared meanings, unexpected difficulties can emerge. For instance, although 'age' is a concept that one might expect to have a shared meaning among all members of any Western society, different age groups may have different ways of thinking about age. Most teenagers, for example, define anyone over forty as old, while most people in their sixties tend to reserve this category for people over eighty. And income seems to be categorised according to locally current norms, so that the meanings associated with different income levels vary from region to region. An income which makes a person wealthy in one community may not make a person wealthy in another.

The above examples serve to highlight a fundamental issue. Minimally, researchers must be confident that they know how respondents will interpret key words. Such confidence might be based either on the fact that they have defined the key concepts for the respondents or the fact that they have taken time to investigate how the respondents interpret the key concepts when left to do so for themselves.

It should not be assumed, however, that it is easy to avoid problems arising from a lack of shared meaning. Even when words are understood as intended, it is possible that unintended nuances can interfere with communication. Sorensen (1972, cited in DeLamater, 1982:24), for example, used the word 'ball' to refer to sexual intercourse in a questionnaire used in an American survey because he assumed it was the word that the respondents would use themselves and that its use would enhance rapport. Certainly, Sorensen's reasoning is in line with Cicourel's (1964:140) view that question wording should reflect the concepts used by the actors. Nevertheless, it has to be recognised that the use of slang runs the risk of transgressing the norms of politeness and of violating respondents' expectations regarding proper 'professional' behaviour. If this happens, using the actors' own concepts might have a negative rather than a positive effect on rapport, and might introduce nuances that escape the researcher — for example, that the researcher is unprofessional and that the research project is not worth the respondents' time.

The choice of which words to stress can also change the meaning of questions. Lazarsfeld (1944, reprinted 1972a) discusses the way in which the meaning of the question, 'Why did you buy that book?' is dependent upon the word that is stressed. Stressing the word 'why' conveys surprise or disapproval; stressing 'buy' conveys a desire for an explanation of that course of action rather than other courses of action (e.g. borrowing it from the library); stressing 'that' implies a particular book rather any book; and stressing 'book' conveys a request for an explanation (e.g. for spending money in that particular way rather than in some other way).

The relative difficulty of words

The relative difficulty of words affects respondents in a number of ways. Sudman and Bradburn (1974:50) suggest that increasing the difficulty of the words in a question increases the degree to which the question threatens respondents and consequently the tendency for respondents to give 'Don't know' responses. And Payne (1951:chapter 8) presents results which suggest that vocabulary difficulty (measured in terms of either the average number of syllables per word or average number of prefixes and suffixes per word) relates to the tendency to endorse the last option mentioned.

Words that may seem simple enough to middle-class, well-educated researchers may not be intelligible to many respondents. Belson (1981), for instance, asked respondents: 'Do you think T.V. news programs are impartial about politics?' In follow-up probes put to fifty-six respondents, twenty-five interpreted the word impartial as the researchers intended; ten overlooked the word altogether; nine interpreted it as 'tending to spend too much time on politics'; five gave an opposite meaning (e.g. unfair, biased); two interpreted it as 'giving too little time to information'; and the remaining seven had no idea what the word meant at all. Similar results were generated by the inclusion of the word proposition in a question. When potentially difficult words (e.g. words that are not commonly used or have a technical meaning) have to be included, it is clear that they should be defined for respondents.

Most survey researchers would subscribe to the rule that 'terms must be simple and comprehensible even to the least educated respondents' (Kidder and
Judd, 1986:245), even if this approach runs the risk of offending some respondents. Payne (1951:117) gives examples of how to do this without insulting the intelligence of better-educated respondents. He claims that it is better to ask: 'How do you feel about the amount you have to pay the government on the money you take in during the year — your income tax, that is?' than to phrase the question in the reverse order: 'How do you feel about your income tax — that is, the amount you have to pay the government on the money you take in during the year?' It is certainly possible that defining a term informally first might lessen any didactic overtones to a particular question formulation while allowing the interviewer two shots at communicating the intended meaning. It is also preferable because it minimises the likelihood that respondents will begin to answer the question before they have fully understood it.

The lack of empirical referents

When more than one plausible interpretation exists, the respondent needs to consider the various possibilities and often must think up and answer an internal questionnaire to help decide which interpretation to accept. Take, for example, this question: 'How many times have you talked to a doctor about your health during the past month?' The respondent may wonder whether to include telephone consultations, whether visits to chiropractors should be included, whether immunisations are part of health, or whether 'past month' refers to the past four weeks or to a calendar month. Whether or not the respondent goes through this process explicitly, he or she must proceed on the basis of assumptions concerning the intended meaning of the question. (Cannell et al., 1981:994)

Bradburn (1982:69) cites the results of an investigation of the interpretation of political symbols (Fee, 1981) which illustrates the point that is being made here. Fee found that the term 'Big Government' is associated with four distinct meanings:
- welfare, socialism and overspending
- government for big business and the wealthy
- federal control and diminished state rights
- bureaucracy and a lack of democratic process

More importantly, Fee found that each of these meanings correlated with quite different attitudes so that, without knowing which meaning respondents ascribe to the term 'Big Government', it is impossible to interpret with any confidence their responses to questions that incorporate this term.

Of course, general questions, by definition, are cast in abstract terms that have, at best, vague empirical referents. It is for this reason that Payne (1951:149) argues that words like 'art', 'business', 'government', and 'world affairs' should be avoided. He maintains that there is so little consensus over what they mean that we may just as well substitute the word 'blab' every time one of them occurs. The following is an example of what Payne calls 'blab-blab' questions: 'Should our country be more active in world affairs?' Such a question, he asserts, means little more than the question: 'Should our bab be more bab in bab?' He asks:

What is conveyed by the word 'country' in this instance — our government as a whole, the State Department, our commercial firms, our industrialists, or what? 'More active?' In what way? How active is our country now? 'World affairs?' Oh brother!

These are the kind of terms that have to be defined explicitly before philosophers can begin to discuss them. Let's not inflict them on the general public with no definitions whatever! That is, let's not do it unless we have some follow-up questions which enable us to realise the multitude of combinations or frames of reference such bab words bring up. (Payne, 1951:150)

There is little doubt, too, that Converse and Presser (1986:31) are correct in arguing that the more general the question, the wider the range of interpretation that will be given to it. To be more precise, words can lack clear meaning in two ways. First, a word may have several meanings associated with it. It is for this reason that Payne (1951) counsels the avoidance of words with multiple meanings — an issue that he says can be settled quickly by reference to a dictionary — on the grounds that the more meanings a word has, the more likely it is to be ambiguous in meaning. The fact that even simple words can have different meanings in different social situations, however, should be kept in mind. Different cultural or subgroup nuances are not likely to be listed in a dictionary. Secondly, a word may lack empirical referents. For example, Rugg and Cantril (1944:25) found that forty respondents gave a variety of interpretations to key words in a question taken from a 'Fortune Poll': 'After the war is over, do you think people will have to work harder, about the same, or not as hard as before?' Twenty-three respondents interpreted 'people' as people in all walks of life, while twelve interpreted 'people' as one class; ten interpreted 'would have to work harder' as produce higher quality work, while six interpreted it as 'work longer hours' and five as 'compete harder'.

Regrettably, there are many abstract words which provide potential pitfalls for the interviewer. Many are very common. The list includes frequency terms such as 'regularly', 'usually', 'often', 'frequently', 'seldom' and 'rarely', as well as evaluative words such as 'good', 'bad', 'approve', 'disapprove', 'agree', 'disagree', 'like', and 'dislike'. All of these words lack clear empirical referents. How many times a year is 'often'? And what is meant by 'usually'? Belson (1981:557) reports the results of investigating 445 interpretations of the word 'usually' where it had been meant to be interpreted as 'ordinarily', 'in the ordinary course of events', 'commonly', 'as a rule', 'generally', 'normally', 'habitually', 'as a routine', 'almost always', 'most', 'typically', and various other interpretations closely similar to these. Eighty-five of the respondents overlooked the word altogether and only 60 per cent of them interpreted it in one of the ways
listed. To make matters more complex, the ways in which many of these words are to be interpreted are dependent upon the topics with which they are paired. Thus, 'once a day' may be 'very often' when one is talking about haircuts but not 'very often' when one is talking about brushing one's teeth.

Generally speaking, abstract words generate problems because they are:

(a) more likely to be affected by contextual influences since respondents are forced to search for interpretative clues;
(b) more likely to produce non-comparable responses from different respondents since respondents can focus upon different dimensions of the concept in question; and
(c) more likely to allow respondents room to change the intended scope of the words in order to answer the questions.

Relevant to the last point, Belson (1981) hypothesises that there is a strong tendency for respondents to interpret broad concepts less broadly as they limit interpretations to their own experience (e.g. 'programs' will become 'adult programs').

The important lesson to be drawn from the foregoing discussion is that the words used by the researcher when formulating questions should be as specific and as concrete as possible. Concepts should be defined in empirical terms. Relevant to this point, Bradburn and Sudman (1979) recommend that frequency questions should be worded in terms of numeric referents per specified time period. In other words, the researcher should ask, 'How many times did you go out to see a movie during the last month?' (1, 2, 3, etc.) rather than asking the question, 'Would you say that you go out to the movies: rarely, once in a while, quite often, often, etc.'? When possible, concrete examples should be given (e.g. following an illustration given by Payne, 1951: 'leafy vegetable' could become 'leafy vegetables like lettuce and cabbage, etc.').

Starting questions with short vignettes giving a little word picture of the situation, so as to clarify the meaning of a key concept, is an extension of the same principle (see Alexander and Becker, 1978). Lemkau and Crocetti (1962), for example, used vignettes to define categories of mental illness for respondents. The following is one that they used:

Now I'd like to describe a certain kind of person and ask you a few questions about her. She is a young woman in her twenties — let's call her Betty Smith. She has never had a job, and she doesn't seem to want to look for one. She is a very quiet girl, she doesn't talk much to anyone — even her own family — and she acts as if she is afraid of people, especially young men her own age. She won't go out with anyone and whenever someone comes to visit the family, she stays in her own room until they leave. She just stays by herself and daydreams all the time and shows no interest in anything or anybody. (Lemkau and Crocetti, 1962:694)

Similarly, Mellinger et al. (1982) used vignettes to define ethical issues. The next vignette was used to define the ethical dilemma posed by the use of placebos in research.

An example of the use of placebos is this research on rheumatoid arthritis. This is a condition that does not kill people but it can be very painful. There's no cure for it but there's a new drug that researchers believe will relieve arthritis pain and keep the disease from getting worse.

In order to test the drug, researchers need to compare it over a long period of time with a placebo. They would give the new drug to one group of arthritis patients, and a placebo to another group. The test would last one year. If the new drug is as good as the researchers think, most of the people taking it would no longer be in pain, and their arthritis would not get worse. Most people taking the placebo would continue to have pain, and their arthritis would steadily get worse. The patients would not be told which they are getting — placebo or new drug.

In this study, then, the possible benefits of the new drug would be withheld from half the subjects for one year. But if the study is successful, a more effective drug would be available for all arthritis sufferers.

Do you think it is OK or not OK to do this study? (Mellinger et al., 1982:109)

The operation of unintended nuances associated with apparently similar words

Occasionally it becomes apparent that words which on the surface appear to be very similar produce very different responses when used in parallel forms of the same question. A classic example of this which has been discussed by a number of writers (e.g. Rugg, 1941; Schuman and Presser, 1981, Hippler and Schwarz, 1986) is a survey experiment conducted by The Roper Public Opinion Research Center in 1940 in which respondents were asked either: 'Do you think the U.S. should allow public speeches against democracy?' or 'Do you think the U.S. should forbid public speeches against democracy?' Whereas 62 per cent said 'no' to the first question, only 46 per cent said 'yes' to the second. Rugg (1941) comments that 'evidently the forbid phrasing makes the implied threat to civil liberties more apparent'. One could also argue that 'allow' and 'forbid' are not exact opposites: that to allow is to permit but not necessarily to encourage while to forbid is not only to discourage but to try to prevent. But how we assess the strength of negatives versus positives is also a factor here. Jordan (1965) reports that negative response options for rating scales are perceived as being stronger than the countervailing positive response options, which explains why respondents' negative responses (e.g. forbid) tend to be less extreme than their positive responses (e.g. allow).

A second, similar, example of how unexpected nuances can differentiate apparently similar words is provided by Murray et al. (1974, cited by Bradburn, 1982:71). They compared the response distributions for two versions of a question aimed at measuring respondents' acceptance of daylight saving. The first version was phrased in terms of whether or not they approved of daylight saving.
The second was phrased in terms of whether or not they liked daylight saving. A cross-tabulation of results indicated that although the two response distributions were highly correlated, it was possible for some people to like daylight saving and not approve of it and vice versa. Of those who liked year round daylight saving very much, 68 per cent approved of it. Of those who disliked it very much, 10 per cent approved of it!

A third illustration of the problems of defining nuance was mentioned in chapter 1 (pp. 5–6), where it was concluded that the terms 'doing a good job', 'being approved of' and 'making the better Prime Minister' appear to refer to quite different things.

The difficulty with all these examples is that the problem they embody is not obvious until after the fact. While much of the trouble is undoubtedly a consequence of the fact that terms such as 'approve', 'like', 'allow' and 'forbid' are ill defined, it is not clear how one could reliably predict either the positive or negative nuances that seem to be associated with them. The best course would seem to be to avoid using terms that have been found to have moral overtones. More generally, the best course would seem to be to avoid ill defined evaluative terms.

Again it seems appropriate to conclude that the problems associated with poorly defined words can be serious. Unfortunately, it must be acknowledged that the advice to use words that are concrete and specific (i.e. defined in empirical terms) is not enough to avoid all of the problems that underlie the construction of clearly interpretable questions. The sorts of problems that have just been looked at are only likely to be flushed out by the use of the in-depth question testing procedures that are discussed in chapter 12.

Structural complexities and the requirement that questions should be interpreted as intended

The overall meaning attributed to a question involves more than the sum of the meanings of the individual words. Interpretations of a question are influenced by its structural complexity. Structural features that have been found to create confusion in respondents include the sheer number of words, the number of grammatical parts and the use of negatives.

The sheer number of words used

Payne (1951) was perhaps the first to present data to support the hypothesis that the level of respondent confusion is related to the length of a question.

Payne concluded that respondents are more affected by the order of alternatives when questions are more than twenty words in length. More recently, Molenaar (1982:53), having reviewed the relevant literature, has concluded that short questions reduce the likelihood of a number of different response effects occurring. The response effects he considered were those which stem from grammatical complexities, task conflicts, and the tendency on the part of some respondents to give either socially desirable or acquiescing responses. Molenaar also concludes that: the greater the number of substantive or informative words used, the more likely it is that the question will be interpreted wrongly.

Grammatical complexities

Asking two or more questions at once

Most texts on question construction assert that questions should be as short and simple as possible. In practice this means that the researcher should avoid asking two or more questions at the same time, as well as avoid adding qualifying clauses, phrases and instructions.

Rather than asking: 'Did you see or hear anything?', it would be better to ask:

Q1 Did you see anything?
Q2 Did you hear anything?

Rather than asking: 'Can you give me some idea of the number, makes and sizes of each size that you have sold to date or have in stock at the moment?' (discussed by Belson, 1981:25), it would be better to ask:

Q1 How many of size X have you sold to date?
Q2 How many of size Y have you sold to date?
Q3 How many of size X have you currently got in stock?
Q4 How many of size Y have you currently got in stock?

and so on.

It should be said, however, that multiple questions are not always obvious. The question: How many of each sized packet have you bought?' (Belson, 1981:25), should be broken down into its implied parts:

Q1 How many size X packets have you bought?
Q2 How many size Y packets have you bought?

and so on.

And it is quite common for a second question to be implied by the way a question is worded. Fowler and Mangione (1990:84) discuss the problem of what they call 'hidden questions' — questions about whether other questions are actually relevant. For example, the question: 'For whom do you think you
will vote in the next election?" is really two questions. The first — "Will you vote in the election?" — is implied. And it would be better to ask it explicitly:

Q1 Will you vote in the next election?
and, if the answer is 'yes',
Q2 Who will you vote for?

Finally, an example discussed by Payne (1951:160) draws attention to the fact that the simple conjunction 'and' can imply a second question because it can be taken as either joining two subjects into one or adding a second subject. Thus the question: 'Is there much rivalry among the boys who sell the soda pop and crackerjack?' can be answered in terms of rivalry within a single group, rivalry between two groups, or rivalry within each of two groups. If the last possibility were intended, one would have to ask:

Q1 Is there rivalry among the boys who sell soda pop?
Q2 Is there rivalry among the boys who sell crackerjack?

Likewise, the question: 'Do you distrust banks and building societies?' can either be interpreted as a question about financial institutions in general or as two questions — one about banks and one about building societies.

The addition of clauses, phrases and instructions

Almost all survey methodologists advocate brevity in questions. It is a rule, nevertheless, that is often broken when researchers succumb to the temptation to add qualifying clauses, phrases or instructions. The problem is forcefully illustrated by the following question: 'Have you bought any in the last seven days not counting today?' (discussed by Belson, 1981:25). The trouble with such questions is that, because the first half makes sense by itself, respondents can begin to answer it before they have heard the second half. As Cannell (1977:27) comments: 'When questions are extremely long and complex, respondents often interrupt at the end of a clause to answer without allowing the interviewer to finish the question.'

Of course, the problem could be partly overcome by putting the qualifying component(s) first (as suggested by Kidder and Judd, 1986:246). It would be better to ask, for example: During the last seven days — not counting today, have you bought any X?

Although the general case for brevity in wording questions would be accepted by most methodologists, it should be pointed out that it is not regarded as a virtue in all quarters. In the last few years, it has been challenged by methodologists concerned with the problem of getting respondents to recall material and by methodologists concerned with the effects of question threat. Laurent (1972) found that questions that began with a repetitive or redundant introduction (e.g. 'The next question is about medicines. We want to know about this. What medicines, if any, did you take during the past four weeks?') induced respondents to give more information than briefer versions of the same questions. And Blair et al. (1977) and Sudman and Bradburn (1982) have reported that long, open questions worded in the respondents' own language increased the level of reporting for socially sensitive or threatening topics. While the results of these studies are provocative, they should not be over emphasised here. For one thing, it is not clear what is really going on in each case. In the use of the 'medicines studies', as Converse and Presser remark (1986:11-12), the additional words may have stimulated respondents to talk more (i.e. a modelling effect), aided recall, or just given respondents more time to formulate an answer. In the case of the studies of questions concerning threatening topics, it is not clear whether the extra words generated a modelling effect (i.e. if the interviewer can talk about the topic without embarrassment, so can I), desensitised respondents to the threat, or just gave respondents more time to formulate an answer. Alternatively, as Hippler and Schwartz (1987:106) suggest, respondents may take the length of the question to be an indication of the importance the researcher places upon the question. Whichever are the correct explanations, it may be better to deal with the implied factors directly — that is, to explicitly provide appropriate recall cues, make the threat as low as possible, make sure that respondents have plenty of time to think, and make sure that respondents know why the researcher wants the information. (Note: we will return to these issues in Chapters 7 and 9.) It is worth adding that none of the suggestions that have just been made actually conflict with the idea that questions should be as brief as possible without lessening the clarity of the definition of key concepts.

The use of negatives

Another structural complexity that has been associated with respondent confusion over the interpretation of questions is the use of negatives. Most writers counsel against the use of double negatives because they have to be translated into positives. For example, the question: 'What is your view about the statement that conservationists should not be so uncooperative with the government?' has to be translated into the question: 'What is your view about the statement that conservationists should cooperate with the government?'

Regrettably, double negatives can sneak into questions without being noticed. Converse and Presser (1986:15-14), for instance, point out that the rating response option 'Disagree' can lead to double negatives as happens with attitude scale items: 'Teachers should not be required to supervise students in the halls — Agree . . . disagree'.

But it is not only double negatives that cause confusion. Even single negatives can create difficulties. Akiyama et al. (1979) provide data that indicate that questions such as: 'Isn't a robin a bird?' or 'Aren't you going to do X?' are harder to answer than positively worded questions (e.g. 'Is a robin a bird?', or 'Are you going to do X?'). The underlying difficulty associated with single negatives
seems to be due to the implied presupposition (e.g. A robin is not a bird) that has to be identified and corrected before an answer can be given. And when an answer is given, it is not clear whether it applies to the original presupposition or to the corrected presupposition.

Linguists (e.g. Kearsey, 1976) have also discussed how the structure of negative questions can imply the answers expected. 'Are you going to do X?' implies a neutral expectation. 'You are going to do X aren't you?' implies a positive expectation (i.e. a 'yes' answer). 'Aren't you going to do X?' implies a negative expectation. And, 'You are going to do X, aren't you?' implies a doubtful expectation.

**Summary**

In this chapter we have looked at the issues that have to be confronted if we want to formulate intelligible requests for information. We saw that both difficult and abstract words are best avoided. If a difficult word has to be used, it is best to define it clearly for respondents. Further, it was recommended that any definition should precede the difficult word rather than follow it, since this course is less likely to insult the intelligence of respondents who already know what the word means. It was suggested that abstract words should be given empirical referents by either providing specific, concrete examples or by providing illustrative vignettes. And it was suggested that words with multiple meanings and words with moral overtones should be avoided. In sum, it was argued that the overriding principles should be brevity, simplicity and concreteness.

We saw that long questions are more likely to elicit biasing response effects. Complicating phrases and clauses are often ignored by respondents — especially when they are placed at the end of a question so that respondents can start to answer before they have heard the complete question. If qualifying phrases or clauses have to be used, they should be put first. For clarity's sake questions should be positively worded. Whenever possible, the use of negatives should be avoided because they make questions and answers more difficult to interpret correctly. Single negatives imply complicating implicit propositions that become confused with the intended question and double negatives have to be translated into positives.

As a way of reinforcing the conclusions that have been reached in this chapter, it is worth ending by referring to the results of a program of research carried out by Belson (1981). Belson and his colleagues began by analysing the questions from 2140 questionnaires developed by twenty-four researchers. Types of questions generally recognised to generate response problems were identified and listed in order of the frequency of their occurrence (see opposite).

Notice that all of the questions involved complexities that conflict with the principles of brevity, grammatical simplicity, specificity and concreteness. Belson and his colleagues then presented variants of the six most frequently occurring types of questions to respondents. All were misunderstood by at least 42% of the respondents: more than half were misunderstood by over 70% per cent.

**Belson's sixteen categories of difficult questions arranged in order of decreasing frequency of occurrence**

1. Two questions presented as one (e.g. 'Which brand do you use or do you change brands frequently?').
2. Questions with a lot of meaningful words (e.g. 'How many of each sized packet have you bought?).
3. Questions which include qualifying phrases or clauses (e.g. 'Have you bought any chocolate in the last 7 days, not counting today?').
4. Questions with multiple ideas or subjects (e.g. 'Which have you heard of or bought at?').
5. Questions that contain difficult or unfamiliar words.
6. Questions that contain one or more instructions (e.g. 'Do not include X in your answer').
7. Questions that start with words that are meant to soften them (e.g. 'Would you mind . . .?').
8. Questions with difficult phrases.
9. Hypothetical questions.
10. Questions that are dependent upon prior questions for meaning (e.g. 'Q1/ Did you buy a copy of X? Q2/ Where is it now?').
11. Questions with negative elements.
12. Inverted questions (e.g. 'The ones you bought last time — what were they?').
13. Questions including either 'if any' or 'if at all' (e.g. 'Which of these, if any, have you bought?').
14. Questions that are too long.
15. Questions that include both present and past tenses.
16. Questions in which singular and plural cases are used.

Summary

Even when the researcher is sure that respondents have been exposed to an event, it is dangerous to assume that they will be able to remember information about it. First, there is evidence that human beings do not attend to, and thus are not able to remember, every stimulus that impinges upon them. Second, respondents are often not consciously aware of all the causes and motives associated with their behaviour. Third, memory traces of events and the cues associated with these fade with time.

Even when it can safely be assumed that respondents would have once had the required information stored in their memories, it makes little sense to ask them to recall it if it was entered into their memories a long time ago. Current thinking is that it makes little sense to ask respondents to recall fine details about either highly salient events that occurred more than a year ago or events of low salience that occurred more than thirty days ago.

Besides taking into account the time that the information has supposedly been stored in memory, the researcher should take steps to facilitate accurate recall. Questions about past events need to be worded in ways that help respondents access the required information. Preliminary questions to elicit answers that in turn cue answers to subsequent questions are useful. Asking respondents to recall sequences of events by starting with the most recent events and working their way back is one application of this principle. Cross-cutting questions that help respondents restructure past events in their minds are also valuable. The use of 'bounded' recall procedures should be considered. And instructions that exhort respondents 'to take their time' and 'to try hard to remember accurately' have been found to increase the validity of the data respondents retrieve from their memories.

Without backing away from the general view that much can be done to help respondents recall material accurately from their memories, it must always be kept in mind that sometimes respondents may be motivated to not cooperate with the researcher. This is especially likely to be the case when respondents feel threatened within the question-answer situation. But we will return to this problem in chapter 9.

Last, but just as important as the other issues that have been raised, is the fact that short-term memory failures can cause comprehension failures in respondents. Some comprehension problems occur when questions are delivered too quickly. Others occur because of the way questions — especially complex questions — are presented to respondents. Effort needs to be directed toward the problem of determining the most effective way of presenting complex questions to respondents so that short-term memory problems are avoided. One possibility that should be entertained is that of letting respondents read them for themselves at the same time as the interviewer reads to them aloud.

Social researchers tend to assume high levels of opinionatedness on the part of respondents, so that little attention is paid to establishing whether or not respondents actually have had the necessary experience upon which an assumed opinion or belief could be based. This is a problem because it has been found that one cannot rely on respondents to indicate when questions are of marginal relevance to them. Respondents typically do their best to answer every question put to them — even questions that they have difficulty understanding or relating to — either by cognitively adjusting the question so they can answer it or by falling back on contextual clues and general attitudes to formulate an appropriate answer (see, e.g., Schuman and Presser, 1980; Bishop et al., 1980; Belson, 1981; Strack and Martin, 1987).

A question or question component that is explicitly offered to a respondent, either to establish the relevance of the question to the respondent, or to emphasise that it is acceptable for the respondent to not answer the question, is generally referred to as a 'filter'. A variety of filters have been invented ranging from the simple 'Don't know', 'Can't recall' and 'No opinion' response options, through the 'Undecided', 'Not sure', 'It depends' and 'Neutral' alternatives, to sets of questions that attempt to establish the extent to which an opinion or belief is based on experience of or knowledge about the topic.

In the process of developing a scale to measure attitudes towards unions, Dunnette et al. (1956) found that twenty-one of the 121 initial items generated rates of 'undecided' responses which varied between 20 and 30 per cent. In order to interpret these responses, they presented the twenty-one items with accompanying five-point rating scales (which included a middle 'Undecided'
response option) to another sample of respondents. A third sample of respondents was exposed to the same items, but this time a sixth response option, 'I don't know enough about this to answer', was added to the rating scales. On the basis of the differences between the response distributions for the three samples, the researchers inferred that almost all of the respondents who indicated that they did not know enough to answer an item, would have endorsed the 'Undecided' category if the sixth response option had not been provided. In other words, in the absence of the knowledge filter, the 'Undecided' responses meant either that respondents were neutral or ambivalent, or that respondents didn’t have opinions because they lacked the information upon which to base them.

This problem was again highlighted when Converse (1964) published an analysis of the attitudes held by American voters. Analysing responses from a sample of respondents who were interviewed three times for the Michigan Survey Research Center during 1956, 1958 and 1960, Converse discovered a high rate (sometimes as high as 80 per cent) of response inconsistency between successive surveys. Converse rejected the hypothesis that true social changes had affected the respondents who changed their opinions over time, because the number of respondents who had changed one way was balanced by the number who had changed the other way. In addition, the correlations between the responses to different questions were as high after four years as they had been after two. Converse discovered, too, that inconsistency was negatively related to education, to political involvement and to the simplicity and concreteness of the questions. These findings led Converse to invent the term 'non-attitudes' to refer to responses that, in his view, were essentially random responses which respondents had arrived at hastily to satisfy the demands of the interview situation.

The problem of 'non-attitudes' received further attention when Schuman and Presser (1980, 1981) reported that between 20 and 50 per cent of respondents who had been asked about an obscure bill before Congress, gave a 'Don't know' response when that option was explicitly available, whereas only 10 per cent of respondents volunteered a 'Don't know' response when provision for this was not explicitly made. They concluded that some 10 to 20 per cent of respondents who would give 'Don't know' responses when such a response is explicitly allowed, give a substantive response when it is not explicitly provided. Schuman and Presser invented the label 'floater' to refer to these respondents.

The 'non-attitude' problem could, of course, be ignored if it could be shown that respondents who give non-substantive responses when they are explicitly allowed but give substantive responses when they are not, are drawn equally from all of the substantive options. If this were the case, even if the absolute size of the percentages for each substantive option were to change, the relative differences between these percentages would not. It would also mean that any patterns of association in any cross-tabulations would not be disturbed. The last point defines what Schuman and Presser call the hypothesis of form-resistant correlations.

Unfortunately, a number of investigations have thrown up examples of questions for which the provision of an explicit, non-substantive response option has had the effect of radically changing the distribution of responses across all categories of the variables being investigated (e.g. Bishop et al., 1980, 1983; Kalton et al., 1980; Schuman and Presser, 1981). In some instances, the changes have been great enough to actually reverse patterns of association with other variables. Although most of the instances in which this has happened involve filters of the 'middle alternative kind' (e.g. 'Makes no difference', 'Undecided'), some involve stronger filters (e.g. 'Don't know' — Bishop et al. 1986:247). Table 8.1 (below) illustrates how the provision of a middle response alternative can affect responses.

In summary, we have learnt that typically up to 20 per cent of respondents will give a non-substantive response if they are allowed to but give a substantive response if a non-substantive option is not offered. We have learnt, too, that the

<table>
<thead>
<tr>
<th>With filter</th>
<th>Without filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Makes no difference' option explicitly provided</td>
<td>'Makes no difference' option not explicitly provided but accepted if volunteered</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>good</th>
<th>males</th>
<th>females</th>
<th>males</th>
<th>females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>64.0%</td>
<td>57.6%</td>
<td>40.1%</td>
<td>22.6%</td>
</tr>
<tr>
<td>9.2</td>
<td>14.0</td>
<td>17.5</td>
<td>18.5</td>
<td></td>
</tr>
<tr>
<td>50.7</td>
<td>63.4</td>
<td>18.5</td>
<td>25.2</td>
<td></td>
</tr>
</tbody>
</table>
| 100% | 100% | 100% | 100%
| (n 217) | (n 541) | (n 189) | (n 524) |
| p<.01 | NS |

Source: Kalton et al., 1980:76, table 3. Title and captions by author.
provision of a filter can completely change the distribution of responses to a question. Both of these findings raise the problem of how non-substantive responses should be interpreted. On what basis do those respondents who give non-substantive responses when these are allowed give substantive answers when they are not?

Work done by Bishop et al. (1980:86) is relevant to the question that has just been asked. They presented respondents with a number of questions about fictitious issues (i.e., issues invented by researchers). The example set out in table 8.2 (below) illustrates the kinds of results produced.

Bishop et al. (1980, 1986) findings are noteworthy because they are very similar to those reported by Schuman and Presser, except their results were obtained using questions concerning fictitious issues. It is one thing for 10–20 percent of respondents to give substantive responses to obscure issues. It is another matter altogether that a similar percentage would give substantive responses to questions concerning topics about which they can know nothing because the researcher has invented them!

Table 8.2 Responses to filtered and unfiltered versions of the question: ‘Do you agree or disagree with the idea that the Public Affairs Act (a fictitious Act) should be repealed?’ by trust in the government or public sector

<table>
<thead>
<tr>
<th>Trust in the government or public sector (measured using the SRC 3 item scale)</th>
<th>Filter condition (table excludes those who said they are not interested in politics)</th>
<th>Unfiltered condition (table includes those who said they are not interested in politics)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Response to question about repeal of PAA</strong></td>
<td><strong>High</strong></td>
<td><strong>Low</strong></td>
</tr>
<tr>
<td>Agree</td>
<td>25.0%</td>
<td>70.0%</td>
</tr>
<tr>
<td>Disagree</td>
<td>75.0</td>
<td>30.0</td>
</tr>
<tr>
<td><strong>High</strong></td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>(n 16)</strong></td>
<td><strong>(n 20)</strong></td>
<td><strong>(n 20)</strong></td>
</tr>
<tr>
<td><strong>p &lt; .05</strong></td>
<td>N.S.</td>
<td></td>
</tr>
</tbody>
</table>

Table 8.2 shows the percentage of respondents who agreed or disagreed with the idea that the Public Affairs Act should be repealed, divided by trust in the government or public sector. The table includes both filtered and unfiltered conditions.

Smith goes on to argue that the results of the relevant studies reported by Hartley (1946), Schuman and Presser (1980) and Bishop et al. (1980) all indicate that such responses can, in the main, be attributed to the third possibility:

Respondents reading into the question and answer in terms of some general predisposition toward the economy, the government, or tolerance. In one sense, when people rely on such predispositions they are showing ideological or constrained attitudes, since they are using general attitudes to supply responses to specific questions. On the Monetary Control Act example of Schuman and Presser, many of the tricked respondents were applying something like the following syllogism: Major premise: I support programs to curb inflation. Minor Premise: The Monetary Control Act is a program to curb inflation. Conclusion: I support the Monetary Control Act. The problem is that the minor premise that they imputed from the bill’s name is wrong. (Smith, 1984a: 225–224)

The phenomenon of responding to fictitious issues is clearly similar to Converse’s concept of ‘non-attitudes’. But, as Smith asserts, they differ in that such responses are not random since they correlate with the answers given to related, more general questions.

Yet even if forced substantive responses are not random, it does not follow that they should be equated to unforced substantive responses that are more likely to be grounded in the respondents’ experiences. It is relevant to note that the best predictors of non-substantive responses have been found to be education and involvement. Francis and Busch’s (1966) analysis of Michigan Survey Research Center data disclosed negative correlations of around 0.4 between non-substantive responses and both education and involvement. And Converse (1976) reports that non-substantive responses in both the Gallup and the Harris polls are negatively associated with respondents’ levels of education. The relevant results reported by Converse are set out in table 8.3 (p. 106).

It is worth noting, too, that after re-evaluating the 1956, 1958 and 1960 Michigan Survey Research Center’s panel data analysed by Converse (1964), Smith comments that:

looking at people who gave one ‘Don’t know’ in 1956 or 1958 and those who gave ‘Don’t know’ both times to the power and housing question, we found that they had...
decidedly less political interest and lower education than respondents who were either consistent or who switched responses. While 52 percent of those who gave substantive responses engaged in no political activity, 67 percent of those with one ‘Don’t know’ and 68 percent with two ‘Don’t knows’ had less than a high school education. Similar patterns appeared on voting, political interest, and political knowledge. (Smith, 1984a:234–235)

Smith (1984a:236) goes on to summarise the respondent characteristics associated with ‘Don’t know’ responses in the SRC’s panel surveys — they were low education, low occupational prestige, low income, female, old, low political activity, low political efficacy, housewife, and member of the general public rather than opinion leader.

The finding that both less well-educated and less involved respondents are the most likely to endorse filter options (when they are provided) is in line with the hypothesis that the main reason for respondents endorsing non-substantive options is that they lack the information sought by the researcher. Of course, it would be dangerous to conclude that this is always the case. Although the available evidence suggests that most non-substantive responses reflect either ambivalence or the lack of a substantive answer, it is possible that some respondents may use explicitly offered non-substantive response options as vehicles for other answers (e.g., ‘I don’t want to answer that question’, ‘I don’t understand what you are asking’, etc.).

Experiments designed to differentiate those respondents who genuinely have no opinion from those who are ambivalent suggest that, typically, between one- and two-thirds of the respondents who endorse an ‘Undecided’ option are really ambivalent, while most of the remainder have no opinion. Dunnette et al. (1956) compared the response distributions for a set of questions that included the filters ‘Undecided’ and ‘I don’t know enough to answer’, with the response distributions for the same set of questions without the ‘I don’t know enough to answer’ filter. Whereas between 20 and 40 per cent of the respondents endorsed the ‘Undecided’ filter when it was the only one provided, the percentage fell to less than 20 per cent when the ‘I don’t know enough to answer’ filter was provided as well. Smith (1984a) re-analysed the non-substantive responses to fifteen of the attitude items used in the Michigan Survey Research Center’s panel surveys in which respondents had been explicitly offered the non-substantive response options ‘Not sure’, ‘It depends’, ‘No opinion’ and ‘Don’t know’, and makes the comment that on average about 34 per cent of the respondents who gave non-substantive responses could be categorised as being ambivalent (because they had either answered ‘Not sure’ or ‘It depends’), and 60 per cent could be categorised as having ‘No opinion’. The meaning of the remaining 6 per cent was less clear but Smith speculates that the ‘Don’t know’ option appears to have been used by people who have an opinion but do not know what it is! Smith (1984a:229) concludes, ‘it is clear that ambivalent responses are common and may, on average, account for a third to one half of all non-substantive responses’.

In sum, it appears that many respondents give substantive answers, if they have to, that are not directly based on experience. Thus, if filters are not used, the meaning of the substantive answers will inevitably be vague. This conclusion reinforces the view that filters should always be employed. It has to be acknowledged, however, that when filters are used, the meaning of their endorsement is itself often unclear. It is pertinent to add that the ambiguity of non-substantive responses usually arises either because respondents are forced to use particular response options to express quite different meanings (as when the researcher has failed to provide a complete array of non-substantive options) or because those non-substantive options that are provided are ill-defined. The sensible course of action is to include two filters: a ‘No opinion’ filter plus an ‘I’m not sure’ or ‘It all depends’ filter.

### Table 8.3: The relationship between non-substantive responses and education

<table>
<thead>
<tr>
<th>% Giving 'No Opinion' &amp; 'Don't Know' responses</th>
<th>Gallup Polls</th>
<th>Harris Polls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest level of education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>High school</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Primary school</td>
<td>15</td>
<td>22</td>
</tr>
</tbody>
</table>


---

**What kind of ‘No opinion’ filter should be used?**

This is not a simple question because the meanings of the commonly used ‘Don’t know’ or ‘No opinion’ filters are not as clear as they might, at first seem. Bishop and his colleagues (1980, 1983) have investigated the endorsement rates for different filters accompanying a list of fictitious items which might be expected to generate high rates of ‘Don’t know’ and ‘No opinion’ responses. The following results are for the three filters endorsed most often when paired with a question about whether or not the respondent agreed that the fictitious Public Affairs Act should be repealed (i.e. a question that should generate very high ‘Don’t know’ or ‘No opinion’ rates):

- 'Have you been interested enough to favour one side or the other?'
- 'Don’t know' = 76.6 per cent.
'Do you have an opinion on this or not?'
'Don't know' = 82.5 per cent.

'Have you thought much about this issue — or haven't you thought much
about it?'
'Don't know' = 85.8 per cent. (Source: Bishop et al. 1980: table 1)
Subsequently, Bishop et al. conclude that a filter:
which asks respondents how interested they are in an issue, or how much they have
thought or read about it, will generally screen out more people than one which asks
simply whether they have an opinion. And the more abstract or remote the issue, the
greater the wording effect will be. (Bishop et al., 1985:545)
It should be added that Bishop et al. do not define what they mean by the
'strength' of a filter. It is possible that, for example, phrases such as: 'or not', 'to
favour one side or the other', 'have thought about it' and 'haven't thought about it'
make it easier or less embarrassing to not give a substantive answer. To
further complicate the matter, Hippler and Schwarz (1989) have reported that
the more strongly a filter is worded, the more strongly it suggests to respon-
dents that the question is important and should be answered only if they can
provide well-considered answers based on fact. This finding is a complication
because it raises the possibility that strong filters are likely to screen out respon-
dents who have weak views which, in spite of being weak, still influence their
daily behaviour, but which they believe are not what the researcher is after, as
well as those respondents who have interpreted the question as indicating that
the researcher is subsequently likely to ask an even more difficult question.

The positioning of filters

Poorly defined non-substantive options are prone to generate format effects.
This is demonstrated by a study carried out by Holdaway (1971). This study
showed that the percentage of respondents using the 'Undecided' option in
5-point, 'Strongly agree' to 'Strongly disagree' rating scales was almost 22 per
cent higher when it was placed in the middle than when it was placed at the end.

Holdaway suggests that respondents were more likely to interpret the
Undecided category as 'neutral' when it was placed in the middle — that is, it
appears that respondents used the position of the non-substantive option as a
cue to its interpretation. A similar finding has been reported by Bishop (1987a).
He and his fellow researchers investigated the effect of positioning the response
option 'Continued at its present level' in the middle versus at the end of a
number of questions. 'Do you think social security should be: increased,
decreased, continued at its present level?' is an example of the questions they
investigated. For four of six questions, the percentage of respondents who
endorsed the middle alternative was greater — on average 3.3 per cent greater
— when it was listed at the end of the question. Bishop suggests that:
the harder it is to choose between two opposite alternatives on an issue . . . the more
likely a person is to select the middle alternative when it is offered. This is particu-
larly so when it is presented at the end of the question, because it appears more
clearly to be a compromise between two equally attractive, or unattractive, alter-
avatives . . . (Bishop, 1987a:227)
Thus the interpretability of responses is not increased a great deal by using
ill-defined non-substantive options.

Having observed that respondents were somewhat more likely to endorse
the middle alternative when it was placed at the end of the question, Bishop
(1987a) does not offer any advice on its ideal positioning. Responses that are so
weak that altering the ordering of the response alternatives affects their like-
lihood should not be confused with substantive responses. The conservative
decision would be to always place the middle option (e.g. ' . . . or continued at its present
level') at the end. For example:

Some people say that the United States needs to develop new power sources from
nuclear energy in order to meet our needs for the future. Other people say that the
danger to the environment and the possibility of accidents is too great. What do you
think — are you in favour of building more nuclear power plants, would you prefer
to see all nuclear power plants closed down, or would you favour operating only those
already built or would you rather say that you 'Don't know' or 'Have never really
thought about it?'

Should a middle category be offered at all?

It is tempting to avoid worrying about the positioning of middle categories by
adopting the course of action suggested by Converse and Presser (1986:36) who
argue that, because the 'intensity' of the views held is the major characteristic
distinguishing respondents who are affected by the presence of a middle cat-

gory from those who are not, the researcher should:
not explicitly provide the middle category, and thereby avoid losing information
about the direction in which some people lean, but follow the question with an
intensity item, thus separating those who definitely occupy a position from those
who only lean toward it. (Converse and Presser, 1986:37)

They recommend measuring 'intensity' by asking the main question and
following it up with a question of the sort: 'How strongly do you feel about that?
— Extremely strongly, Very strongly, Somewhat strongly, or Not at all strongly’. One shortcoming of this approach, however, is that respondents who are truly neutral or ambivalent are equated with those who hold a substantive answer but indicate that they do not hold it very strongly. The use of both a ‘Don’t know’ or ‘No opinion’ filter and an ‘I’m not sure’ or ‘It all depends’ filter in conjunction with some sort of ‘intensity’ measure would avoid this objection. Whether the researcher should use the ‘intensity’ filter suggested by Converse and Presser (1986) or choose some other filter, however, is not so clear. In 1944 Katz found that an accompanying ‘sureness’ scale (i.e. ‘How sure are you that your opinion is right?: Not sure, Fairly sure, Very sure’) produced higher correlations between responses to questions related to the focal substantive question and responses to the focal question itself than a ‘How strongly do you feel on this issue?’ scale. More recently, Schuman and Presser (1981:234–245) have reported the results of a comparison between an ‘intensity’ measure (‘Compared to how you feel about other public issues, are your feelings on this issue: Extremely strong, Fairly strong, Not very strong?’); a measure of commitment to the issue in question (‘Have you ever written a letter or given money in defence of the issue?’); and a centrality measure (‘How important is this issue to you when you decide how to vote . . .?’). In conjunction with questions about gun control, they found that the ‘intensity’ measure was not as successful as either of the other two measures. They conclude that the ‘intensity’ measure failed because it was too easy for respondents to claim that they felt strongly about the issue being investigated. Thus, both the Katz and Schuman and Presser studies suggest that either a ‘sureness’ measure or an ‘importance’ measure is to be preferred over an ‘intensity’ measure.

Summary

It has been observed that it is a common occurrence for 10 to 20 per cent of respondents to answer a question in a substantive fashion when a non-substantive response option is not explicitly offered, but to shift to a non-substantive option when one is provided. Further, it appears that the researcher cannot rely upon respondents to indicate when some sort of filter is required. In other words, the fact that most respondents will answer a question, if forced, cannot be interpreted to mean that the question is equally relevant to each respondent. Further, since it has been found that the use of filters often changes the pattern of results, the researcher cannot simply ignore the problem of whether or not to allow non-substantive responses.

Both Kalton et al. (1980) and Schuman and Presser (1981:311–312) have argued that a ‘Don’t know’ response option should not be included if a general indication of community values, ideological biases, or stereotypes is required.

Yet this suggestion is not well founded. There is no justification for the practice of collecting and pooling answers that are uninterpretable in the sense that, although empirically the same, they mean quite different things to different respondents. For this reason, Sudman and Bradburn’s (1982:141) suggestion that ‘Don’t know’ and ‘No opinion’ options should always be included, because they generate additional information, is a much more defensible position. Following Hippler and Schwartz (1989), the researcher might consider using moderately strong filters that are less likely than very strong ones to inhibit respondents from answering even when they could give legitimate answers. On the other side of the coin, Kahn and Cannell (1957) are probably correct in suggesting that filters should be chosen to minimise the psychological impact of not being able to answer. For example, saying to respondents: ‘Many people have not had the opportunity to learn about X while others have picked up information about it — do you happen to know about X?’ is probably better than requiring respondents to admit that they do not have an answer because they have never heard about X and have never thought about it.

Furthermore, it makes no sense to make a filter do several jobs at once. The task of a filter should be to establish the relevance of the question to each respondent. Middle alternatives (e.g. ‘Neutral’, ‘Undecided’, ‘Ambivalent’) should be treated as legitimate substantive options and explicitly offered to respondents along with the standard ‘Don’t know’ option. For this reason, it makes sense to include an unambiguous ‘Undecided’ or ‘It all depends’ option. Again there is little justification in collecting uninterpretable answers, and even less in pooling them together.

In conclusion, it should not surprise anyone that Andrews (1984), after applying a sophisticated, multi-correlational, structural statistical modelling procedure to assess the validity of the data collected in six different surveys (totalling 7706 respondents), reaches the conclusion that:

The second most important survey characteristic is whether the answer categories include an explicit ‘Don’t know’ option. The effect of this design matter is clear and consistent: inclusion of an explicit ‘Don’t know’ category was associated with better data, higher validity . . . . The idea that one should let respondents ‘opt out’ if they lack the requisite information receives strong endorsement . . . (Andrews, 1984:431)

It should be added that Andrews looked only at the effect of the inclusion of the ‘Don’t know’ filter. There can be little doubt that his conclusions would have been even more emphatic had he been able to analyse data collected with the use of questions which included filters and middle alternatives that generate clearly interpretable answers!
Chapter 10

THE OPEN VS. CLOSED QUESTIONS DEBATE

Coding responses to open questions and formulating sets of response options for closed questions

Since the initial development of modern social survey techniques at the start of the twentieth century, social scientists have fallen into two camps: those who have been attracted to the use of open questions, and those favouring the use of closed questions. These two camps have become more entrenched with the passing of the decades.

On the one hand, advocates of open questions have been fortified by the thoughts of a number of the early sociologists (such as Max Weber and William Thomas), ethnographical anthropologists (such as Clifford Geertz) and more recently by ethnographers (such as Harold Garfinkel, Aaron Cicourel, Jack Douglas and Charles Briggs). All of these social scientists have stressed the cultural relativity of meaning — that is, that respondents' acts must be understood in terms of the meanings that they assign to them — coupled with the idea that the meanings that are ascribed to an act in one situation can be very different to the meanings ascribed to it in another situation.

On the other hand, although advocates of closed questions were initially led by those involved in the development of large scale, national survey techniques and by market researchers who needed to conduct surveys quickly (e.g. Paul Lazarsfeld and George Gallup), during the last few decades the emergence of computer technology has, if anything, increased the attractiveness of closed questions for them.

It is also probably fair to say that the arguments put forward by the two opposing camps have often looked more like 'articles of faith' than conclusions based on reason and evidence. Yet the methodological issues raised by the use of open and closed questions are fundamental and deserve careful consideration.

So that we will not lose our way or become bogged down in the 'fire' between the camps, the rest of this chapter has been structured according to the following plan:

1. The principal advantages and disadvantages associated with the two formats.
2. An evaluation of the assumptions underlying the use of open questions.
3. Problems associated with 'probing' inadequate answers to open questions.
4. Problems associated with coding responses to open questions.
5. An evaluation of the assumptions underlying the use of closed questions.
6. Problems associated with developing sets of response options for closed questions.
7. Problems associated with recording responses to closed questions.
8. Problems associated with interpreting responses to closed questions.
9. An evaluation of three additional uses claimed for open questions.

The principal advantages and disadvantages associated with the two formats

Proponents of the use of open questions argue that they allow respondents to say what is really on their minds without being influenced by suggestions from the researcher, whereas closed questions lock respondents into arbitrarily limited alternatives. They see closed questions as being typically decontextualised and as typically providing incomplete and arbitrarily closed sets of response options which are almost bound to distort respondents' answers. Survey researchers, on the other hand, think that open questions tend to produce material that is extremely variable, of low reliability and difficult to code.

Of course, not all social researchers have taken such clearcut positions. There have been those who have tried to adopt a compromise position. Lazarsfeld (1944), for instance, suggested that open questions should be used at the initial stage of a project, so that the appropriate response categories could be identified for use with closed questions, and also, at later stages in the research, to throw light on apparently deviant answers to the closed questions. In fact, most researchers would at least pay lip service to the first of these suggestions. The current normative position is implied in the following excerpt from Schuman and Presser (1979a:710):
[D]ifferences (in response distributions for open and closed versions of the same questions) will be minimised if investigators begin with open questions on large samples of the target population and use these responses to construct closed alternatives that reflect the substance and wording of what people say spontaneously. This point is so obvious as to be embarrassing to state yet it is probably violated in survey research more often than it is practised.

The most important of the claims that have been made regarding the two formats are summarised in table 10.1 (below). Although the relative merits of open and closed questions have been debated for much of this century, few attempts have actually been made to compare the kinds of answers actually produced by the two formats. Moreover, those that have been made (see Link, 1943; Crutchfield and Gordon, 1947; Dohrenwend, 1965; Schuman and Presser, 1979a) have reached conclusions which seem to be grounded more firmly on 'common sense' than on empirical findings.

Table 10.1  The most important claims that have been made regarding open and closed questions

<table>
<thead>
<tr>
<th>Open questions</th>
<th>Closed questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Allow respondents to express themselves in their own words.</td>
<td>(a) Allow respondents to answer the same question so that answers can be meaningfully compared.</td>
</tr>
<tr>
<td>(b) Do not suggest answers</td>
<td>(b) Produce less variable answers.</td>
</tr>
<tr>
<td>— indicate respondent's level of information</td>
<td>(c) Present a recognition, as opposed to a recall, task to respondents and for this reason respondents find them much easier to answer.</td>
</tr>
<tr>
<td>— indicate what is salient in the respondent's mind</td>
<td></td>
</tr>
<tr>
<td>— indicate strength of respondent's feelings</td>
<td></td>
</tr>
<tr>
<td>(c) Avoid format effects.</td>
<td>(d) Produce answers that are much easier to computerise and analyse.</td>
</tr>
<tr>
<td>(d) Allow complex motivational influences and frames of reference to be identified.</td>
<td></td>
</tr>
<tr>
<td>(e) Are a necessary prerequisite for the proper development of sets of response options for closed questions.</td>
<td></td>
</tr>
<tr>
<td>(f) Aid in the interpretation of deviant responses to closed questions.</td>
<td></td>
</tr>
</tbody>
</table>

An evaluation of the assumptions underlying the use of open questions

Open questions do not suggest answers to respondents

Perhaps the most persistent criticism of closed questions is that pre-set response options are likely to cause respondents to give answers they would not give if they had to provide them for themselves. The difference between the workings of open and closed questions in this respect is indicated by the results of a study that was referred to in chapter 1. Schuman and Presser (1979a) report that, whereas 22 per cent of the respondents who answered an open question about 'the most important problems facing the country,' mentioned food and energy shortages, only one out of 592 respondents used the 'other' category to give this answer to a corresponding closed question that did not include it as an option. What is especially interesting about these results is that they were obtained just after a severe winter that had caused food and energy shortages which had received a lot of media attention. Schuman and Presser are driven to remark:

[These] ... results bring home an obvious and yet profound point about survey questions: almost all respondents work within the substantive framework of priorities provided by the investigators, whether or not it fits their own priorities ... (Schuman and Presser, 1979a:707)

Even so, the implied suggestion that open questions do not suggest answers to respondents is not necessarily valid. Indeed, the common practice of using 'probes' to clarify the meaning of responses to open questions comes close to turning them into closed ones. (Note that this problem is taken up again below, pp. 154–8.) Further, the implied view that answers to open questions are sometimes more valid than answers to closed questions can be correct only if the answers to open questions can be interpreted and coded properly. Both of these requirements assume a lot, as will be made clear later (see pp. 158–9).

Answers to open questions indicate respondents' levels of knowledge about the topic

This assumption is suspect on a number of grounds. Most generally, open questions rest on the assumptions that they are relevant to respondents and that respondents can give correct answers within the question-answer situation. But it is easy to imagine that respondents can forget appropriate answers in the heat of the moment and there is little evidence that respondents necessarily mention the things that are most important to them first. Post-Freud, it is widely taken
for granted that people often repress, or refuse to disclose, psychologically or socially sensitive concerns. The overall situation is much more complex than is implied by the assumption that answers to open questions indicate respondents' levels of information. In fact this assumption rests on three more basic assumptions:

(a) respondents will answer an open question if they 'know the answer';
(b) respondents will not try to answer an open question if they do not know the answer; and,
(c) respondents will answer closed questions even when they do not know the answer.

It should be obvious that not one of these assumptions is likely to be invariably true. Consequently, the assumption that open questions indicate respondents' levels of knowledge about the topic will seldom be true.

**Answers to open questions indicate the salience of the topic in the respondents' minds**

Most methodologists have unquestioningly accepted the claim that answers to open questions indicate what issues are salient in the respondents' minds (e.g. Lazarsfeld, 1944; Dohrenwend, 1965; Schuman and Presser, 1979a; Converse and Presser, 1986; Schwarz and Hippler, 1987; Fowler and Mangione, 1990).

Few, if any, methodologists, however, have bothered to define the concept of 'salience'. This would be unproblematic if the meaning of the concept were clear, but it is not. Should it be taken to mean: 'importance to the respondent', 'most easily remembered', or 'central to the respondent's thoughts at the moment'? Fowler and Mangione (1990:90) beg the issue by recommending that the variability in the number of points made by respondents in response to an open question should be controlled by asking for the 'main' reason rather than 'all' reasons.

Assuming that salience is taken to mean 'importance to the respondent', there is little reason to believe that the first items mentioned by respondents in response to an open question are among the most important to them. In fact, some issues may not only be important but so threatening that respondents either repress or avoid talking about them altogether, while other issues may be so salient that respondents do not bother to mention them simply because they think that they are too obvious to mention.

The last possibility probably explains why responses to the 'twenty statements test' used in symbolic interactionist studies of the 'self', typically lack references to social roles that one would think should be central to respondents' self concepts (see Zurcher, 1977). The twenty statements test requires respondents to make up to twenty statements in answer to the open question, 'Who am I?' Respondents are told to list their answers in the order that they occur to them, not to worry about importance, and to go along 'fairly fast'. Although, a respondent might be expected to list such things as: (I am) male, a student, 20 years of age, ... and so on, typical response patterns have been found to include a surprisingly low proportion of references to gender and to work roles, even though there is no evidence to think that these have lessened in importance in Western societies during the last few decades. Hence, it is difficult to look at answers to the twenty statements test and not suspect that respondents have neglected to list such things as gender and work roles because they have assumed that these are too obvious to mention within the test situation (i.e. because it can be assumed that the researcher already knows these things).

Following on from this observation, respondents may also either mention or not mention items because the test situation either accentuates or does not accentuate them. McQuire and McQuire (1982), for instance, report that the likelihood that a respondent will mention a particular status factor in answers to the twenty statements test increases if the respondent is different in terms of that status factor from the others taking the test (e.g. is the only female, is the only black person, is the youngest or oldest person, etc.).

The commonly observed phenomenon of job applicants walking out of interviews 'kicking themselves' for forgetting to make 'this' point or 'that' point also indicates how situational pressures and distractions can change the salience of different issues.

**Respondents' answers to an open question indicate the strength of their feelings about the topic**

Although this assumption is as widely held as the 'salience' assumption is, it too has little, if any, published evidence to substantiate it.

**Open questions avoid the format effects that have been associated with closed questions**

The format effects that have been of concern include: the tendency to tick in the middle of rating scales, the tendency to endorse the most socially desirable or acceptable answer, the tendency to endorse the first response option seen, the tendency to endorse the last option heard, and the tendency for judgements to be affected by end items or anchor points.

Some of these effects (e.g. the tendency to endorse the first response option...
seen or the last one heard) are easily controlled by varying the order in which response options for closed questions are presented to respondents. And anchor effects (i.e. contextual effects) can be minimised by making sure that the range of responses that is offered to respondents is appropriate. Schwarz and Hippler (1987:174) may be right when they suggest that the range covered by response options can be a source of information for respondents about the sorts of answers that the researcher sees as 'normal'. But in these cases the researcher has only to make sure that the range of responses offered sensibly encompasses the likely range of respondents' positions in regard to the topic so that they are not made to feel deviant to minimise the problem. (Note this was discussed in chapter 5 and is discussed again in chapter 11.)

As for any tendency for respondents to avoid strong evaluations and to endorse socially desirable response options, it has never been shown — nor is it clear how it could be shown — that closed questions are likely to suffer more than open questions from such problems. Indeed, if the researcher does not specify the array of possible answers, it would seem to be impossible to assess the strength of such tendencies.

Thus the criticism that closed questions are more susceptible to format effects than open questions has little basis.

**Answers to open questions allow complex motivational influences and frames of reference to be identified**

Lazarsfeld (1944) was one of the first to specifically suggest that open questions are useful for:

(a) clarifying the meaning of respondents' answers;

(b) discerning influences on opinions (for example, the dimensions of the topic and personal motivations that have influenced respondents); and

(c) clarifying the nature of the relationship between variables.

These presumed benefits of questions are clearly not independent. More to the point, they imply that the answers to an open question indicate:

(a) how a respondent has interpreted it;

(b) the underlying motivation(s) influencing the respondent's orientation to the topic; and

(c) the frame of reference that the respondent has employed.

Each of these implications awaits critical evaluation.

An answer to an open question indicates the way in which a respondent has interpreted it

In chapters 3, 4, 5 and 6 we looked at the conditions that must be met if questions are to elicit interpretable answers. It was argued that each respondent must both define the topic in the same way as the researcher and also give the kind of answer the researcher requires. In regard to the latter requirement, it was argued that there are a number of degrees of freedom that respondents can exercise when formulating answers to either open or closed questions and that the only way to avoid having different respondents give different kinds of answers is to specify the kinds of answers required. It was argued that this has to be done because the ways in which respondents have defined a topic and the perspectives they have adopted are not necessarily apparent in their answers. If anything, however, this argument is even more pertinent to open questions than closed questions because open questions afford fewer clues as to what kind of answer the researcher expects. In other words, there are good reasons for rejecting the hypothesis that answers to an open question indicate the ways in which respondents have interpreted it.

An answer to an open question indicates the motivation(s) that have influenced the respondent's orientation to the topic

Many researchers have been unwilling to accept the assumption that open questions are useful for exploring complex motivational states:

[The] normal human being in Western civilisation is usually quite unable to realise — even to himself — what are the motivations for many of his actions . . . many of these motives are in fact relatively simple and may be ascertained if the problem is tackled scientifically . . . (e.g. with projective tests) . . . but to attempt to ascertain them by asking point blank: Why do you do this? is not only a waste of time and money but also a course of proceeding which any reputable practitioner or market researcher should regard as unethical. (Henry, 1971:294)

As was noted in chapter 7, Henry reports the results of several market research experiments that suggest respondents had been influenced by factors of which they had not been aware. It should be added that, even if respondents had been aware of their motivations, there is no reason to expect them to disclose them unless they had been explicitly asked to do so.
The answer to an open question indicates the frame of reference used by the respondent.

The fact that respondents can formulate responses from a large (if not infinite) array of perspectives (see chapters 4 and 5), makes it theoretically nonsensical to pick on any one and treat it as the only 'real' explanation. Requests for explanations are generally likely to simply pressure respondents into giving answers that they think will satisfy the researcher. Not necessarily at odds with this argument is the tendency, identified by attribution theorists in social psychology (see, e.g., Monson and Snyder, 1977) for respondents to give answers from their own standpoint. Mentel (1978) gives an example that illustrates how this works. He notes that soldiers questioned about the reasons for the My Lai massacre might have talked about how it was impossible to distinguish the Viet Cong from the South Vietnamese and how they had merely carried out their orders. An equally good explanation (though less likely to be actually given according to attribution theory) could have been cast in terms of the political processes current in the United States at that time — for example, the pressure to produce evidence to demonstrate that the war was being won (e.g. such as documenting rising daily body counts). In the final analysis, causes that are taken into account reflect both implicit and explicit interpretative decisions made by both the researcher and the respondents. This implies that the response framework that should be used needs to be specified by the researcher or that the researcher must attempt to delineate the types of explanations which different kinds of respondents typically give in different kinds of situations, if answers are to be both interpretable and comparable. The need to identify the nature of the perspectives used by the respondents stands as a problem that cannot be avoided.

Problems associated with probing inadequate answers to open questions

Perhaps because open questions require respondents to first decide 'how' to answer before answering, answers to open questions are much more variable than answers to closed questions (e.g. Cantril and Fried, 1944:11; Schuman and Presser 1979a:704; Potter and Mulkay, 1985). This puts pressure on interviewers to continually ask themselves whether or not particular answers are of the sort they require: can they understand the answers respondents have given; have the respondents provided the information they want?

The solution usually suggested, both for the tendency of respondents to stray from the topic when answering open questions and for the high frequency of incomplete answers, is to make sure that interviewers clearly understand the researchers' goals so that they can 'probe' respondents' minds to clarify obscure answers. Campbell (1965:347), comments that the open question:

is by no means an infallible instrument: the freedom of response which it permits can lead to ambiguity rather than clarity. The successful use of this technique requires precise and thoughtful formulation of questions and careful, intelligent interviewing . . .

Converse and Presser (1986:67) echo Campbell's view when they assert that rough codes for the possible responses to open questions should be designed in advance and explained to the interviewers so that they can recognise when a question has been adequately answered. And Zeisel (1985:205) writes:

The major interviewing problem derives from the difference between what the respondent thinks is a satisfactory answer and what the interviewer regards as satisfactory. The interviewer must be familiar with the accounting scheme because she must try to obtain answers for each dimension.

To this purpose, the interviewer must help her respondent along by encouraging more specific or more complete answers or by asking for the resolution of contradictions . . .

The irony of this line of advice is that it begs the question of whether open questions are better or worse than closed questions because, in essence, it turns them into closed questions.

Just as important as the principle that interviewers should know the required sorts of answers is the principle that interviewers should attempt to get adequate answers. Fowler and Mangione (both of the Massachusetts Center for Survey Research, 1990:41) draw attention to the fact that existing manuals vary greatly in the advice they give about how inadequate answers should be dealt with: 'Some organisations seem to accept, or even encourage, interviewers to find a variety of conversational ways to get respondents to clarify or elaborate their answers'. They go on to say that:

Our preference is to have interviewers stick with a very small list of probes.

In fact, we train interviewers that in addition to repeating the question, they only need to use three probes:
1. How do you mean that?
2. Tell me more about that.
3. Anything else?

These three probes are easy to remember. They are non directive. they do not give interviewers any opportunity to innovate in ways that would make their interviews different across respondents or interviewers.

The interviewer's task is to decide which of these probes is appropriate, and that involves analyzing the respondent's answer. The four probes, including repeating
the question, correspond to the four ways in which a respondent’s answer can be inadequate:
1. The response can fail to answer the question; it answers some other question. The interviewer should repeat the question.
2. The answer contains unclear concepts or terms that make its meaning ambiguous. The interviewer should probe, saying, ‘How do you mean (that)?’
3. The answer is not detailed enough or specific enough. The interviewer should probe saying, ‘Could you tell me more about (that)?’
4. A perfectly appropriate answer has been given, but there is a possibility that there are additional points that the respondent could make in answer to the question. The interviewer should ask, ‘Is there anything else?’ (Fowler and Mangione, 1990:41–42)

The idea that probes should be non-directive is obviously crucial. Yet how well do the limited, standardised probes suggested by Fowler and Mangione meet this requirement? An analysis of an illustrative example that Fowler and Mangione (1990:42–43) discuss at length is instructive. A precis of this example follows:

Q. From your point of view, what are the best things about living in this neighbourhood?
A1. In the last neighbourhood in which we lived, it was very transient. People didn’t care about keeping up the neighbourhood.

[Decision — answer does not answer question — use probe: repeat question]

A2. The people.

[Decision — answer is unintelligible — use probe: Tell me more about that.]

A3. The people are good neighbours.

[Decision — answer needs further clarification — use probe: How do you mean ‘good neighbours’?]

A4. They keep to themselves. They leave you alone. You don’t have to worry about being sociable and you don’t have to worry about what they think.

Notice that in this example the respondent’s focus of attention in A1 is quite different from that in A4. In A1 the focus is on the way the neighbours keep up the neighbourhood; in A4, it is on the way the neighbours keep to themselves. Fowler and Mangione claim that A1 is not relevant. But it is not clear why it is not relevant. It is true that the focus is upon the last neighbourhood the respondent lived in rather than upon the new one. Nevertheless, the respondent may very well be saying something about the new neighbourhood by comparing it to the last one. Likewise, although A4 is overtly about the new neighbourhood, it may nevertheless contain elements of comparison with the last

neighbourhood. In other words, it is not necessarily the case that A4 is any more relevant to the question than A1. What must be said, however, is that the dimension focused upon in A1 is quite different from that being focused upon in A4. Just why this shift of focus occurs is not clear. But the most likely explanation would seem to be that the interviewer’s probing caused it to happen. It is as if the probes signalled to the respondent that the answer given first is somehow inadequate so that the respondent has ‘tried again’ to find an answer that will satisfy the interviewer. The most obvious explanation for this is that the respondent was not told, when the question was first presented, what sort of answer was required.

If the task of ensuring that answers to open questions are relevant poses methodological difficulties, the task of ensuring that they are complete raises similar problems. Link captures the problem succinctly:

[The] depth interviewer is supposed to use his own ingenuity in asking impromptu questions, and to keep up this process until the particular subject seems to have been wrung dry . . .

Obviously, from this description, the responses obtained by a depth interviewer reflect the mind of the interviewer as well as the mind of the person interviewed. When we come to evaluate such interviews, we are at once confronted by this problem, namely: which of the ideas expressed represent the real thoughts of the respondent, and which represent the thoughts which the interviewer, by suggestion, has construed to elicit? (Link, 1943:269)

Similarly, Smith and Hyman (1950) report the results of a study which indicate a tendency on the part of interviewers to record the results that they expect to hear rather than those they are actually given.

In addition, a related difficulty arises because different respondents give different amounts of information. Some respondents are capable of giving lengthy, multi-stranded answers while others respond with the briefest of answers to the same questions. Adverbial questions — that is, ‘how’, ‘where’, and ‘why’ — questions are particularly prone to eliciting answers of varying length. Fowler and Mangione (1990:90–91) suggest that the researcher can avoid this problem by narrowing the discretion of both interviewers and respondents about the kinds of answers that count. They recommend translating general, adverbial questions into more specific questions. For example, ‘How do you get to work?’ might become:

‘What kind of transport do you use to get to work?’; or

‘By what route do you proceed to work?’

And, as has already been noted, they suggest limiting the number of responses to ‘why’ questions by asking for the main reason rather than for all reasons, so that each respondent only gives one answer.
Again, no matter how sensible the advice that has been given about how interviewers should 'probe' to improve the adequacy of respondents' answers, such advice avoids the question of whether or not open questions are more useful than closed questions as indicators of what issues are important to respondents and of how their answers should be interpreted. The problem is that the use of 'probes' inevitably comes close to turning open questions into loosely formulated closed questions.

**Problems associated with coding responses to open questions**

Advocates of open questions usually hold that a satisfactory coding schema can be formulated by going through a sample of responses several times to get a sense of the sort of categories into which the responses naturally fall (e.g. Silvey, 1975:88; Montgomery and Crittendon, 1977; Mostyn, 1985). Once a set of consistent categories has emerged, the remaining responses are coded. If one adopts this view, one implicitly assumes that respondents' answers speak for themselves. But this simple approach ignores all of the issues that have been raised in chapters 4 and 5, and in earlier sections of this chapter.

Consider, for instance, the deceptively simple question: 'What is your father's occupation?' The author had a sample of students answer this question. The answers given are set out in table 10.2 (p. 189). How should such answers be coded? The coding system chosen should reflect the reason that the research was carried out in the first place. If the intention had been to use the answers as indications of social class, they could be coded according to the ranking of each occupation mentioned on an established occupational prestige scale. If the intention had been to look at the impact of the fathers' occupations on the respondents' lives, the information might be classified according to the hours worked. If the intention was to classify fathers in terms of their probable relationship to different economic interests, the answers would need to be classified in yet another way. And so on.

Looking at table 10.2 it can be seen that most answers contain insufficient information to allow them to be classified properly in terms of any one of the possible ways suggested above. Even for the relatively simple goal of classifying paternal occupational prestige, many of the answers do not contain all the information that would allow this to be done. The obvious conclusion that has to be drawn here is that a question should be put to respondents in such a way that it elicits complete answers of the kind required. More generally, in the absence of clear guidelines about what sorts of answers are required, respondents can, and will, focus upon such a variety of aspects of the topic that their answers are neither comparable nor codable.

<table>
<thead>
<tr>
<th>Table 10.2</th>
<th>A sample of answers to an open question regarding father's occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The question:</strong></td>
<td>Your father's occupation is (or was)? <em>Give details:</em></td>
</tr>
<tr>
<td><strong>Nature of business:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Size of business:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Owner</strong></td>
<td><strong>Employee</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A sample of the answers given:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Real estate agent</td>
<td>Employee</td>
</tr>
<tr>
<td>2 Tailor</td>
<td>Employee</td>
</tr>
<tr>
<td>3 Taxi</td>
<td>Medium</td>
</tr>
<tr>
<td>4 Motor mechanic</td>
<td>Unknown</td>
</tr>
<tr>
<td>5 Timber merchant</td>
<td>Small</td>
</tr>
<tr>
<td>6 Commercial manager telecommunications</td>
<td>Philips</td>
</tr>
<tr>
<td>7 Farming</td>
<td>3000 acres</td>
</tr>
<tr>
<td>8 Research organisation</td>
<td>Small</td>
</tr>
<tr>
<td>9 Bus driver</td>
<td></td>
</tr>
<tr>
<td>10 Exploration company</td>
<td>Large</td>
</tr>
<tr>
<td>11 Stipendiary magistrate</td>
<td></td>
</tr>
<tr>
<td>12 Builder</td>
<td>Small-medium</td>
</tr>
<tr>
<td>13 Plumber</td>
<td>2 people</td>
</tr>
<tr>
<td>14 Car manufacturer</td>
<td>Very Large</td>
</tr>
<tr>
<td>15 Foreman</td>
<td>Large</td>
</tr>
<tr>
<td>16 Small bus. plastic components manufacturer</td>
<td>Small</td>
</tr>
<tr>
<td>17 Doctor/ GP</td>
<td></td>
</tr>
<tr>
<td>18 Tool setter</td>
<td>Big</td>
</tr>
<tr>
<td>19 Builder</td>
<td>Medium</td>
</tr>
<tr>
<td>20 Garage proprietor</td>
<td>1 employee</td>
</tr>
<tr>
<td>21 Shop assistant</td>
<td></td>
</tr>
<tr>
<td>22 Imports and exports</td>
<td>Medium</td>
</tr>
<tr>
<td>23 Manufacturing firm</td>
<td>200 people</td>
</tr>
<tr>
<td>24 Teacher/library asst</td>
<td></td>
</tr>
<tr>
<td>25 Photography</td>
<td></td>
</tr>
<tr>
<td>26 Wheat production</td>
<td>Family</td>
</tr>
<tr>
<td>27 Clerical</td>
<td>Multi-national</td>
</tr>
<tr>
<td>28 Teacher/catering</td>
<td></td>
</tr>
</tbody>
</table>

*Note: Responses to this question are set out here to demonstrate the general point that has been made. Although more complex question sequences have been developed to establish respondents' occupational status — see chapter 3, pp. 27-8 — it could be argued that they, too, suffer from the same problems as the question used here.*
An evaluation of the assumptions underlying the use of closed questions

Assumption 1: Because all respondents answer the question in the same way, the answers can be meaningfully compared

A great deal of evidence on the falsity of this assumption has been presented in earlier chapters. In chapters 3 and 4 it was noted that if different respondents are to interpret a question in the same way they must give the same meaning to key words. Yet even when they have done this, respondents seldom interpret questions literally. They adopt a perspective that includes, among other things, assumptions about the sort of information the researcher ‘really’ wants. If different respondents give different meanings to key concepts and adopt different perspectives, they will, in fact, be answering different questions. In other words, the fact that every respondent has been exposed to the same words is no guarantee that they will have understood the question in the same way. Thus it is clear that the researcher should try to ensure that respondents will both define key words in the same way and adopt the same perspective. While this may require some effort on the part of the researcher, it is not an impossible task (see chapters 3 to 6).

It is undoubtedly the case that closed questions go a long way towards limiting the kinds of answers respondents give because they require respondents to select responses from sets of response options. The provision of response options necessarily means that respondents are provided with clues as to the intended focus of the question, the required level of social generality of the answer, and so on. Nevertheless, it must be recognised that providing arrays of response options will not guarantee that key terms are defined in the same way or that the overall perspectives adopted by respondents are the same.

Assumption 2: Respondents find closed questions easier to answer

This assumption is easily supported. Besides helping respondents decide what sort of responses are appropriate, preset response options inevitably act as prompts which help respondents recall information that they might otherwise forget. For this reason, it is not surprising that Dohrenwend (1965:180) reports that the proportion of usable responses declined during the course of experimental interviews for open questions while remaining constant for closed questions. Likewise it is not surprising that Loftus (1982) has reported that when respondents are allowed to testify freely in legal cases they produce accounts that are far less complete.

Assumption 3: Answers to closed questions are more easily analysed

The claim that closed questions produce data that are more easily analysed than open questions is at one level almost true by definition. Respondents’ answers are, after all, constrained by the categories that are provided and are therefore less variable. True, that is, if one is willing to assume that every respondent who selects a particular category has interpreted it in the same way. We have seen, however, that there are no guarantees that this is always the case. Indeed, it was because of this that Lazarsfeld (1944:48) argued that open questions are useful aids for making sense of apparently deviant responses elicited by closed questions. Discussing the way in which political scientists had used open questions, he noted:

The general pattern of these studies proceeds from an empirical correlation which is not very high. We take cases which do not follow the majority pattern and try to gain an impression or to account for their irregularity. The political scientist is used to such a procedure. He knows, for instance, that the more poor people and Catholics live in a given precinct of a big city, the more Democratic votes he can expect. But here is a precinct which qualifies on both scores, and still it went Republican. What accounts for this deviation? Is the Democratic machine inefficient? Has a special local grievance developed? Was there a recent influx of people with different political traditions? This is quite analogous to what we are trying to do when we are faced with individual cases which went statistically out of line. With the help of open interviews we try to discover new factors which, if properly introduced, would improve our multiple correlation. (Lazarsfeld, 1944:48)

The point is, if closed questions can not be interpreted without the researcher having to fall back on to open questions, the superiority of closed questions may be more apparent than real.

Problems associated with developing response categories for closed questions

In the past, most methodologists have accepted the idea that the use of open questions is unavoidable during the initial stages of question design. Lazarsfeld (1944) argued that the respondents’ motivations could be adequately explored
with closed questions provided that they were well grounded in the results of pilot work based on open questions. Schuman and Presser (1979a), too, contend that closed questions do not suggest invalid answers to questions as long as the alternatives that are presented to respondents have been properly developed through pilot work with open questions followed by pre-testing to ensure that the set of response options is appropriate in the sense that all of the required categories, and no inappropriate categories, have been included. Fowler and Mangione (1990:18–19), too, assert that:

Exploratory research usually is not done best using standardized interviews. By design, in a standardized interview one only learns the answers to the questions that are asked. At the exploratory stages of research, finding out which questions to ask is a major goal. Restricting or structuring answers should not be done until the researcher is sure the answer options are comprehensive and appropriate.

It is worth observing again that few methodologists have raised any objections to Lazarsfeld’s (1944) recommendation that researchers should begin with open questions in pilot work and use the resulting responses as a basis for developing meaningful and exhaustive sets of response alternatives for closed questions.

Clearly, the completeness of the response set that accompanies a closed question is a critically important issue. Moreover, it is an issue that is made even more important by the finding that the widespread practice of including an ‘Other’ category does not seem to suffice (see Belson and Duncan, 1962; Schuman and Presser, 1979a:707).

Cantril and Fried (1944:11) present a convincing example of what happens when response options are not appropriate. They cite the results of a national survey conducted by the United States Office of Public Opinion Research in which the question was asked, ‘If Russia should defeat Germany, which of these things do you think Russia would then try to do — try to spread Communism all through Europe, or work with Britain and the U.S. in making it possible for the countries of Europe to choose their own form of government?’ Forty-one per cent thought Russia would try to spread Communism, 56 per cent thought she would try to work with Britain and the United States and 23 per cent had no opinion. Cantril and Fried then note that the Office of Public Opinion Research subsequently had a sample of forty respondents answer an open version of the closed survey question — that is, the forty respondents were asked, ‘If Russia should defeat Germany, what do you think Russia would try to do next?’ Only three of the 40 gave answers that could be said to fall into either of the survey alternatives. Seven (17.5%) couldn’t answer. The remaining thirty respondents’ answers were spread quite evenly over another sixteen categories. As Cantril and Fried remark: ‘With respect to issues that are not clear-cut... armchair attempts to design alternatives may prove extremely misleading’.

The danger is that it is very easy to invent a set of categories on an a priori basis which appear to be self-evidently adequate. Another example will serve to reinforce this lesson. A researcher might feel that the categories ‘Very good’, ‘Good’, ‘Fair’, or ‘Poor’, might be appropriate response options for the question, ‘How would you rate the school your child attends?’ What should one conclude, however, if a respondent answers as follows?

That is complicated for me. My child is in the second grade. I do not think they are doing a very good job in areas such as numbers and reading skills. However, at this age, I do not think that the content of what they learn is all that important. Kids can catch up in that respect. On the other hand, I think they do an excellent job in recreational areas such as gym and recess, where there is good equipment and a lot of opportunities for the kids to do things together. I think that is very important at this age. (Fowler and Mangione, 1990:15)

And the following example illustrates the same problem. While most respondents in Western societies can be made to place themselves in one of the following age categories: 10–19, 20–29, 30–59, etc., Carter (1971:24 — cited by Pawson, 1989:43) has suggested that the categories: ‘Infancy’, ‘Adolescence’, ‘Young adulthood’, and ‘Old age’ might better reflect the age-related distinctions that are actually made; and the categories: ‘Child’, ‘Warrior’, and ‘Elder’, might be preferable in certain pre-industrial societies.

It must be emphasised that it is all too easy to formulate inappropriate categories for use with closed questions. It is also very easy to fail to properly list a complete set of appropriate response options. One class of questions for which this basic requirement is commonly overlooked is discussed by Payne (1951: chapter 4). Payne refers to this class as ‘two-way’, ‘dichotomous’ or ‘bifurcated’ questions. Questions of this type suggest only two possible alternatives (e.g. yes or no, approve or disapprove, favour or oppose, etc.). Payne observes that:

This type of question is by far the most commonly used of all. It appears to fit the largest number of situations. It reduces issues to their simplest terms and its advocates say that it comes closest to duplicating the types of decisions that people are accustomed to making. And whether we realise it or not, it is probably correct that even our complicated decisions are broken down into many separate two-way issues. (Payne, 1951:55)

The inherent fault with dichotomous questions is that they imply rather than explicitly state one of the alternatives and thus underemphasise it. For example, the question, ‘Are you going to the game — yes or no?’ does not give equal weight to the possibility that the respondent is not going to the game (i.e., ‘Are you going to the game or not?’). Payne reports the results of an investigation into the effect of adding ‘or not’ to such questions. Whereas 77 per cent of one sample of respondents agreed with the statement, ‘Do you think anything could be done to make it easier for people to pay doctor’s or hospital bills?, 74 per
cent of another similar sample of respondents, for whom the phrase 'or not' was added, responded in the same way. No doubt the proportion of respondents agreeing would have been even larger if the second option had been given its full weight (e.g., ‘. . . or do you think there is nothing that could be done to make it easier for people to pay doctor or hospital bills?’). Given this, one might predict that simple yes-no questions would be biased towards 'Yes' answers because the negative option is not given full weight. In fact, this is the conclusion that Molenaar (1982:58–59) reaches on the basis of a review of the literature. Molenaar expresses the view that there is a greater tendency for respondents to endorse a response option when it is presented by itself in a yes-no format than when it is presented with fully explicated contrasting alternatives.

But to return to the basic issue that is being addressed in this section: is it true that a set of response options for a closed question will be both more appropriate and more complete if it is based on responses to an open question? Few studies have been conducted to address this claim. Of those that have been carried out, one in particular is especially revealing. It is Schuman and Presser's (1979a) attempt to use the responses to an open question to improve the response set for a corresponding closed question. Schuman and Presser begin by comparing the responses elicited by a closed question (first used by Lenski in 1963 to gauge what people value in their jobs) with the responses elicited by an open version of the same question. The two questions and associated response distributions are set out in table 10.3 (p. 145). The first thing to notice is that the responses to the open question are far more diverse than are the responses to the closed question. Indeed, the five closed question response categories account for less than half of the responses given to the open question. Second, it should be noticed that the differences between two response distributions vary greatly over the first five categories. In fact, they are so different that they correlate quite differently with both gender and education.

Schuman and Presser admit that it is not clear which question should be seen to have produced the most valid results. Nevertheless, they entertain the possibility that the discrepancies between the two response distributions might have occurred because the response options accompanying the closed question had not been appropriate. To test this possibility, they revised the response categories for the closed question to make them more like the codes that had been developed to classify the answers to the open question. In doing this, they dropped the third and fourth options because they had been endorsed by tiny percentages of respondents answering the open question, and they added the fifth and sixth codes for the open question responses. The eighth code for the open question was not added to the revised closed question response options even though it accounted for 17 percent of the open question responses because it was a vaguely defined, general category. Finally, they relabelled the

| Table 10.3 First comparison of response distributions for open and closed versions of the work values question |
|-----------------------------------------------------------|-----------------------------------------------------------|
| **Closed question:**                                      | **Open question:**                                         |
| This next question is on the subject of work.             | This next question is on the subject of work.              |
| Please look at this card and tell me which thing on this list you would prefer in a job? | People look for different things in a job. What would you most prefer in a job? — |

**Answers coded:**

<table>
<thead>
<tr>
<th></th>
<th>Percentage</th>
<th></th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) High income</td>
<td>12.4%</td>
<td>(1) Pay</td>
<td>11.5%</td>
</tr>
<tr>
<td>(2) No danger of being fired</td>
<td>7.2%</td>
<td>(2) Security — steady employment</td>
<td>6.7%</td>
</tr>
<tr>
<td>(3) Working hours are short: lots of free time</td>
<td>3.0%</td>
<td>(5) Short hours — time for other things</td>
<td>0.9%</td>
</tr>
<tr>
<td>(4) Chances for advancement</td>
<td>17.2%</td>
<td>(4) Opportunity for promotion</td>
<td>1.8%</td>
</tr>
<tr>
<td>(5) The work is important and gives a feeling of accomplishment</td>
<td>59.1%</td>
<td>(5) Stimulating work — Work that is challenging and gives a sense of accomplishment</td>
<td>21.3%</td>
</tr>
<tr>
<td>(6) Pleasant — enjoyable work</td>
<td>15.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7) Conditions include control over work &amp; physical conditions</td>
<td>14.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8) Liking the job — non-specific — not codable as 5 or 6</td>
<td>17.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9) Responses specific to particular job</td>
<td>3.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10) More than 1 codable response</td>
<td>1.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(11) Other</td>
<td>2.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) Don't know/no opin.</td>
<td>1.1%</td>
<td>(12) Don't know/no opin.</td>
<td>4.1%</td>
</tr>
</tbody>
</table>

100% (n = 460) 100% (n = 456)

categories to more closely reflect the nature of the cases that had been subsumed under each of the codes for the open question. Two new surveys were then conducted to test the revised set of response options and codes. The response distributions for the revised procedures are set out in table 10.4 (p. 147).

A number of things are worth noting about the new results. In the first place, although the revised response options for the closed question account for a higher percentage of the responses to the open version, a large proportion (42%) of the respondents answering the open version still gave responses that were not anticipated by the response options for the closed question. Next, relabelling the response options for the closed question appears to have changed the percentage of respondents endorsing 'security' (from 7.2% up to 20.3%) and 'accomplishment' (from 59.1% down to 51.0%). In the third place, whereas 14.9 per cent gave answers that were coded under 'work conditions' in the first survey, a total of only 7.7 per cent fall into the two derivative categories 'control over work' and 'working conditions' in the second survey. And fourth, the fact that the increase in percentage of respondents endorsing the closed question option 'security' goes up would seem to be incongruent with the hypothesis that the closed question format per se generates social desirability responses (discussed earlier in this chapter).

In trying to make sense of the differences between the two response distributions, Schuman and Presser suggest that the 'pay' responses to the open question equate to the 'pay' plus 'security' responses to the closed question. As evidence for this suggestion, they note that the 'pay' responses to the closed question do not correlate with levels of education, while the 'security' responses correlate negatively. It is relevant to point out, though, that the 'pay' responses accounted for 24.3 per cent of the responses to the open question while the 'pay' and 'security' responses accounted for a total of 83.5 per cent of the closed question responses (i.e. 9.2 per cent more).

Similarly, Schuman and Presser suggest that the 'satisfaction' responses to the open question equate to the closed question 'accomplishment' plus 'pleasant' responses. In support of this suggestion, they cite the results of a small scale follow-up study in which eighteen of twenty-six respondents, whose responses to the original open question had been coded 'satisfaction', endorsed 'accomplishment' when responding to the closed question. Again, it is relevant to note that whereas a total of 44.6 per cent of the open question respondents' answers were coded 'accomplishment', 'pleasant' or 'satisfaction', a total of 50.8 per cent of the responses to the closed question were accounted for by the 'accomplishment' and 'pleasant' response alternatives (i.e. 6.2 per cent more).

Finally, Schuman and Presser interpret the fact that the response options for the revised closed question picked up 15.7 per cent more respondents than the response options for the original closed question as support for the hypothesis

**Table 10.4 Comparison of response distributions for the revised response options for the closed question and the revised codes for the open question**

<table>
<thead>
<tr>
<th>Closed question:</th>
<th>Open question:</th>
</tr>
</thead>
<tbody>
<tr>
<td>This next question is on the subject of work.</td>
<td>This next question is on the subject of work.</td>
</tr>
<tr>
<td>People look for different things in a job, which of the following things would you most prefer in a job?</td>
<td>People look for different things in a job. What would you most prefer in a job? (Codes 1–5 given high priority.) —</td>
</tr>
</tbody>
</table>

**Answers coded:**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Closed question</th>
<th>Open question</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>Work that pays well</td>
<td>15.2%</td>
<td>16.7%</td>
</tr>
<tr>
<td>(2)</td>
<td>Work that gives a feeling of accomplishment</td>
<td>31.0%</td>
<td>14.5%</td>
</tr>
<tr>
<td>(3)</td>
<td>Work where there is not too much supervision &amp; you make most decisions</td>
<td>11.7%</td>
<td>14.5%</td>
</tr>
<tr>
<td>(4)</td>
<td>Work that is pleasant; other people are nice to work with</td>
<td>19.8%</td>
<td>7.6%</td>
</tr>
<tr>
<td>(5)</td>
<td>Work that is steady with little chance of being laid off</td>
<td>20.3%</td>
<td>1.6%</td>
</tr>
<tr>
<td>(6)</td>
<td>Physical working conditions</td>
<td>3.1%</td>
<td>2.3%</td>
</tr>
<tr>
<td>(7)</td>
<td>Short hours/free time</td>
<td>15.6%</td>
<td></td>
</tr>
<tr>
<td>(8)</td>
<td>Opportunity for promotion</td>
<td>1.0%</td>
<td></td>
</tr>
<tr>
<td>(9)</td>
<td>Security</td>
<td>7.6%</td>
<td></td>
</tr>
<tr>
<td>(10)</td>
<td>Control of work</td>
<td>4.6%</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>Pleasant work (sociability &amp; enjoyment responses)</td>
<td>4.6%</td>
<td></td>
</tr>
<tr>
<td>(12)</td>
<td>Benefits</td>
<td>2.3%</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.4%</td>
<td>3.6%</td>
<td></td>
</tr>
<tr>
<td>Don't know/no opinion</td>
<td>3.6%</td>
<td>7.0%</td>
<td></td>
</tr>
<tr>
<td>(n=1194)</td>
<td>100%</td>
<td>(n=1155)</td>
<td></td>
</tr>
<tr>
<td>(n=1194)</td>
<td>100%</td>
<td>(n=1155)</td>
<td></td>
</tr>
</tbody>
</table>

that the categories for the original version had not been properly grounded in responses to the initial open question. But this conclusion must be kept in perspective. It is significant that the revised response options still failed to encompass over 40 per cent of the responses to the corresponding open question. In addition, the fact the answers to the corresponding open question were still much more variable than the answers to the closed question should not be forgotten. Further, although Schuman and Presser are clearly of the view that open questions are essential for obtaining the frames of reference used by respondents for formulating appropriate response sets (Schuman and Presser, 1979a:704 and 710), much of their discussion of the differences between the response distributions for the revised questions focuses on interpreting the responses to the open question — especially the ‘pay’, ‘security’ and ‘satisfaction’ responses. In other words, one could just as well argue that the closed question responses were necessary to interpret responses to the open question as argue that the closed question was improved by grounding its response alternatives in the responses to an open question.

Problems associated with recording responses to closed questions

When a question calls for a respondent to choose answers from a list and then the respondent has not done so, the interviewer’s job is to explain to the respondent that choosing one answer from the list is the way to answer the question (called training the respondent) and to read the list of responses again. (Fowler and Mangione, 1990:59)

Fowler and Mangione continue by suggesting that there are two kinds of mistakes that interviewers can make. The interviewer can:

(a) wrongly record answers; or

(b) neglect to read all of the response options.

For example, if a question reads ‘How would you rate your schools — Very good, Good, Fair, Poor’, and a respondent answers, ‘Not very good’, the interviewer should read the whole question again rather than just saying, ‘Well would you say, Fair or Poor?’

Fowler and Mangione insist that closed questions should be asked exactly as they are written. They reason that if interviewers do this and respondents have trouble answering the question the researcher at least, knows that any differences in respondents’ answers have arisen out of the respondents’ interpretations of the original question rather than out of the interviewers’ behaviour, and that the measurement process has been consistent across respondents. For these reasons, Fowler and Mangione recommend that interviewers be programmed with standardised explanations to teach respondents to use the response options that are offered. They say that this can be achieved by:

(a) starting an interview with a general instruction:

Since many people have never been in an interview exactly like this, let me read you a paragraph that tells you a little about how it works. I am going to read you a set of questions exactly as they are worded so that every respondent in the survey is answering the same questions. You’ll be asked to answer two kinds of questions. In some cases, you’ll be asked to answer in your own words. For those questions, I will have to write down your answers word for word. In other cases, you will be given a list of answers and asked to choose the one that fits best. If at any time during the interview you are not clear about what is wanted, be sure to ask me. (Fowler and Mangione, 1990:51)

(b) Programming the interviewer to explain why it is necessary ask a question, even if the respondent has partly answered it already:

The next question is one you have dealt with to some extent. However, the way the interview works I need to have you answer each question specifically, so that we can compare the answers you give with the answers everyone else gives. Also, sometimes we find the answer is different to a specific question, even though it seems that the question has been answered before. So let me read the question as it is worded here, and I would like you to make sure we have it right. (Fowler and Mangione, 1990:51)

And programming the interviewer to merely repeat a question if the respondent appears to have difficulty with it:

I see what your problem is with the question. Even though these questions are carefully tested, sometimes we have one that is not quite clear to some people, or doesn’t quite fit everybody’s situation. Again though, the way the survey works, we need people’s best answers to the questions as they are written. That way we can compare your answers with other people’s. If we change the question for each respondent, we wouldn’t be able to analyze the answers. Let me read the question again, and you give me the best, most accurate answer you can, given the way it is written. (Fowler and Mangione, 1990:51–52)

or:

With this kind of question, answers are analyzed according to which of these alternatives people choose. I need to have you choose one of the specific answers so that we can compare your response with those others give. We know that in some cases none of the answers will fit the way you feel exactly; but other people will have that problem, too. The important thing is that we keep the question-and-answer process consistent across everybody, so we can see similarities and differences in the answers people give. (Fowler and Mangione, 1990:52)
The problem with instructions like these is that they clearly pressure respondents into answering in terms of the alternatives, whether or not these are appropriate for them.

Problems associated with interpreting responses to closed questions

It has been repeatedly stressed in previous chapters that a key assumption underlying the use of questions in social research is that the answers respondents give can be meaningfully compared with one another. This assumption is forcefully expressed by Crutchfield and Gordon:

In order that opinion poll answers be valid, respondents must interpret the question in the same way. Respondents who read different meanings into a question are, in effect, replying to different questions. When this happens the investigator cannot make a valid analysis of his results, for he has no assurance that the tabulations would be similar if all respondents had construed the questions in the same way. (Crutchfield and Gordon, 1947:1)

It was argued in chapters 3 to 6 that unless respondents have a clear understanding of what the question is about and are told what perspective to adopt when framing an answer, different respondents will answer the same question in quite different ways. This is as true for open questions as it is for closed questions. Even though it might be added that the response options which make up a closed question put some restrictions on the kinds of answers respondents can give, it must be appreciated that the implied framework will be, at best, partial in nature. Implicated response frameworks typically fail to provide respondents with information about what the researcher wants the information, what the researcher already knows or assumes, the perspectives that should be adopted, the required social generality of answers, and the standards of comparison which should be used if evaluations are called for. In other words, closed questions can be as deficient as open questions when it comes to the issues of the comparability and the interpretability of respondents’ answers. The real problem is not in the different formats (i.e. ‘open’ vs. ‘closed’ formats) per se but in the failure to properly specify a response framework for respondents.

The inescapable, overall conclusion that one is forced to reach is that it is dangerous to argue that open questions necessarily produce more valid results than closed questions. On the other side of the coin, it is equally dangerous to argue that closed questions are more efficient — especially if the time taken to develop appropriate sets of response options is taken into account.

An evaluation of three additional uses claimed for open questions

Three more virtues have been claimed for open questions, each of which merits brief discussion.

First, Kidder and Judd (1986:248), for example, argue that open questions allow respondents to convey the fine shades of their attitudes to their own satisfaction instead of forcing them to choose one of several statements that may all seem more or less unsatisfactory. For this reason, they suggest that open questions may be more motivating for respondents. This suggestion aside, it is just as likely that respondents are more likely to feel frustrated rather than liberated if the meaning of an open question is obscure — that is, if the respondents are not told what sorts of answers they should give.

Second, Schuman and Presser (1979a:711) suggest that open questions are useful when a set of meaningful alternatives is too large and complex to present to respondents (presumably by reading them out to the respondents because length would seem to be less of a problem if the options are presented on a card). Unfortunately, this suggestion runs contrary to everything we know about human memory. If respondents are asked to search their memories for the most appropriate answer out of a large range of possible answers, there is no guarantee that their memories will not fail or that they will not give the first answer that occurs to them that seems to answer the question.

Third, Schuman and Presser (1979a:711) suggest that it may be necessary to use open questions when the researcher has reason to suspect that rapidly shifting external events will affect answers (i.e., when the relevance or adequacy of preset sets of response options are likely to be affected). If one takes the line that researchers have to make the most of all the opportunities that are open to them, rather than the line that open questions are problem free, this is plainly a defensible suggestion.

Summary

Open and closed versions of the same questions have been found to typically generate quite different response distributions (e.g. Schuman and Presser 1979a) and it is not obvious which format produces the most valid data. The meaning of responses to open questions can be just as obscure as the meaning of responses to closed questions. This is especially true when respondents are allowed to wander from the topic, which seems to be almost an endemic problem with open questions (Campbell, 1945; Dohrenwend, 1965). It is also true
when respondents are not told what kinds of answers the researcher requires. In addition, answers to open questions are often less complete than answers to corresponding closed questions.

Methodologists who have considered the issues have tended to settle on the compromise position that a judicious mix of open and closed questions is best (e.g. Kahn and Cannell, 1957:158; Gallup, 1947). It has been widely taken for granted that qualitative, in-depth interviews should precede the formulation of fixed-choice questions so that response categories will reflect the respondents' worlds rather than the researchers' (e.g. Cicourel, 1982; Converse, 1984; Converse and Presser 1986). One might almost say that since the publication of Lazarsfeld's paper in 1944 every survey researcher has accepted the idea that open questions can play a useful part in interpreting responses to closed questions.

But the issues are not so easily settled. The observation that the distributions of answers to open questions often differ markedly from the distributions of answers to corresponding closed versions of the same questions is often taken as evidence that the response options for the closed questions must be inappropriate; but such an outcome can just as easily be the result of respondents having to guess what kinds of answers the researcher wants in response to open questions. The central issue is not which format produces the most valid responses but whether or not respondents know what kinds of answers they should give. And this is an issue that applies equally to both open and closed questions.

Researchers are often tempted to design and use questions that will allow them to measure the strength of respondent attributes instead of just noting their presence or absence. Working with non-dichotomous variables is a step toward greater precision which in turn allows the formulation and testing of more complex hypotheses using sophisticated statistical procedures (such as correlational and analysis of variance procedures).

One class of attributes in particular — that of 'attitudes' — has received a great deal of attention from methodologists. And a great many question devices have been invented to measure respondents' attitudes; see Sudman and Bradburn (1982:158–173) for a concise discussion of the most commonly used techniques. The list includes:

(a) Simple open ended questions (e.g. 'What are your feelings about X?')
(b) Simple rating scales which are presented as representing the attitude continuum underlying topics with instructions for the respondent to place ticks at the points on the scales which best indicate their attitudes toward the topics.
(c) The Michigan Survey Center's adaptation of Katz's (1944) 'ladder' and 'thermometer' rating devices, which require respondents to indicate where they would place themselves on pictorial representations of the attitude continuum. The ladder and thermometer presentations are attempts to make the rating task less abstract for respondents (figure 11.2 — 5 and 6, p. 157).