8

Sampling

Chapter outline

Introduction to survey research 184

Introduction to sampling 186

Sampling error 188

Types of probability sample 190

- Simple random sample 190
- Systematic sample 191
- Stratified random sampling 192
- Multi-stage cluster sampling 193

The qualities of a probability sample 195

Sample size 197

- Absolute and relative sample size 197
- Time and cost 198
- Non-response 199
- Heterogeneity of the population 200
- Kind of analysis 201

Types of non-probability sampling 201

- Convenience sampling 201
- Snowball sampling 202
- Quota sampling 203

Limits to generalization 205

Error in survey research 205

Key points 206

Questions for review 206

Chapter guide

This chapter and the three that follow it are very much concerned with principles and practices associated with social survey research. Sampling principles are not exclusively concerned with survey research; for example, they are relevant to the selection of documents for content analysis (see Chapter 13). However, in this chapter the emphasis will be on sampling in connection with the selection of people who would be asked questions by interview or questionnaire. The chapter explores:

- the role of sampling in relation to the overall process of doing survey research;
- the related ideas of generalization (also known as external validity) and of a representative sample; the latter allows the researcher to generalize findings from a sample to a population;
- the idea of a *probability sample*—that is, one in which a random selection process has been employed;
- the main types of probability sample: the simple random sample; the systematic sample; the stratified random sample; and the multi-stage cluster sample;
- the main issues involved in deciding on sample size;
- different types of non-probability sample, including quota sampling, which is widely used in market research and opinion polls;
- potential sources of error in survey research.

Introduction to survey research

This chapter is concerned with some important aspects of conducting a survey, but it presents only a partial picture, because there are many other steps. In this chapter we are concerned with the issues involved in selecting individuals for survey research, although the principles involved apply equally to other approaches to quantitative research, such as content analysis. Chapters 9, 10, and 11 deal with the data-collection aspects of conducting a survey, while Chapters 15 and 16 deal with issues to do with the analysis of data.

Figure 8.1 aims to outline the main steps involved in doing survey research. Initially, the survey will begin with general research issues that need to be investigated. These are gradually narrowed down so that they become research questions, which may take the form of hypotheses, but this need not necessarily be the case. The movement from research issues to research questions is likely to be the result of reading the literature relating to the issues, such as relevant theories and evidence (see Chapters 1 and 4).

Once the research questions have been formulated, the planning of the fieldwork can begin. In practice, decisions relating to sampling and the research instrument will overlap, but they are presented in Figure 8.1 as part of a sequence. The figure is meant to illustrate the main phases of a survey, and these different steps (other than those to do with sampling, which will be covered in this chapter) will be followed through in Chapters 9–11 and 15–16.

The survey researcher needs to decide what kind of population is suited to the investigation of the topic and also needs to formulate a research instrument and how it should be administered. By 'research instrument' is meant simply something like a **structured interview** schedule or a **self-completion questionnaire**. Moreover, there are several different ways of administering such instruments. Figure 8.2 outlines the main types that are likely to be encountered. Types 1 through 4 are covered in Chapter 9. Types 5 and 6 are covered in Chapter 10. Types 7 through 9 are covered in Chapter 28 in the context of the use of the Internet generally.

Figure 8.1

Steps in conducting a social survey

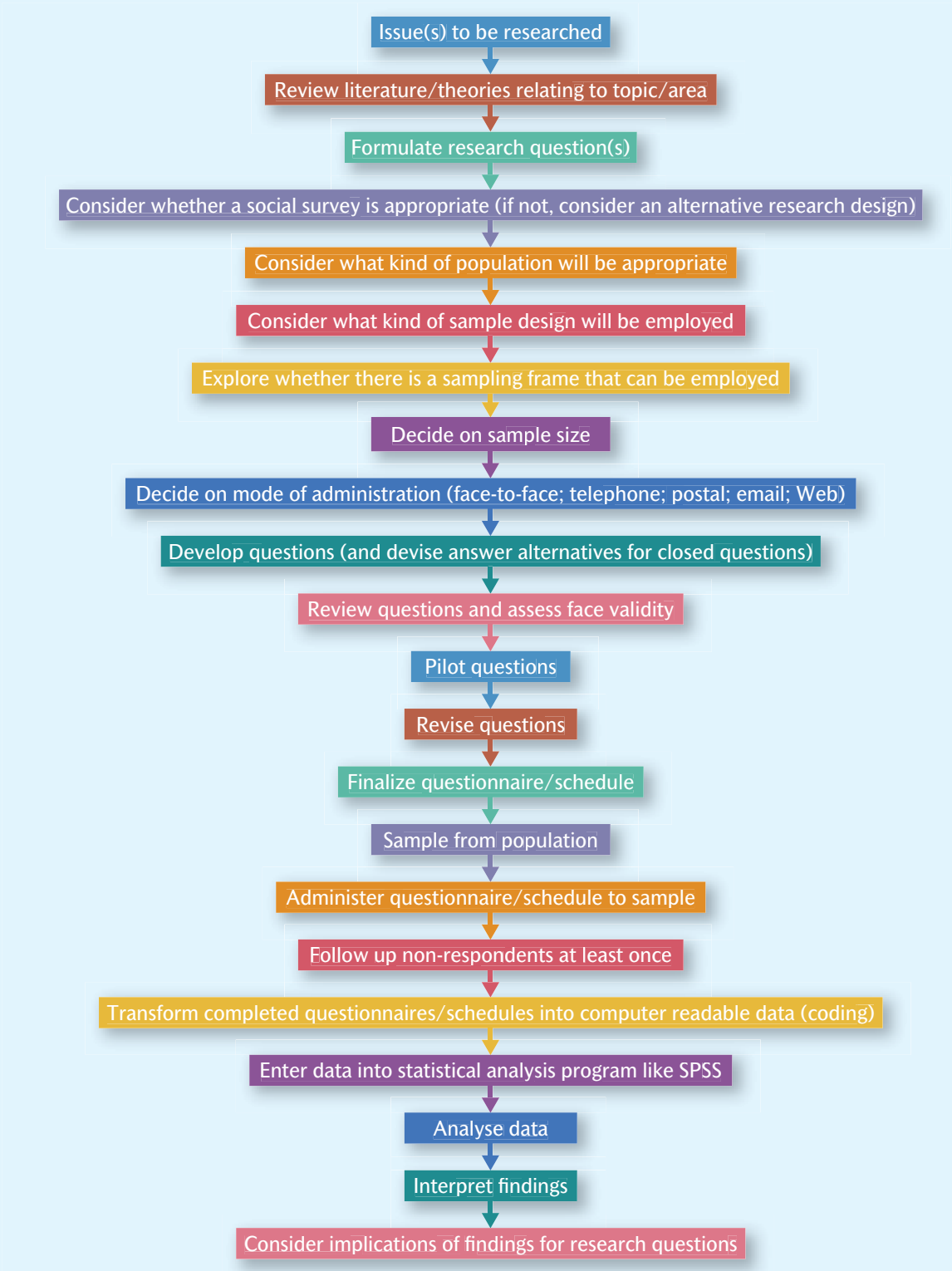
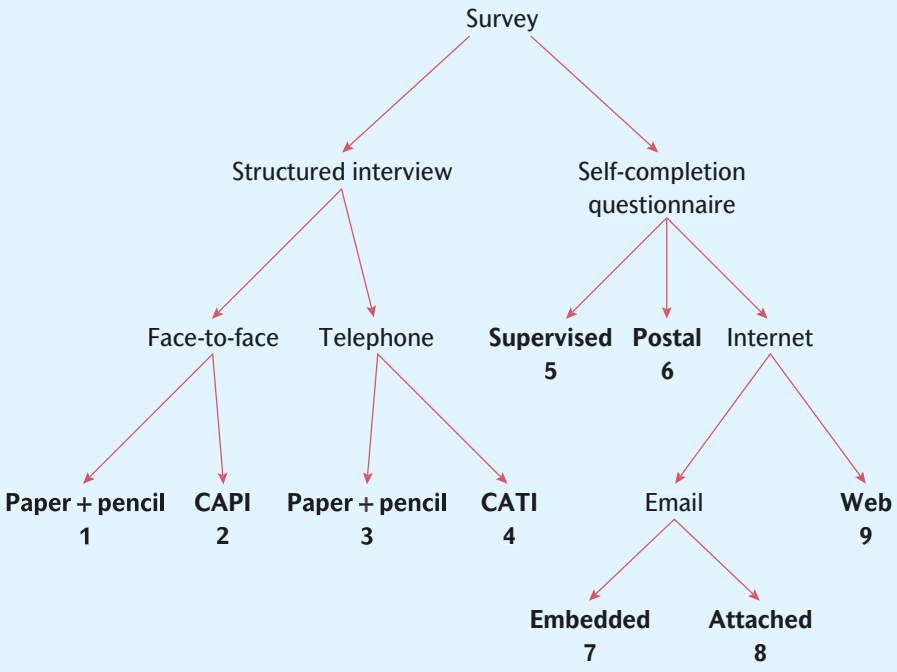


Figure 8.2

Main modes of administration of a survey



Notes: CAPI is computer-assisted personal interviewing; CATI is computer-assisted telephone interviewing.

Introduction to sampling

Many of the readers of this book will be university or college students. At some point in your stay at your university (I will use this term from now on to include colleges) you may have wondered about the attitudes of your fellow students to various matters, or about their behaviour in certain areas, or something about their backgrounds. If you were to decide to examine any or all of these three areas, you might consider conducting structured interviews or sending out questionnaires in order to find out about their behaviour, attitudes, and backgrounds. You will, of course, have to consider how best to design your interviews or questionnaires, and the issues that are involved in the decisions that need to be made about designing these research instruments and administering them will be the focus of Chapters 9–11. However, before getting to that point you are likely to be confronted with a problem. Let us say that your university is quite large and has around 9,000 students. It is extremely unlikely that you will have the time and resources to conduct a survey of all these students. It is unlikely that you would be able to send questionnaires to all 9,000 and even more unlikely that you would be able to interview all of them, since conducting survey research by interview is considerably more expensive and time consuming, all things being equal, than by postal questionnaire (see Chapter 10). It is almost certain that you will need to *sample* students from the total population of students in your university.

The need to sample is one that is almost invariably encountered in quantitative research. In this chapter I will be almost entirely concerned with matters relating to sampling in relation to social survey research involving data collection by structured interview or questionnaire. Other methods of quantitative research involve sampling considerations, as will be seen in Chapters 12 and 13, when we will examine structured observation and content analysis respectively. The principles of sampling involved are more or less identical in connection with

these other methods, but frequently other considerations come to the fore as well.

But will any old sample suffice? Would it be sufficient to locate yourself in a central position on your campus (if it has one) and then interview the students who come past you and whom you are in a position to interview? Alternatively, would it be sufficient to go around your student union asking people to be interviewed? Or again to send questionnaires to everyone on your course?

The answer, of course, depends on whether you want to be able to *generalize* your findings to the entire student body in your university. If you do, it is unlikely that any of the three sampling strategies proposed in the previous paragraph would provide you with a *representative sample* of all students in your university. In order to be able to generalize your findings from your sample to the population from which it was selected, the sample must be representative. See Key concept 8.1 for an explanation of key terms concerning sampling.

Key concept 8.1
Basic terms and concepts in sampling

- *Population*: basically, the universe of units from which the sample is to be selected. The term 'units' is employed because it is not necessarily people who are being sampled—the researcher may want to sample from a universe of nations, cities, regions, firms, etc. Finch and Hayes (1994), for example, based part of their research upon a random sample of wills. Their population, therefore, was a population of wills. Thus, 'population' has a much broader meaning than the everyday use of the term, whereby it tends to be associated with a nation's entire population.
- *Sample*: the segment of the population that is selected for investigation. It is a subset of the population. The method of selection may be based on a probability or a non-probability approach (see below).
- *Sampling frame*: the listing of all units in the population from which the sample will be selected.
- *Representative sample*: a sample that reflects the population accurately so that it is a microcosm of the population.
- *Sampling bias*: a distortion in the representativeness of the sample that arises when some members of the population (or more precisely the sampling frame) stand little or no chance of being selected for inclusion in the sample.
- *Probability sample*: a sample that has been selected using random selection so that each unit in the population has a known chance of being selected. It is generally assumed that a *representative sample* is more likely to be the outcome when this method of selection from the population is employed. The aim of probability sampling is to keep *sampling error* (see below) to a minimum.
- *Non-probability sample*: a sample that has not been selected using a random selection method. Essentially, this implies that some units in the population are more likely to be selected than others.
- *Sampling error*: error in the findings deriving from research due to the difference between a sample and the population from which it is selected. This may occur even though probability sampling has been employed.
- *Non-sampling error*: error in the findings deriving from research due to the differences between the population and the sample that arise either from deficiencies in the sampling approach, such as an inadequate sampling frame or **non-response** (see below), or from such problems as poor question wording, poor interviewing, or flawed processing of data.
- *Non-response*: a source of non-sampling error that is particularly likely to happen when individuals are being sampled. It occurs whenever some members of the sample refuse to cooperate, cannot be contacted, or for some reason cannot supply the required data (for example, because of mental incapacity).
- *Census*: the enumeration of an entire population. Thus, if data are collected in relation to all units in a population, rather than in relation to a sample of units of that population, the data are treated as census data. The phrase '*the* census' typically refers to the complete enumeration of all members of the population of a nation state—that is, a national census. This form of enumeration currently occurs once every ten years in the UK, although there is some uncertainty at the time of writing about whether another census will take place. However, in a statistical context, like the term *population*, the idea of a census has a broader meaning than this.

Why might the strategies for sampling students previously outlined be unlikely to produce a representative sample? There are various reasons, of which the following stand out.

- The first two approaches depend heavily upon the availability of students during the time or times that you search them out. Not all students are likely to be equally available at that time, so the sample will not reflect these students.
- They also depend on the students going to the locations. Not all students will necessarily pass the point where you locate yourself or go to the student union, or they may vary hugely in the frequency with which they do so. Their movements are likely to reflect such things as where their halls of residence or accommodation are situated, or where their departments are located, or their social habits. Again, to rely on these locations would mean missing out on students who do not frequent them.
- It is possible, not to say likely, that your decisions about which people to approach will be influenced by your judgements about how friendly or cooperative the people concerned are likely to be or by how comfortable you feel about interviewing students of the same (or opposite) gender to yourself, as well as by many other factors.
- The problem with the third strategy is that students on your course by definition take the same subject as each other and therefore will not be representative of all students in the university.

In other words, in the case of all of the three sampling approaches, your decisions about whom to sample are influenced too much by personal judgements, by prospective respondents' availability, or by your implicit criteria for inclusion. Such limitations mean that, in the language of survey sampling, your sample will be *biased*. A biased sample is one that does not represent the population from which the sample was selected. Sampling bias will occur if some members of the population have little or no chance of being selected for inclusion in the sample. As far as possible, bias should be removed from the selection of your sample. In fact, it is incredibly difficult to remove bias altogether and to derive a truly representative sample. What needs to be done is to ensure that steps are taken to keep bias to an absolute minimum.

Three sources of sampling bias can be identified (see Key concept 8.1 for an explanation of key terms).

1. *If a non-probability or non-random sampling method is used*. If the method used to select the sample is not random, there is a possibility that human judgement will affect the selection process, making some members of the population more likely to be selected than others. This source of bias can be eliminated through the use of probability/random sampling, the procedure for which is described below.
2. *If the sampling frame is inadequate*. If the sampling frame is not comprehensive or is inaccurate or suffers from some other kind of similar deficiency, the sample that is derived cannot represent the population, even if a random/probability sampling method is employed.
3. *If some sample members refuse to participate or cannot be contacted—in other words, if there is non-response*. The problem with non-response is that those who agree to participate may differ in various ways from those who do not agree to participate. Some of the differences may be significant to the research question or questions. If the data are available, it may be possible to check how far, when there is non-response, the resulting sample differs from the population. It is often possible to do this in terms of characteristics such as gender or age, or, in the case of something like a sample of university students, whether the sample's characteristics reflect the entire sample in terms of faculty membership. However, it is usually impossible to determine whether differences exist between the population and the sample after non-response in terms of 'deeper' factors, such as attitudes or patterns of behaviour.

Sampling error

In order to appreciate the significance of sampling error for achieving a representative sample, consider Figures 8.3–8.7. Imagine we have a population of 200 people and we want a sample of 50. Imagine as well that one of the variables of concern to us is whether people watch soap operas and that the population is equally divided

Figure 8.3

Watching soap operas in a population of 200

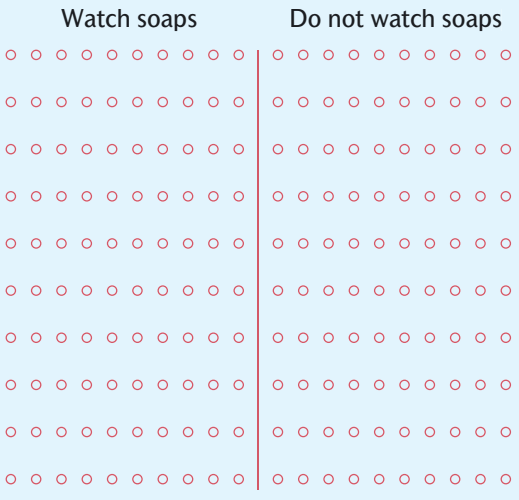


Figure 8.4

A sample with no sampling error

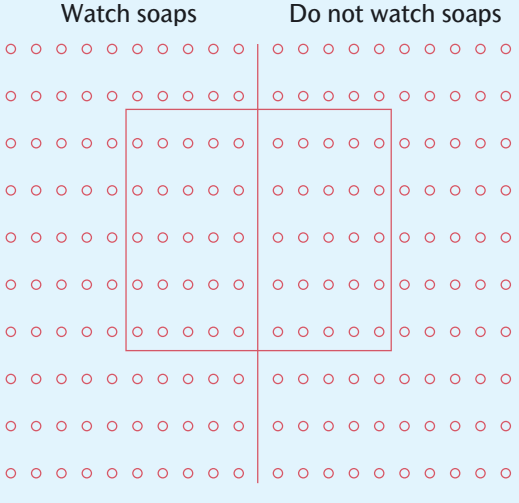


Figure 8.5

A sample with very little sampling error

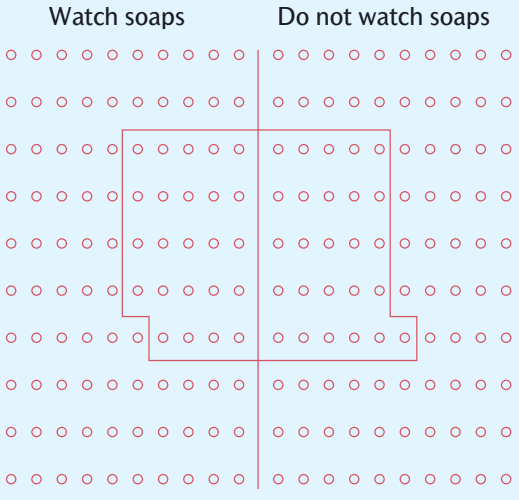
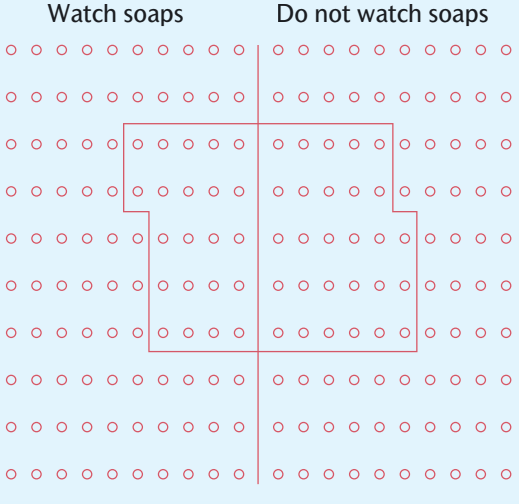


Figure 8.6

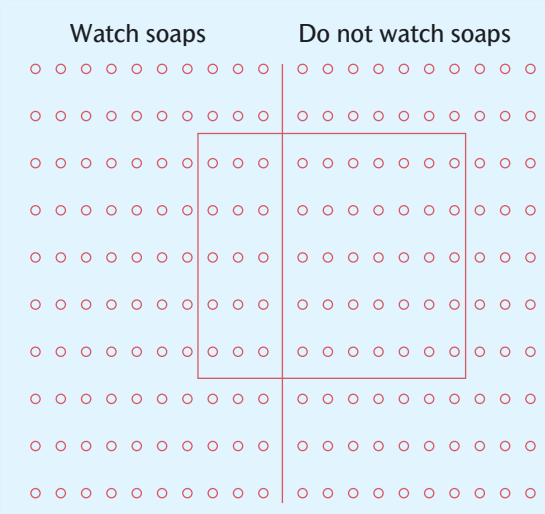
A sample with some sampling error



between those who do and those who do not. This split is represented by the vertical line that divides the population into two halves (Figure 8.3). If the sample is representative we would expect our sample of 50 to be equally split in terms of this variable (Figure 8.4). If there is a small amount of sampling error, so that we have one person too many who does not watch soap operas and one too few who does, it will look like Figure 8.5. In Figure 8.6 we see a rather more serious degree of over-representation of people who do not watch soaps. This

Figure 8.7

A sample with a lot of sampling error



time there are three too many who do not watch them and three too few who do. In Figure 8.7 we have a very serious over-representation of people who do not watch soaps, because there are 35 people in the sample who do not watch them, which is much larger than the 25 who should be in the sample.

It is important to appreciate that, as suggested above, probability sampling does not and cannot eliminate sampling error. Even with a well-crafted probability sample, a degree of sampling error is likely to creep in. However, probability sampling stands a better chance than non-probability sampling of keeping sampling error in check so that it does not end up looking like the outcome featured in Figure 8.7. Moreover, probability sampling allows the researcher to employ tests of statistical significance that permit inferences to be made about the sample from which the sample was selected. These will be addressed in Chapter 15.

Types of probability sample

Imagine that we are interested in levels of alcohol consumption among university students and the variables that relate to variation in levels of drinking. We might decide to conduct our research in a single nearby university. This means that our population will be all students in that university, which will in turn mean that we will be able to generalize our findings only to students of that university. We simply cannot assume that levels of alcohol consumption and their correlates will be the same in other universities. We might decide that we want our research to be conducted only on full-time students, so that part-time students are omitted. Imagine too that there are 9,000 full-time students in the university.

Simple random sample

The **simple random sample** is the most basic form of probability sample. With random sampling, each unit of the population has an equal probability of inclusion in the sample. Imagine that we decide that we have enough money to interview 450 students at the university. This means that the probability of inclusion in the sample is

$$\frac{450}{9{,}000}, \text{ i.e. } 1 \text{ in } 20$$

This is known as the *sampling fraction* and is expressed as

$$\frac{n}{N}$$

where n is the sample size and N is the population size.

The key steps in devising our simple random sample can be represented as follows.

1. Define the population. We have decided that this will be all full-time students at the university. This is our N and in this case is 9,000.
2. Select or devise a comprehensive sampling frame. It is likely that the university will have an office that keeps records of all students and that this will enable us to exclude those who do not meet our criteria for inclusion—i.e. part-time students.
3. Decide your sample size (n). We have decided that this will be 450.

4. List all the students in the population and assign them consecutive numbers from 1 to *N*. In our case, this will be 1 to 9,000.
5. Using a table of random numbers, or a computer program that can generate random numbers, select *n* (450) different random numbers that lie between 1 and *N* (9,000).
6. The students to which the *n* (450) random numbers refer constitute the sample.

Two points are striking about this process. First, there is almost no opportunity for human bias to manifest itself. Students would not be selected on such subjective criteria as whether they looked friendly and approachable. The selection of whom to interview is entirely mechanical. Second, the process is not dependent on the students' availability. They do not have to be walking in the interviewer's proximity to be included in the sample. The process of selection is done without their knowledge. It is not until they are contacted by an interviewer that they know that they are part of a social survey.

Step 5 mentions the possible use of a table of random numbers. These can be found in the appendices of many statistics books. The tables are made up of columns of five-digit numbers, such as:

09188
90045
73189
75768
54016
08358
28306
53840
91757
89415

The first thing to notice is that, since these are five-digit numbers and the maximum number that we can sample from is 9,000, which is a four-digit number, none of the random numbers seems appropriate, except for 09188 and 08358, although the former is larger than the largest possible number. The answer is that we should take just four digits in each number. Let us take the last four digits. This would yield the following:

9188
0045
3189
5768
4016
8358
8306
3840
1757
9415

However, two of the resulting numbers—9188 and 9415—exceed 9,000. We cannot have a student with either of these numbers assigned to him or her. The solution is simple: we ignore these numbers. This means that the student who has been assigned the number 45 will be the first to be included in the sample; the student who has been assigned the number 3189 will be next; the student who has been assigned the number 5768 will be next; and so on.

However, this somewhat tortuous procedure may be replaced in some circumstances by using a *systematic sampling* procedure (see next section) and more generally can be replaced by enlisting the computer for assistance (see Tips and skills 'Generating random numbers').

Systematic sample

A variation on the simple random sample is the **systematic sample**. With this kind of sample, you select units directly from the sampling frame—that is, without resorting to a table of random numbers.

We know that we are to select 1 student in 20. With a systematic sample, we would make a random start between 1 and 20 inclusive, possibly by using the last two digits in a table of random numbers. If we did this with the ten random numbers above, the first relevant one would be 54016, since it is the first one where the last two digits yield a number of 20 or below, in this case 16. This means that the sixteenth student on our sampling frame is the first to be in our sample. Thereafter, we take every twentieth student on the list. So the sequence will go:

16, 36, 56, 76, 96, 116, etc.

Tips and skills
Generating random numbers

The method for generating random numbers described in the text is what might be thought of as the classic approach. However, a far neater and quicker way is to generate random numbers on the computer. For example, the following website provides an online random generator which is very easy to use: **www.psychicscience.org/random.aspx** (accessed 9 August 2010).

If we want to select 450 cases from a population of 9,000, specify 450 after Generate, the digit 1 after random integers between and then 9000 after and. You will also need to specify from a drop-down menu 'with no repeats'. This means that no random number will be selected more than once. Then simply click on GO and the 450 random numbers will appear in a box below OUTPUT. You can then copy and paste the random numbers into a document.

This approach obviates the need to assign numbers to students' names and then to look up names of the students whose numbers have been drawn by the random selection process. It is important to ensure, however, that there is no inherent ordering of the sampling frame, since this may bias the resulting sample. If there is some ordering to the list, the best solution is to rearrange it.

Stratified random sampling

In our imaginary study of university students, one of the features that we might want our sample to exhibit is a proportional representation of the different faculties to which students are attached. It might be that the kind of discipline a student is studying is viewed as relevant to a wide range of attitudinal features that are relevant to the study of drinking. Generating a simple random sample or a systematic sample *might* yield such a representation, so that the proportion of humanities students in the sample is the same as that in the student population and so on. Thus, if there are 1,800 students in the humanities faculty, using our sampling fraction of 1 in 20, we would expect to have *90* students in our sample from this faculty. However, because of sampling error, it is unlikely that this will occur and that there will be a difference, so that there may be, say, 85 or 93 from this faculty.

Because it is almost certain that the university will include in its records the faculty in which students are based, or indeed may have separate sampling frames for each faculty, it will be possible to ensure that students are accurately represented in terms of their faculty membership. In the language of sampling, this means stratifying the population by a criterion (in this case, faculty membership) and selecting either a simple random sample or a systematic sample from each of the resulting strata. In the present example, if there are five faculties we would have five strata, with the numbers in each stratum being one-twentieth of the total for each faculty, as in Table 8.1, which also shows a hypothetical outcome of using a simple random sample, which results in a distribution of students across faculties that does not mirror the population all that well.

Table 8.1

The advantages of stratified sampling

Faculty	Population	Stratified sample	Hypothetical simple random or systematic sample
Humanities	1,800	90	85
Social sciences	1,200	60	70
Pure sciences	2,000	100	120
Applied sciences	1,800	90	84
Engineering	2,200	110	91
TOTAL	9,000	450	450

The advantage of stratified random sampling in a case like this is clear: it ensures that the resulting sample will be distributed in the same way as the population in terms of the stratifying criterion. If you use a simple random or systematic sampling approach, you *may* end up with a distribution like that of the stratified sample, but it is unlikely. Two points are relevant here. First, you can conduct stratified sampling sensibly only when it is relatively easy to identify and allocate units to strata. If it is not possible or it would be very difficult to do so, stratified sampling will not be feasible. Second, you can use more than one stratifying criterion. Thus, it may be that you would want to stratify by both faculty and gender or

faculty and whether students are undergraduates or postgraduates. If it is feasible to identify students in terms of these stratifying criteria, it is possible to use pairs of criteria or several criteria (such as faculty membership plus gender plus undergraduate/postgraduate).

Stratified sampling is really feasible only when the relevant information is available. In other words, when data are available that allow the ready identification of members of the population in terms of the stratifying criterion (or criteria), it is sensible to employ this sampling method. But it is unlikely to be economical if the identification of population members for stratification purposes entails a great deal of work because there is no available listing in terms of strata.

Student experience
Probability sampling for a student project

Joe Thompson describes the sampling procedure that he and the other members of his team used for their study of students living in halls of residence at the University of East Anglia as a **stratified random sample**. The following description suggests that they employed a systematic sampling approach for finding students within halls.

> Stratified random sampling was used to decide which halls of residence each member of the research team would go to and obtain questionnaire responses. This sampling method was the obvious choice as it meant there could be no fixing/bias to which halls the interviewee would go to and also maintained the representative nature of the research.
>
> The stratified random sampling method known as the 'random walk process' was used when conducting the interviews. Each member of the research group was assigned a number between 4 and 8 as a sampling fraction gap: I was assigned the number 7 and 'Coleman house block 1' as my accommodation block. This meant that, when conducting my interviews, I would go to Coleman house and knock on the 7th door, and then the 14th door, adding 7 each time, until I had completed five interviews. If I encountered a lack of response from the 6th door, I would return to the first flat but add one each time to avoid periodicity. This sampling method was determined by the principles of standardization, reliability, and validity.

To read more about Joe's research experiences, go to the Online Resource Centre that accompanies this book at:
www.oxfordtextbooks.co.uk/orc/brymansrm4e/

Multi-stage cluster sampling

In the example we have been dealing with, students to be interviewed are located in a single university. Interviewers will have to arrange their interviews with the sampled students, but, because they are all close together (even in a split-site university), they will not be involved in a lot of travel. However, imagine that we wanted a *national* sample of students. It is likely that interviewers would have to travel the length and breadth of the UK to interview the sampled students. This would add a great deal to the time and cost of doing the research. This kind of problem occurs whenever the aim is to interview a sample that is to be drawn from a widely dispersed population, such as a national population, or a large region, or even a large city.

One way in which it is possible to deal with this potential problem is to employ **cluster sampling**. With cluster sampling, the primary sampling unit (the first stage of the sampling procedure) is not the units of the population to be sampled but groupings of those units. It is the latter groupings or aggregations of population units that are known as *clusters*. Imagine that we want a nationally representative sample of 5,000 students. Using simple random or systematic sampling would yield a widely dispersed sample, which would result in a great deal of travel for interviewers. One solution might be to sample universities and then students from each of the sampled universities. A probability sampling method would need to be employed at each stage. Thus, we might randomly sample ten universities from the entire population of universities, thus yielding ten clusters, and we would then interview 500 randomly selected students at each of the ten universities.

Now imagine that the result of sampling ten universities gives the following list:

- Glasgow Caledonian
- Edinburgh
- Teesside
- Sheffield
- University College Swansea
- Leeds Metropolitan
- University of Ulster
- University College London
- Southampton
- Loughborough

This list is fine, but interviewers could still be involved in a great deal of travel, since the ten universities are quite a long way from each other. North American and Australian readers who examine this last comment by looking at a map of the United Kingdom may view the universities as in fact very close to each other!

One solution is likely to be to group all UK universities by standard region (see Research in focus 8.1 for an example of this kind of approach) and randomly to sample two standard regions. Five universities might then be sampled from each of the two lists of universities and then 500 students from each of the ten universities. Thus, there are separate stages:

- group UK universities by standard region and sample two regions;
- sample five universities from each of the two regions;
- sample 500 students from each of the ten universities.

Research in focus 8.1
An example of a multi-stage cluster sample

For their study of social class in modern Britain, Marshall et al. (1988: 288) designed a sample 'to achieve 2,000 interviews with a random selection of men aged 16–64 and women aged 16–59 who were not in full-time education'.

- Sampling parliamentary constituencies
  - — Parliamentary constituencies were ordered by standard region (there are eleven).
  - — Constituencies were allocated to one of three population density bands within standard regions.
  - — These subgroups were then reordered by political party voted to represent the constituency at the previous general election.
  - — These subgroups were then listed in ascending order of percentage in owner–occupation.
  - — 100 parliamentary constituencies were then sampled.
  - — Thus, parliamentary constituencies were stratified in terms of four variables: standard region; population density; political party voted for in last election; and percentage of owner–occupation.
- Sampling polling districts
  - — Two polling districts were chosen from each sampled constituency.
- Sampling individuals
  - — Nineteen addresses from each sampled polling district were systematically sampled.
  - — One person at each address was chosen according to a number of pre-defined rules.

In a sense, cluster sampling is always a multi-stage approach, because one always samples clusters first, and then something else—either further clusters or population units—is sampled.

Many examples of multi-stage cluster sampling entail stratification. We might, for example, want to stratify universities in terms of whether they are 'old' or 'new' universities—that is, those that received their charters after the 1991 White Paper for Higher Education, *Higher Education: A New Framework*. In each of the two regions, we would group universities along the old/new university criterion and then select two or three universities from each of the two strata per region.

Research in focus 8.1 provides an example of a multi-stage cluster sample. It entailed three stages: the sampling of parliamentary constituencies, the sampling of polling districts, and the sampling of individuals. In a way, there are four stages, because addresses are

sampled from polling districts and then individuals are sampled from each address. However, Marshall et al. (1988) present their sampling strategy as involving just three stages. Parliamentary constituencies were stratified by four criteria: standard region, population density, voting behaviour, and owner–occupation.

The advantage of multi-stage cluster sampling should be clear by now: it allows interviewers to be far more geographically concentrated than would be the case if a simple random or stratified sample were selected. The advantages of stratification can be capitalized upon because the clusters can be stratified in terms of strata. However, even when a very rigorous sampling strategy is employed, sampling error cannot be avoided, as the example in Research in focus 8.2 shows.

Research in focus 8.2
The 1992 British Crime Survey

The British Crime Survey (BCS) is a regular survey, funded by the Home Office, of a national sample drawn from the populations of England and Wales. The survey was conducted on eight occasions between 1982 and 2000 and has been conducted annually since 2001. In each instance, over 10,000 people have been interviewed. The main object of the survey is to glean information on respondents' experiences of being victims of crime. There is also a self-report component in which a selection of the sample are interviewed on their attitudes to crime and to report on crimes they have committed. Before 1992, the BCS used the electoral register as a sampling frame. Relying on a register of the electorate as a sampling frame is not without problems in spite of appearing robust: it omits any persons who are not registered, a problem that was exacerbated by the Community Charge (poll tax), which resulted in a significant amount of non-registration, as some people sought to avoid detection in order not to have to pay the tax. In 1992 the Postcode Address File was employed as a sampling frame and has been used since then. Its main advantage over the electoral register as a sampling frame is that it is updated more frequently. It is not perfect, because the homeless will not be accessible through it. The BCS sample itself is a stratified multi-stage cluster sample. The sampling procedure produced 13,117 residential addresses. Like most surveys, there was some non-response, with 23.3 per cent of the 13,117 addresses not resulting in a 'valid' interview. Just under half of these cases were the result of an outright refusal. In spite of the fact that the BCS is a rigorously selected and very large sample, an examination of the 1992 survey by Elliott and Ellingworth (1997) shows that there is some sampling error. By comparing the distribution of survey respondents with the 1991 census, they show that certain social groups are somewhat under-represented, most notably: owner–occupiers, households in which no car is owned, and male unemployed. However, Elliott and Ellingworth show that, as the level of property crime in postcode address sectors increases, the response rate (see Key concept 8.2) decreases. In other words, people who live in high-crime areas tend to be less likely to agree to be interviewed. How far this tendency affects the BCS data is difficult to determine, but the significance of this brief example is that, even when a sample of this quality is selected, the existence of sampling and non-sampling error cannot be discounted. The potential for a larger spread of errors when levels of sampling rigour fall short of a sample like that selected for the BCS is, therefore, considerable.

The qualities of a probability sample

The reason why probability sampling is such an important procedure in social survey research is that it is possible to make inferences from information about a random sample to the population from which it was selected. In other words, we can generalize findings derived from a sample to the population. This is not to say that we treat the population data and the sample data as the same. If we take the example of the level of alcohol consumption in our sample of 450 students, which we will treat as the number of units of alcohol consumed in

the previous seven days, we will know that the **mean** number of units consumed by the sample ($\bar{x}$) can be used to estimate the population mean (μ) but with known margins of error. The mean, or more properly the **arithmetic mean**, is the simple average.

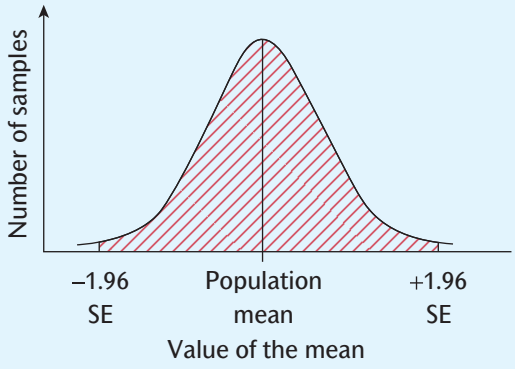
In order to address this point it is necessary to use some basic statistical ideas. These are presented in Tips and skills 'Generalizing from a random sample to the population' and can be skipped if just a broad idea of sampling procedures is required.

Tips and skills
Generalizing from a random sample to the population

Let us say that the sample mean is 9.7 units of alcohol consumed (the average amount of alcohol consumed in the previous seven days in the sample). A crucial consideration here is: how confident can we be that the mean level of alcohol consumption of 9.7 units is likely to be found in the population, even when probability sampling has been employed? If we take an infinite number of samples from a population, the sample estimates of the mean of the variable under consideration will vary in relation to the population mean. This variation will take the form of a bell-shaped curve known as a *normal distribution* (see Figure 8.8). The shape of the distribution implies that there is a clustering of sample means at or around the population mean. Half the sample means will be at or below the population mean; the other half will be at or above the population mean. As we move to the left (at or lower than the population mean) or the right (at or higher than the population mean), the curve tails off, implying fewer and fewer samples generating means that depart considerably from the population mean. The variation of sample means around the population mean is the *sampling error* and is measured using a statistic known as the **standard error of the mean**. This is an estimate of the amount that a sample mean is likely to differ from the population mean.

Figure 8.8

The distribution of sample means



Notes: 95 per cent of sample means will lie within the shaded area. SE = standard error of the mean.

This consideration is important because sampling theory tells us that 68 per cent of all sample means will lie between + or – 1 standard error from the population mean and that 95 per cent of all sample means will lie between + or – 1.96 standard errors from the population mean. It is this second calculation that is crucial, because it is at least implicitly employed by survey researchers when they report their statistical findings.

They typically employ 1.96 standard errors as the crucial criterion in how confident they can be in their findings. Essentially, the criterion implies that you can be 95 per cent certain that the population mean lies within + or – 1.96 sampling errors from the sample mean.

If a sample has been selected according to probability sampling principles, we know that we can be 95 per cent certain that the population mean will lie between the sample mean + or – 1.96 multiplied by the standard error of the mean. This is known as the *confidence interval*. If the mean level of alcohol consumption in the previous seven days in our sample of 450 students is 9.7 units and the standard error of the mean is 1.3, we can be 95 per cent certain that the population mean will lie between

$$9.7 + (1.96 \times 1.3)$$

and

$$9.7 - (1.96 \times 1.3)$$

i.e. between 12.248 and 7.152.

If the standard error was smaller, the range of possible values of the population mean would be narrower; if the standard error was larger, the range of possible values of the population mean would be wider.

If a stratified sample is selected, the standard error of the mean will be smaller because the variation between strata is essentially eliminated because the population will be accurately represented in the sample in terms of the stratification criterion or criteria employed. This consideration demonstrates the way in which stratification injects an extra increment of precision into the probability sampling process, since a possible source of sampling error is eliminated.

By contrast, a cluster sample without stratification exhibits a larger standard error of the mean than a comparable simple random sample. This occurs because a possible source of variability between students (i.e. membership of one university rather than another, which may affect levels of alcohol consumption) is disregarded. If, for example, some universities had a culture of heavy drinking in which a large number of students participated, and if these universities were not selected because of the procedure for selecting clusters, an important source of variability would have been omitted. It also implies that the sample mean would be on the low side, but that is another matter.

Sample size

As someone who is known as a teacher of research methods and a writer of books in this area, I often get asked questions about methodological issues. One question that is asked almost more than any other relates to the size of the sample—'how large should my sample be?' or 'is my sample large enough?' The decision about sample size is not a straightforward one: it depends on a number of considerations, and there is no one definitive answer. This is frequently a source of great disappointment to those who pose such questions. Moreover, most of the time decisions about sample size are affected by considerations of time and cost. Therefore, invariably decisions about sample size represent a compromise between the constraints of time and cost, the need for precision, and a variety of further considerations that will now be addressed.

Absolute and relative sample size

One of the most basic considerations, and one that is possibly the most surprising, is that, contrary to what you might have expected, it is the *absolute* size of a sample that is important not its *relative* size. This means that a national probability sample of 1,000 individuals in the UK has as much validity as a national probability sample of 1,000 individuals in the USA, even though the latter has a much larger population. It also means that increasing the size of a sample increases the precision of a sample.

This means that the 95 per cent confidence interval referred to in Tips and skills 'Generalizing from a random sample to the population' narrows. However, a large sample cannot *guarantee* precision, so that it is probably better to say that increasing the size of a sample increases the *likely* precision of a sample. This means that, as sample size increases, sampling error decreases. Therefore, an important component of any decision about sample size should be how much sampling error one is prepared to tolerate. The less sampling error one is prepared to tolerate, the larger a sample will need to be. Fowler (1993) warns against a simple acceptance of this criterion. He argues that in practice researchers do not base their decisions about sample size on a single estimate of a variable. Most survey research is concerned to generate a host of estimates—that is, of the variables that make up the research instrument that is administered. He also observes that it is not normal for survey researchers to be in a position to specify in advance 'a desired level of precision' (Fowler 1993: 34). Moreover, since sampling error will be only one component of any error entailed in an estimate, the notion of using a desired level of precision as a factor in a decision about sample size is not realistic. Instead, to the extent that this notion does enter into decisions about sample size, it usually does so in a general rather than in a calculated way.

Tips and skills
Sample size and probability sampling

As I have said in the text, the issue of sample size is the matter that most often concerns students and others. Basically, this is an area where size really does matter—the bigger the sample, the more representative it is likely to be (provided the sample is randomly selected), regardless of the size of the population from which it is drawn. However, when doing projects, students clearly need to do their research with very limited resources. You should try to find out from your department whether there are any guidelines about whether samples of a minimum size are expected. If there are no such guidelines, you will need to conduct your mini-survey in such a way as to maximize the number of interviews you can manage or the number of postal questionnaires you can send out, given the amount of time and resources available to you. Also, in many if not most cases, a truly random approach to sample selection may not be open to you. The crucial point is to be clear about and to justify what you have done. Explain the difficulties that you would have encountered in generating a random sample. Explain why you really could not include any more in your sample of respondents. But, above all, do not make claims about your sample that are not sustainable. Do not claim that it is representative or that you have a random sample when it is clearly not the case that either of these is true. In other words, be frank about what you have done. People will be much more inclined to accept an awareness of the limits of your sample design than claims about a sample that are patently false. Also, it may be that there are lots of good features about your sample—the range of people included, the good response rate, the high level of cooperation you received from the firm. Make sure you play up these positive features at the same time as being honest about its limitations.

Time and cost

Time and cost considerations become very relevant in this context. In the previous paragraph it is clearly being suggested that the larger the sample size the greater the precision (because the amount of sampling error will be less). However, by and large, up to a sample size of around 1,000, the gains in precision are noticeable as the sample size climbs from low figures of 50, 100, 150, and so on upwards. After a certain point, often in the region of 1,000, the sharp increases in precision become less pronounced, and, although it does not plateau, there is a slowing-down in the extent to which precision increases (and hence the extent to which the sample error of the mean declines). Considerations of sampling size are likely to be profoundly affected by matters of time and cost at such a juncture, since striving for smaller and smaller increments of precision becomes an increasingly uneconomic proposition. As Hazelrigg (2004: 85) succinctly puts it: 'The larger the size of the sample drawn from a population the more likely ($\bar{x}$) converges to μ; but the convergence occurs at a decelerating rate (which means that very large samples are decreasingly cost efficient).'

Non-response

However, considerations about sampling error do not end here. The problem of **non-response** should be borne in mind. Most sample surveys attract a certain amount of non-response. Thus, it is likely that only some members of our sample will agree to participate in the research. If it is our aim to ensure as far as possible that 450 students are interviewed and if we think that there may be a 20 per cent rate of non-response, it may be advisable to sample 540–50 individuals, on the grounds that approximately 90 will be non-respondents.

The issue of non-response, and in particular of refusal to participate, is of particular significance, because it has been suggested by some researchers that response rates to social surveys (see Key concept 8.2) are declining in many countries. This implies that there is a growing tendency towards people refusing to participate in social survey research. In 1973 an article in the American magazine *Business Week* carried an article ominously entitled 'The Public Clams up on Survey Takers'. The magazine asked survey companies about their experiences and found considerable concern about declining response rates. Similarly, in Britain, a report from a working party on the Market Research Society's Research and Development Committee in 1975 pointed to similar concerns among market research companies. However, an analysis of this issue by T. W. Smith (1995) suggests that, contrary to popular belief, there is no consistent evidence of such a decline. Moreover, Smith shows that it is difficult to disentangle general trends in response rates from such variables as the subject matter of the research, the type of respondent, and the level of effort expended on improving the number of respondents to individual surveys. However, an overview of non-response trends in the USA based on non-response rates for various continuous surveys suggests that there is a decline in the preparedness of households to participate in surveys (Groves et al. 2004). Further evidence comes from a study by Baruch (1999) of questionnaire-based articles published in 1975, 1985, and 1995 in five academic journals in the area of management studies. This article found an average (mean) response rate of 55.6 per cent, though with quite a large amount of variation around this average. The average response rate over the three years was 64.4 per cent in 1975, 55.7 per cent in 1985, and 48.4/52.2 per cent in 1995. Two percentages were provided for 1995 because the larger figure includes a journal that publishes a lot of research based on top managers, who tend to produce a poorer response rate. Response rates were found that were as low as 10 per cent and 15 per cent.

Key concept 8.2
What is a response rate?

The notion of a response rate is a common one in social survey research. When a social survey is conducted, whether by structured interview or by self-completion questionnaire, it is invariably the case that some people who are in the sample refuse to participate (referred to as non-response). The response rate is, therefore, the percentage of a sample that does, in fact, agree to participate. However, the calculation of a response rate is a little more complicated than this. First, not everyone who replies will be included: if a large number of questions are not answered by a respondent or if there are clear indications that he or she has not taken the interview or questionnaire seriously, it is better to employ only the number of *usable* interviews or questionnaires as the numerator. Similarly, it also tends to occur that not everyone in a sample turns out to be a suitable or appropriate respondent or can be contacted. Thus the response rate is calculated as follows:

$$\frac{\text{number of usable questionnaires}}{\text{total sample} - \text{unsuitable or uncontactable members of the sample}} \times 100$$

A further interesting issue in connection with non-response is that of how far researchers should go in order to boost their response rates. In Chapter 10, a number of steps that can be taken to improve response rates to postal questionnaires, which are particularly prone to poor response rates, are discussed. However, boosting response rates to interview-based surveys can prove expensive. Teitler et al. (2003) present a discussion of the steps taken to boost the response rate of a US sample that was hard to reach—namely, both parents of newly

born children, where most of the parents were not married. They found that, although there was evidence that increasing the response rate from an initial 68 per cent to 80 per cent meant that the final sample resembled more closely the population from which the sample had been taken, diminishing returns undoubtedly set in. In other words, the improvements in the characteristics of the sample necessitated a disproportionate outlay of resources. However, this is not to say that steps should not be taken to improve response rates. For example, following up respondents who do not initially respond to a postal questionnaire invariably results in an improved response rate at little additional cost. A study based on a survey of New Zealand residents by Brennan and Charbonneau (2009) provides unequivocal evidence of the improvement in response rate that can be achieved by at least two follow-up mailings to respondents to postal questionnaire surveys, which tend to achieve lower response rates than comparable interview-based surveys. A chocolate sent with the questionnaire helps too apparently!

As the previously mentioned study of response rates by Baruch (1999) suggests, there is wide variation in the response rates that social scientists achieve when they conduct surveys. It is difficult to arrive at clear indications of what is expected from a response rate. Baruch's study focused on research in business organizations, and, as he notes, when top managers are the focus of a survey, the response rate tends to be noticeably lower. In the survey component of the Cultural Capital and Social Exclusion (CCSE) project referred to in Research in focus 2.9, the initial main sample constituted a 53 per cent response rate (Bennett et al. 2009). The researchers decided to supplement the initial sample in various ways, one of which was to have an ethnic boost sample, in large part because the main sample did not include sufficient numbers of ethnic-minority members. However, the response rate from the ethnic boost sample was substantially below that achieved for the main sample. The researchers write: 'In general, ethnic boosts tend to have lower response rates than cross-sectional surveys' (Thomson 2004: 10). There is a sense, then, that what might be anticipated to be a reasonable response rate varies according to the type of sample and the topics covered by the interview or questionnaire. While it is obviously desirable to do one's best to maximize a response rate, it is also important to be open about the limitations of a low response rate in terms of the likelihood that findings will be biased. In the future, it seems likely that, given that there are likely to be limits on the degree to which a survey researcher can boost a response rate, more and more effort will go into refining ways of estimating and correcting for anticipated biases in findings (Groves 2006).

Research in focus 8.3
The problem of non-response

In December 2006 an article in *The Times* reported that a study of the weight of British children had been hindered because many families declined to participate. The study was commissioned by the Department of Health and found that, for example, among those aged 10 or 11, 14 per cent were overweight and 17 per cent were obese. However, *The Times* writer notes that a report compiled by the Department of Health on the research suggests that such figures are 'likely systematically to underestimate the prevalence of overweight and obesity' (quoted in Hawkes 2006: 24). The reason for this bias in the statistics is that parents were able to refuse to let their children participate, and those whose children were heavier were more likely to do so. As a result, the sample was biased towards those who were less heavy. The authors of the report drew the inference about sampling bias because they noted that more children were recorded as obese in areas where there was a poorer response rate.

Heterogeneity of the population

Yet another consideration is the homogeneity and heterogeneity of the population from which the sample is to be taken. When a population is very heterogeneous, like a whole country or city, a larger sample will be needed to reflect the varied population. When it is relatively homogeneous, such as a population of students or of members of an occupation, the amount of variation is less and therefore the sample can be smaller. The implication of this is that, the greater the heterogeneity of a population, the larger a sample will need to be.

Kind of analysis

Finally, researchers should bear in mind the *kind of analysis* they intend to undertake. A case in point here is the **contingency table**. A contingency table shows the relationship between two variables in tabular form. It shows how variation in one variable relates to variation in another variable. To understand this point, consider the basic structure of a table in the study by Marshall et al. (1988) of social class in Britain. This research was referred to in Research in focus 8.1. The table is based on the 589 cohabiting couples (1,178 people) of the sample in which both partners are employed in paid work. The authors aim to show in the table how far couples are of the same or a different social class in terms of Goldthorpe's seven-category scheme for classifying social class. The result is a table in which, because each variable comprises 7 categories, there are 49 **cells** in the table (i.e. 7×7). In order for there to be an adequate number of cases in each cell, a fairly large sample was required. Imagine that Marshall et al. had conducted a survey on a much smaller sample in which they ended up with just 150 couples. If the same kind of analysis as Marshall et al. carried out was conducted, it would be found that these 150 couples would be very dispersed across the 49 cells of the table. It is likely that many of the cells would be empty or would have very small numbers in them, which would make it difficult to make inferences about what the table showed. In fact, quite a lot of the cells in the actual table in Marshall et al. have very small numbers in them (8 cells contain 1 or 0). This problem would have been even more pronounced if they had ended up with a much smaller sample of couples. Consequently, considerations of sample size should be sensitive to the kinds of analysis that will be subsequently required, such as the issue of the number of cells in a table. In a case such as this, a larger sample will be necessitated by the nature of the analysis to be conducted as well as the nature of the variables in question.

Types of non-probability sampling

The term 'non-probability sampling' is essentially an umbrella term to capture all forms of sampling that are not conducted according to the canons of probability sampling outlined above. It is not surprising, therefore, that the term covers a wide range of different types of sampling strategy, at least one of which—the **quota sample**—is claimed by some practitioners to be almost as good as a probability sample. In this section we will cover three main types of non-probability sample: the **convenience sample**; the **snowball sample**; and the quota sample.

Convenience sampling

A convenience sample is one that is simply available to the researcher by virtue of its accessibility. Imagine that a researcher who teaches education at a university is interested in the kinds of features that teachers look for in their headmasters. The researcher might administer a questionnaire to several classes of students, all of whom are teachers taking a part-time master's degree in education. The chances are that the researcher will receive all or almost all of the questionnaires back, so that there will be a good response rate. The findings may prove quite interesting, but the problem with such a sampling strategy is that it is impossible to generalize the findings, because we do not know of what population this sample is representative. They are simply a group of teachers who are available to the researcher. They are almost certainly not representative of teachers as a whole—the very fact they are taking this degree programme marks them off as different from teachers in general.

This is not to suggest that convenience samples should never be used. Let us say that our lecturer/researcher is developing a battery of questions that are designed to measure the leadership preferences of teachers. It is highly desirable to pilot such a research instrument before using it in an investigation, and administering it to a group that is not a part of the main study may be a legitimate way of carrying out some preliminary analysis of such issues as whether respondents tend to answer in identical ways to a question, or whether one question is often omitted when teachers respond to it. In other words, for this kind of purpose, a convenience sample may be acceptable though not ideal. A second kind of context in which it may be at least fairly acceptable to use a convenience sample is when the chance presents itself to gather data from a convenience sample and it represents too good an opportunity to miss. The data will not allow definitive findings to be generated, because of the

problem of generalization, but they could provide a springboard for further research or allow links to be forged with existing findings in an area.

It also perhaps ought to be recognized that convenience sampling probably plays a more prominent role than is sometimes supposed. Certainly, in the field of organization studies it has been noted that convenience samples are very common and indeed are more prominent than are samples based on probability sampling (Bryman 1989a: 113–14). Social research is also frequently based on convenience sampling. Research in focus 8.4 contains an example of the use of convenience samples in social research. Probability sampling involves a lot of preparation, so that it is frequently avoided because of the difficulty and costs involved.

Research in focus 8.4
A convenience sample

Miller et al. (1998) were interested in theories concerning the role of shopping in relation to the construction of identity in modern society. Since many discussions of this issue have been concerned with shopping centres (malls), they undertook a study that combined quantitative and qualitative research methods in order to explore the views of shoppers at two London shopping centres: Brent Cross and Wood Green. One phase of the research entailed structured interviews with shoppers leaving the centres. The interviews were conducted mainly during weekdays in June and July 1994. Shoppers were chiefly questioned as they left the main exits, though some questioning at minor exits also took place. The authors tell us: 'We did not attempt to secure a quota [see below] or random sample but asked every person who passed by, and who did not obviously look in the other direction or change their path, to complete a questionnaire' (Miller et al. 1998: 55). Such a sampling strategy produces a convenience sample because only people who are visiting the centre and who are therefore self-selected by virtue of their happening to choose to shop at these times can be interviewed.

Snowball sampling

In certain respects, snowball sampling is a form of convenience sample, but it is worth distinguishing because it has attracted quite a lot of attention over the years. With this approach to sampling, the researcher makes initial contact with a small group of people who are relevant to the research topic and then uses these to establish contacts with others. I used an approach like this to create a sample of British visitors to Disney theme parks (Bryman 1999).

Research in focus 8.5 describes the generation of a snowball sample of marijuana-users for what is often regarded as a classic study of drug use. Becker's comment on this method of creating a snowball sample is interesting: 'The sample is, of course, in no sense

Research in focus 8.5
A snowball sample: Becker's study of marijuana-users

In an article first published in 1953, Becker (1963) reports on how he generated a sample of marijuana-users. He writes:

> I conducted fifty interviews with marijuana users. I had been a professional dance musician for some years when I conducted this study and my first interviews were with people I had met in the music business. I asked them to put me in contact with other users who would be willing to discuss their experiences with me. . . . Although in the end half of the fifty interviews were conducted with musicians, the other half covered a wide range of people, including laborers, machinists, and people in the professions. (Becker 1963: 45–6)

"random"; it would not be possible to draw a random sample, since no one knows the nature of the universe from which it would have to be drawn' (Becker 1963: 46). What Becker is essentially saying here (and the same point applies to my study of Disney theme park visitors) is that there is no accessible sampling frame for the population from which the sample is to be taken and that the difficulty of creating such a sampling frame means that a snowball sampling approach is the only feasible one. Moreover, even if one could create a sampling frame of marijuana-users or of British visitors to Disney theme parks, it would almost certainly be inaccurate straight away, because this is a shifting population. People will constantly be becoming and ceasing to be marijuana-users, while new theme park visitors are arriving all the time.

The problem with snowball sampling is that it is very unlikely that the sample will be representative of the population, though, as I have just suggested, the very notion of a population may be problematic in some circumstances. However, by and large, snowball sampling is used not within a quantitative research strategy, but within a qualitative one: both Becker's and my study were carried out within a qualitative research framework. Concerns about external validity and the ability to generalize do not loom as large within a qualitative research strategy as they do in a quantitative research one (see Chapter 17). In qualitative research, the orientation to sampling is more likely to be guided by a preference for *theoretical sampling* than with the kind of statistical sampling that has been the focus of this chapter (see Key concept 18.3). There is a much better 'fit' between snowball sampling and the theoretical sampling strategy of qualitative research than with the statistical sampling approach of quantitative research. This is not to suggest that snowball sampling is entirely irrelevant to quantitative research: when the researcher needs to focus upon or to reflect relationships between people, tracing connections through snowball sampling may be a better approach than conventional probability sampling (Coleman 1958).

Quota sampling

Quota sampling is comparatively rarely employed in academic social research, but is used intensively in commercial research, such as market research and political opinion polling. The aim of quota sampling is to produce a sample that reflects a population in terms of the relative proportions of people in different categories, such as gender, ethnicity, age groups, socio-economic groups, and region of residence, and in combinations of these categories. However, unlike a stratified sample, the sampling of individuals is not carried out randomly, since the final selection of people is left to the interviewer. Information about the stratification of the UK population or about certain regions can be obtained from sources like the census and from surveys based on probability samples such as the General Household Survey, British Social Attitudes, and the British Household Panel Survey.

Once the categories and the number of people to be interviewed within each category (known as *quotas*) have been decided upon, it is then the job of interviewers to select people who fit these categories. The quotas will typically be interrelated. In a manner similar to stratified sampling, the population may be divided into strata in terms of, for example, gender, social class, age, and ethnicity. Census data might be used to identify the number of people who should be in each subgroup. The numbers to be interviewed in each subgroup will reflect the population. Each interviewer will probably seek out individuals who fit several subgroup quotas. Accordingly, an interviewer may know that among the various subgroups of people he or she must find, and interview, five Asian, 25–34-year-old, lower-middle-class females in the area in which the interviewer has been asked to work (say, the Wirral). The interviewer usually asks people who are available to him or her about their characteristics (though gender will presumably be self-evident) in order to determine their suitability for a particular subgroup. Once a subgroup quota (or a combination of subgroup quotas) has been achieved, the interviewer will no longer be concerned to locate individuals for that subgroup.

The choice of respondents is left to the interviewer, subject to the requirement of all quotas being filled, usually within a certain time period. Those of you who have ever been approached on the street by a person toting a clipboard and interview schedule and have been asked about your age, occupation, and so on, before being asked a series of questions about a product or whatever, have almost certainly encountered an interviewer with a quota sample to fill. Sometimes, he or she will decide not to interview you because you do not meet the criteria required to fill a quota. This may be due to a quota already having been filled or to the criteria for exclusion meaning that a person with a certain characteristic you possess is not required.

A number of criticisms are frequently levelled at quota samples.

- Because the choice of respondent is left to the interviewer, the proponents of probability sampling

argue that a quota sample cannot be representative. It may accurately reflect the population in terms of superficial characteristics, as defined by the quotas. However, in their choice of people to approach, interviewers may be unduly influenced by their perceptions of how friendly people are or by whether the people make eye contact with the interviewer (unlike most of us, who look at the ground and shuffle past as quickly as possible because we do not want to be bothered in our leisure time).

- People who are in an interviewer's vicinity at the times he or she conducts interviews, and are therefore available to be approached, may not be typical. There is a risk, for example, that people in full-time paid work may be under-represented and that those who are included in the sample are not typical.
- The interviewer is likely to make judgements about certain characteristics in deciding whether to approach a person, in particular, judgements about age. Those judgements will sometimes be incorrect—for example, when someone who is eligible to be interviewed, because a quota that he or she fits is unfilled, is not approached because the interviewer makes an incorrect judgement (for example, that the person is older than he or she looks). In such a case, a possible element of bias is being introduced.
- It has also been argued that the widespread use of social class as a quota control can introduce difficulties, because of the problem of ensuring that interviewees are properly assigned to class groupings (Moser and Kalton 1971).
- It is not permissible to calculate a standard error of the mean from a quota sample, because the non-random method of selection makes it impossible to calculate the range of possible values of a population.

All this makes the quota sample look a poor bet, and there is no doubt that it is not favoured by academic social researchers. It does have some arguments in its favour, however.

- It is undoubtedly cheaper and quicker than an interview survey on a comparable probability sample. For example, interviewers do not have to spend a lot of time travelling between interviews.
- Interviewers do not have to keep calling back on people who were not available at the time they were first approached.
- Because calling back is not required, a quota sample is easier to manage. It is not necessary to keep track of people who need to be recontacted or to keep track of refusals. Refusals occur, of course, but it is not necessary (and indeed it is not possible) to keep a record of which respondents declined to participate.
- When speed is of the essence, a quota sample is invaluable when compared to the more cumbersome probability sample. Newspapers frequently need to know how a national sample of voters feel about a certain topic or how they intend to vote at that time. Alternatively, if there is a sudden major news event, such as a terrorist incident like the London bombs of July 2005, the news media may seek a more or less instant picture of the nation's views about personal security or people's responses more generally. Again, a quota sample will be much faster.
- As with convenience sampling, it is useful for conducting development work on new measures or on research instruments. It can also be usefully employed in relation to exploratory work from which new theoretical ideas might be generated.
- Although the standard error of the mean should not be computed for a quota sample, it frequently is. As Moser and Kalton (1971) observe, some writers argue that the use of a non-random method in quota sampling should not act as a barrier to such a computation because its significance as a source of error is small when compared to other errors that may arise in surveys (see Figure 8.9). However, they go on to argue that at least with random sampling the researcher can calculate the amount of sampling error and does not have to be concerned about its potential impact.

There is some evidence to suggest that, when compared to random samples, quota samples often result in biases. They under-represent people in lower social strata, people who work in the private sector and manufacturing, and people at the extremes of income, and they over-represent women in households with children and people from larger households. On the other hand, it has to be acknowledged that probability samples are often biased too—for example, it is often suggested that they under-represent men and those in employment (Marsh and Scarbrough 1990; Butcher 1994).

Limits to generalization

One point that is often not fully appreciated is that, even when a sample has been selected using probability sampling, any findings can be generalized only to the population from which that sample was taken. This is an obvious point, but it is easy to think that findings from a study have some kind of broader applicability. If we take our imaginary study of alcohol consumption among students at a university, any findings could be generalized only to that university. In other words, you should be very cautious about generalizing to students at other universities. There are many factors that may imply that the level of alcohol consumption is higher (or lower) than among university students as a whole. There may be a higher (or lower) concentration of pubs in the university's vicinity, there may be more (or fewer) bars on the campus, there may be more (or less) of a culture of drinking at this university, or the university may recruit a higher (or lower) proportion of students with disposable income. There may be many other factors too. Similarly, we should be cautious of overgeneralizing in terms of locality. Lunt and Livingstone's (1992: 173) study of consumption habits was based on a postal questionnaire sent to '241 people living in or around Oxford during September 1989'. While the authors' findings represent a fascinating insight into modern consumption patterns, we should be cautious about assuming that they can be generalized beyond the confines of Oxford and its environs.

There could even be a further limit to generalization that is implied by the Lunt and Livingstone (1992) sample. They write that the research was conducted in September 1989. One issue that is rarely discussed in this context and that is almost impossible to assess is whether there is a time limit on the findings that are generated. Quite aside from the fact that we need to appreciate that the findings cannot (or at least should not) be generalized beyond the Oxford area, is there a point at which we have to say, 'well, those findings applied to the Oxford area then but things have changed and we can no longer assume that they apply to that or any other locality'? We are, after all, used to thinking that things have changed when there has been some kind of prominent change. To take a simple example: no one would be prepared to assume that the findings of a study in 1980 of university students' budgeting and personal finance habits would apply to students in the early twenty-first century. Quite aside from changes that might have occurred naturally, the erosion and virtual dismantling of the student grant system has changed the ways students finance their education, including perhaps a greater reliance on part-time work (Lucas 1997), a greater reliance on parents, and the use of loans. But, even when there is no definable or recognizable source of relevant change of this kind, there is none the less the possibility (or even likelihood) that findings are temporally specific. Such an issue is impossible to resolve without further research (Bryman 1989b).

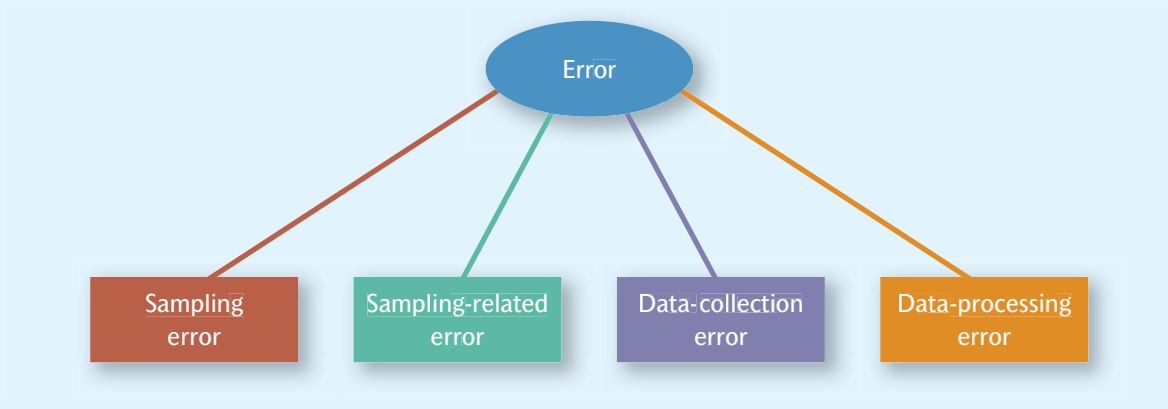
Error in survey research

We can think of 'error', a term that has been employed on a number of occasions, as being made up of four main factors (Figure 8.9).

1. *Sampling error*. See Key concept 8.1 for a definition. This kind of error arises because it is extremely unlikely that one will end up with a truly representative sample, even when probability sampling is employed.
2. We can distinguish what might be thought of as *sampling-related error*. This is error that is subsumed under the category *non-sampling error* (see Key concept 8.1) but that arises from activities or events that are related to the sampling process and that are connected with the issue of generalizability or external validity of findings. Examples are an inaccurate sampling frame and non-response.
3. There is also error that is connected with the implementation of the research process. We might call this *data-collection error*. This source of error includes such factors as: poor question wording in self-completion questionnaires or structured interviews; poor interviewing techniques; and flaws in the administration of research instruments.

Figure 8.9

Four sources of error in social survey research



4. Finally, there is *data-processing error*. This arises from faulty management of data, in particular, errors in the *coding* of answers.

The third and fourth sources of error relate to factors that are not associated with sampling and instead relate much more closely to concerns about the validity of measurement, which was addressed in Chapter 7. However, the kinds of steps that need to be taken to keep these sources of error to a minimum in the context of social survey research will be addressed in Chapters 9–11.

Key points

- Probability sampling is a mechanism for reducing bias in the selection of samples.
- Ensure you become familiar with key technical terms in the literature on sampling such as: representative sample; random sample; non-response; population; sampling error; etc.
- Randomly selected samples are important because they permit generalizations to the population and because they have certain known qualities.
- Sampling error decreases as sample size increases.
- Quota samples can provide reasonable alternatives to random samples, but they suffer from some deficiencies.
- Convenience samples may provide interesting data, but it is crucial to be aware of their limitations in terms of generalizability.
- Sampling and sampling-related error are just two sources of error in social survey research.

Questions for review

- What do each of the following terms mean: population; probability sampling; non-probability sampling; sampling frame; representative sample; and sampling and non-sampling error?
- What are the goals of sampling?
- What are the main areas of potential bias in sampling?

Sampling error

- What is the significance of sampling error for achieving a representative sample?

Types of probability sample

- What is probability sampling and why is it important?
- What are the main types of probability sample?
- How far does a stratified random sample offer greater precision than a simple random or systematic sample?
- If you were conducting an interview survey of around 500 people in Manchester, what type of probability sample would you choose and why?
- A researcher positions herself on a street corner and asks 1 person in 5 who walks by to be interviewed. She continues doing this until she has a sample of 250. How likely is she to achieve a representative sample?

The qualities of a probability sample

- A researcher is interested in levels of job satisfaction among manual workers in a firm that is undergoing change. The firm has 1,200 manual workers. The researcher selects a simple random sample of 10 per cent of the population. He measures job satisfaction on a Likert scale comprising ten items. A high level of satisfaction is scored 5 and a low level is scored 1. The mean job satisfaction score is 34.3. The standard error of the mean is 8.58. What is the 95 per cent confidence interval?

Sample size

- What factors would you take into account in deciding how large your sample should be when devising a probability sample?
- What is non-response and why is it important to the question of whether you will end up with a representative sample?

Types of non-probability sample

- Are non-probability samples useless?
- In what circumstances might you employ snowball sampling?
- 'Quota samples are not true random samples, but in terms of generating a representative sample there is little difference between them, and this accounts for their widespread use in market research and opinion polling.' Discuss.

Limits to generalization

- 'The problem of generalization to a population is not just to do with the matter of getting a representative sample.' Discuss.

Error in survey research

- 'Non-sampling error, as its name implies, is concerned with sources of error that are not part of the sampling process.' Discuss.

Online Resource Centre

www.oxfordtextbooks.co.uk/orc/brymansrm4e/

Visit the Online Resource Centre that accompanies this book to enrich your understanding of sampling. Consult web links, test yourself using multiple choice questions, and gain further guidance and inspiration from the Student Researcher's Toolkit.

9

Structured interviewing

Chapter outline

Introduction 209

The structured interview 209

- Reducing error due to interviewer variability 210
- Accuracy and ease of data processing 211
- Other types of interview 212

Interview contexts 213

- More than one interviewee 213
- More than one interviewer 214
- In person or by telephone? 214
- Computer-assisted interviewing 216

Conducting interviews 217

- Know the schedule 217
- Introducing the research 217
- Rapport 218
- Asking questions 219
- Recording answers 219
- Clear instructions 219
- Question order 220
- Probing 223
- Prompting 224
- Leaving the interview 225
- Training and supervision 225

Problems with structured interviewing 227

- Characteristics of interviewers 227
- Response sets 227
- The problem of meaning 228
- The feminist critique 228

Key points 229

Questions for review 230

Chapter guide

The structured interview is one of a variety of forms of research interview, but it is the one that is most commonly employed in survey research. The goal of the structured interview is for the interviewing of respondents to be standardized so that differences between interviews in any research project are minimized. As a result, there are many guidelines about how structured interviewing should be carried out so that variation in the conduct of interviews is small. The chapter explores:

- the reasons why the structured interview is a prominent research method in survey research; this issue entails a consideration of the importance of standardization to the process of measurement;
- the different contexts of interviewing, such as the use of more than one interviewer and whether the administration of the interview is in person or by telephone;
- various prerequisites of structured interviewing, including: establishing rapport with the interviewee; asking questions as they appear on the interview schedule; recording exactly what is said by interviewees; ensuring there are clear instructions on the interview schedule concerning question sequencing and the recording of answers; and keeping to the question order as it appears on the schedule;
- problems with structured interviewing, including: the influence of the interviewer on respondents and the possibility of systematic bias in answers (known as response sets); the feminist critique of structured interview, which raises a distinctive cluster of problems with the method, is also examined.

Introduction

The interview is a common occurrence in social life, because there are many different forms of interview. There are job interviews, media interviews, social work interviews, police interviews, appraisal interviews. And then there are research interviews, which represent the kind of interview that will be covered in this and other chapters (such as Chapters 20 and 21). These different kinds of interview share some common features, such as the eliciting of information by the interviewer from the interviewee and the operation of rules of varying degrees of formality or explicitness concerning the conduct of the interview.

In the social research interview, the aim is for the interviewer to elicit from the interviewee or *respondent*, as he or she is frequently called in survey research, all manner of information: interviewees' own behaviour or that of others; attitudes; norms; beliefs; and values. There are many different types or styles of research interview, but the kind that is primarily employed in survey research is the structured interview, which is the focus of this chapter. Other kinds of interview will be briefly mentioned in this chapter but will be discussed in greater detail in later chapters.

The structured interview

The research interview is a prominent data-collection strategy in both quantitative and qualitative research. The survey is probably the chief context within which social researchers employ the structured interview (see Key concept 9.1) in connection with quantitative research, and it is this form of the interview that will be emphasized in this chapter. The **structured interview** is one of the two main ways of administering a survey research instrument, and its main forms are briefly outlined in Figure 8.2. This figure should be consulted as a background to this chapter and Chapter 10.

Key concept 9.1
What is a structured interview?

A structured interview, sometimes called a *standardized interview*, entails the administration of an interview schedule by an interviewer. The aim is for all interviewees to be given exactly the same context of questioning. This means that each respondent receives exactly the same interview stimulus as any other. The goal of this style of interviewing is to ensure that interviewees' replies can be aggregated, and this can be achieved reliably only if those replies are in response to identical cues. Interviewers are supposed to read out questions exactly and in the same order as they are printed on the schedule. Questions are usually very specific and very often offer the interviewee a fixed range of answers (this type of question is often called *closed*, *closed ended*, *pre-coded*, or *fixed choice*). The structured interview is the typical form of interview in survey research.

Thinking deeply 9.1
Common sources of error in survey research

There are many sources of error in survey research in addition to those associated with sampling. This is a list of the principal sources of error:

1. a poorly worded question;
2. the way the question is asked by the interviewer;
3. misunderstanding on the part of the interviewee;
4. memory problems on the part of the interviewee;
5. the way the information is recorded by the interviewer;
6. the way the information is processed, either when answers are coded or when data are entered into the computer.

The reason why survey researchers typically prefer the structured interview is that it promotes standardization of *both* the asking of questions *and* the recording of answers. This feature has two closely related virtues from the perspective of quantitative research: reducing error due to variation in the asking of questions, and greater accuracy in and ease of processing respondents' answers.

Reducing error due to interviewer variability

The standardization of both the asking of questions and the recording of answers means that, if the interview is properly executed, variation in people's replies will be due to 'true' or 'real' variation and not due to the interview context. To take a simple illustration, when we ask a question that is supposed to be an indicator of a concept, we want to keep error to a minimum, an issue that was touched on at the end of Chapter 8. We can think of the answers to a question as constituting the values that a variable takes. These values, of course, exhibit variation. This could be the question on alcohol consumption among students that was a focus of Chapter 8 at certain points. Students will vary in the number of alcohol units they consume (as in Figure 9.1). However, some respondents may be inaccurately classified in terms of the variable. There are a number of possible reasons for this (see Thinking deeply 9.1).

Most variables will contain an element of error, so that it is helpful to think of variation as made up of two components: true variation and error. In other words:

variation = true variation + variation due to error.

The aim is to keep the error component to a minimum (see Figure 9.2), since error has an adverse effect on the

Figure 9.1

A variable

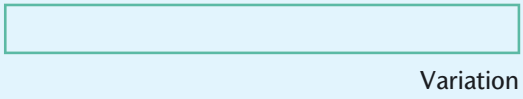


Figure 9.2

A variable with little error

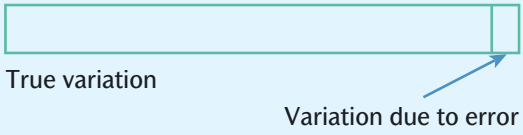
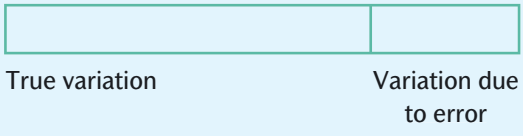


Figure 9.3

A variable with considerable error



validity of a measure. If the error component is quite high (see Figure 9.3), validity will be jeopardized. The significance for error of standardization in the structured interview is that two sources of variation due to error—the second and fifth in Thinking deeply 9.1—are likely to be less pronounced, since the opportunity for variation in interviewer behaviour in these two areas (asking questions and recording answers) is reduced.

The significance of standardization and of thereby reducing interviewer variability is this: assuming that there is no problem with an interview question due to such things as confusing terms or ambiguity (an issue that will be examined in Chapter 11), we want to be able to say as far as possible that the variation that we find is connected with true variation between interviewees and not to variation in the way a question was asked or the answers recorded in the course of the administration of a survey by structured interview. Variability can occur in either of two ways. First, *intra-interviewer variability*, whereby an interviewer is not consistent in the way he or she asks questions and/or records answers. Second, when there is more than one interviewer, there may be *inter-interviewer variability*, whereby interviewers are not consistent with each other in the ways they ask questions and/or record answers. Needless to say, these two sources of variability are not mutually exclusive; they can coexist, compounding the problem even further. In view of the significance of standardization, it is hardly surprising that some writers prefer to call the structured interview a *standardized interview* (e.g. Oppenheim 1992) or *standardized survey interview* (e.g. Fowler and Mangione 1990).

Accuracy and ease of data processing

Like self-completion questionnaires, most structured interviews contain mainly questions that are variously referred to as *closed*, *closed ended*, *pre-coded*, or *fixed choice*. This issue will be covered in detail in Chapter 11. However, this type of question has considerable relevance to the current discussion. With the **closed question**, the respondent is given a limited choice of possible answers. In other words, the interviewer provides respondents with two or more possible answers and asks them to select which one or ones apply. Ideally, this procedure will simply entail the interviewer placing a tick in a box by the answer(s) selected by a respondent or circling the selected answer or using a similar procedure. The advantage of this practice is that the potential for interviewer variability is reduced: there is no problem of whether the interviewer writes down everything that the respondent says or of misinterpretation of the reply given. If an *open* or *open-ended* question is asked, the interviewer may not write down everything said, may embellish what is said, or may misinterpret what is said.

However, the advantages of the closed question in the context of survey research go further than this, as we will see in Chapter 11. One advantage that is particularly significant in the context of the present discussion is that closed questions greatly facilitate the processing of data. When an open question is asked, the answers need to be sifted and *coded* in order for the data to be analysed quantitatively. Not only is this a laborious procedure, particularly if there is a large number of open questions and/or of respondents; it also introduces the potential for another source of error, which is the sixth in Thinking deeply 9.1: it is quite likely that error will be introduced as a result of variability in the coding of answers. When open questions are asked, the interviewer is supposed to write down as much of what is said as possible. Answers

can, therefore, be in the form of several sentences. These answers have to be examined and then categorized, so that each person's answer can be aggregated with other respondents' answers to a certain question. A number will then be allocated to each category of answer, so that the answers can then be entered into a computer database and analysed quantitatively. This general process is known as coding and will be examined in greater detail in Chapter 11.

Coding introduces yet another source of error. First, if the rules for assigning answers to categories, collectively known as the **coding frame**, are flawed, the variation that is observed will not reflect the true variation in interviewees' replies. Second, there may be variability in the ways in which answers are categorized. As with interviewing, there can be two sources: *intra-coder variability*, whereby the coder varies over time in the way in which the rules for assigning answers to categories are implemented, and *inter-coder variability*, whereby coders differ from each other in the way in which the rules for assigning answers to categories are implemented. If either (or both) source(s) of variability occur, at least part of the variation in interviewees' replies will not reflect true variation and instead will be caused by error.

The closed question sidesteps this problem neatly, because respondents allocate *themselves* to categories. The coding process is then a simple matter of attaching a different number to each category of answer and of entering the numbers into a computer database. It is not surprising, therefore, that this type of question is often referred to as pre-coded, because decisions about the coding of answers are typically undertaken as part of the design of the schedule—that is, before any respondents have actually been asked questions. There is very little opportunity for interviewers or coders to vary in the recording or the coding of answers. Of course, if some respondents misunderstand any terms in the alternative answers with which they are presented, or if the answers do not adequately cover the appropriate range of possibilities, the question will not provide a valid measure. However, that is a separate issue and one that will be returned to in Chapter 11. The chief point to register about closed questions for the moment is that, when compared to open questions, they reduce one potential source of error *and* are much easier to process for quantitative data analysis.

Other types of interview

The structured interview is by no means the only type of interview, but it is certainly the main type that is likely to be encountered in survey research and in quantitative research generally. Unfortunately, a host of different terms have been employed by writers on research methodology to distinguish the diverse forms of research interview. Key concept 9.2 represents an attempt to capture some of the major terms and types.

All the forms of interview outlined in Key concept 9.2, with the exception of the *structured interview* and the *standardized interview*, are primarily used in connection with qualitative research, and it is in that context that they will be encountered again later in this book. They are rarely used in connection with quantitative research, and survey research in particular, because the absence of standardization in the asking of questions and recording of answers makes respondents' replies difficult to aggregate and to process. This is not to say that they have no role at all. For example, as we will see in Chapter 11, the unstructured or semi-structured interview can have a useful role in relation to developing the fixed-choice alternatives with which respondents are provided in the kind of closed question that is typical of the structured interview.

Key concept 9.2
Major types of interview

- *Structured interview*. See Key concept 9.1.
- *Standardized interview*. See Key concept 9.1.
- *Semi-structured interview*. This is a term that covers a wide range of instances. It typically refers to a context in which the interviewer has a series of questions that are in the general form of an interview schedule but is able to vary the sequence of questions. The questions are frequently somewhat more general in their frame of reference from that typically found in a structured interview schedule. Also, the interviewer usually has some latitude to ask further questions in response to what are seen as significant replies.

- *Unstructured interview*. The interviewer typically has only a list of topics or issues, often called an interview guide or *aide-mémoire*, that are to be covered. The style of questioning is usually informal. The phrasing and sequencing of questions will vary from interview to interview.
- *Intensive interview*. This term is employed by Lofland and Lofland (1995) as an alternative term to the unstructured interview. Spradley (1979) uses the term ethnographic interview to describe a form of interview that is also more or less synonymous with the unstructured interview.
- *Qualitative interview*. For some writers, this term seems to denote an unstructured interview (e.g. Mason 1996), but more frequently it is a general term that embraces interviews of both the semi-structured and unstructured kind (e.g. Rubin and Rubin 1995).
- *In-depth interview*. Like the term 'qualitative interview', this one sometimes refers to an unstructured interview but more often refers to both semi-structured and unstructured interviewing. The use of this term seems to be increasing.
- *Focused interview*. This is a term devised by Merton et al. (1956) to refer to an interview using predominantly open questions to ask interviewees questions about a specific situation or event that is relevant to them and of interest to the researcher.
- *Focus group*. This is the same as the focused interview, but interviewees discuss the specific issue in groups. See Key concept 21.1 for a more detailed definition.
- *Group interview*. Some writers see this term as synonymous with the focus group, but a distinction may be made between the latter and a situation in which members of a group discuss a variety of matters that may be only partially related.
- *Oral history interview*. This is an unstructured or semi-structured interview in which the respondent is asked to recall events from his or her past and to reflect on them. There is usually a cluster of fairly specific research concerns to do with a particular epoch or event, so there is some resemblance to a focused interview (see the section on 'Life history and oral history interviewing' in Chapter 20.).
- *Life history interview*. This is similar to the oral history interview, but the aim of this type of unstructured interview is to glean information on the entire biography of each respondent (see the section on 'Life history and oral history interviewing' in Chapter 20.)

Interview contexts

In an archetypal interview, an interviewer stands or sits in front of the respondent asking the latter a series of questions and writing down the answers. However, there are several possible departures from it, although this archetype is the most usual context for an interview.

More than one interviewee

In the case of group interviews or focus groups, there is more than one, and usually quite a few more than one, respondent or interviewee. Nor is this the only context in which more than one person is interviewed. McKee and Bell (1985), for example, interviewed couples in their study of the impact of male unemployment, while, in my research on visitors to Disney theme parks, not just couples but often their children took part in the interview as well (Bryman 1999). However, it is very unusual for structured interviews to be used in connection with this kind of questioning. In survey research, it is almost always a specific individual who is the object of questioning. Indeed, in survey interviews it is very advisable to discourage as far as possible the presence and intrusion of others during the course of the interview. Investigations in which more than one person is being interviewed tend to be exercises in qualitative research, though this is not always the case: Pahl's (1990) study of patterns of control of money among couples employed structured interviewing of couples and of husbands and wives separately.

More than one interviewer

This is a very unusual situation in social research, because of the considerable cost that is involved in dispatching two (or indeed more than two) people to interview someone. Bechhofer et al. (1984) describe research in which two people interviewed individuals in a wide range of occupations. However, while their approach achieved a number of benefits for them, their interviewing style was of the unstructured kind that is typically employed in qualitative research, and they argue that the presence of a second interviewer is unlikely to achieve any added value in the context of structured interviewing.

In person or by telephone?

A third way in which the archetype may not be realized is that interviews may be conducted by telephone rather than face-to-face. While telephone interviewing is quite common in commercial fields like market research, where it usually takes the form of computer-assisted telephone interviewing (CATI; see below), it is still far more customary to read reports of studies based on face-to-face interviews in academic social research, but see Research in focus 9.1 for an interesting example.

Research in focus 9.1
A telephone survey of the unemployed in Sweden

Nordenmark and Strandh (1999) report the findings of an interesting study of mental well-being among the unemployed in Sweden. Early in 1996 a national random sample of 3,500 was drawn from a register of all unemployed persons that is maintained by the Swedish Labour Market Board. A telephone survey was conducted with members of the sample. The response rate was 74 per cent. The interview schedule included questions on such issues as 'mental well-being, the economy, work involvement, belief in the future, wage demands and job search behaviour' (Nordenmark and Strandh 1999: 585). Nearly two years later, those who had participated were re-interviewed by telephone with very similar questions. This is, therefore, an example of a *panel study*. The authors inform us that only part (around 6 per cent) of the 26 per cent who did not respond was due to a refusal to participate; the remainder was due to problems of contacting respondents.

There are several advantages of telephone over personal interviews.

- On a like-for-like basis, they are far cheaper and also quicker to administer. This arises because, for personal interviews, interviewers have to spend a great deal of time and money travelling between respondents. This factor will be even more pronounced when a sample is geographically dispersed, a problem that is only partially mitigated for personal interview surveys by strategies like cluster sampling. Of course, telephone interviews take time and hired interviewers have to be paid, but the cost of conducting a telephone interview will still be lower than a comparable personal one. Moreover, the general efficiency of telephone interviewing has been enhanced with the advent and widespread use in commercial circles of computer-assisted telephone interviewing (CATI).
- The telephone interview is easier to supervise than the personal interview. This is a particular advantage when there are several interviewers, since it becomes easier to check on interviewers' transgressions in the asking of questions, such as rephrasing questions or the inappropriate use of probes by the interviewer. Interviews can be tape-recorded so that data quality can be assessed, but this raises issues that relate to data protection and confidentiality, so that this procedure has to be treated cautiously.
- Telephone interviewing has a further advantage, which is to do with evidence (which is not as clear-cut as one might want) that suggests that, in personal interviews, respondents' replies are sometimes affected by characteristics of the interviewer (for example, class, ethnicity) and indeed by his or her mere presence (implying that the interviewees may reply in ways they feel will be deemed desirable by interviewers). The remoteness of the interviewer in telephone interviewing removes this potential source of bias to a significant extent. The interviewer's personal characteristics cannot be seen, and the fact that he or she is

not physically present may offset the likelihood of respondents' answers being affected by the interviewer.

Telephone interviewing suffers from certain limitations when compared to the personal interview.

- People who do not own or who are not contactable by telephone obviously cannot be interviewed by telephone. Since this characteristic is most likely to be a feature of poorer households, the potential for sampling bias exists. Also, many people choose to be ex-directory—that is, they have taken action for their telephone numbers not to appear in a telephone directory. Again, these people cannot be interviewed by telephone. One possible solution to this last difficulty is *random digit dialling*. With this technique, the computer randomly selects telephone numbers within a predefined geographical area. Not only is this a random process that conforms to the rules about probability sampling examined in Chapter 8; it also stands a chance of getting at ex-directory households. But it cannot, of course, gain access to those without a telephone at all.
- Respondents with hearing impairments are likely to find telephone interviewing much more difficult for them than personal interviewing.
- The length of a telephone interview is unlikely to be sustainable beyond 20–25 minutes, whereas personal interviews can be much longer than this (Frey 2004).
- The question of whether response rates (see Key concept 8.2) are lower with surveys by telephone interview than with surveys by personal interview is unclear, in that there is little consistent evidence on this question. However, there is a general belief that telephone interviews achieve slightly lower rates than personal interviews (Frey and Oishi 1995; Shuy 2002; Frey 2004).
- There is some evidence to suggest that telephone interviews fare less well for the asking of questions about sensitive issues, such as drug and alcohol use, income, tax returns, and health. However, the evidence is not entirely consistent on this point, though it is probably sufficient to suggest that, when many questions of this kind are to be used, a personal interview may be superior (Shuy 2002).
- Developments in telephone communications such as the growing use of answerphones and other forms of call screening and of mobile phones have almost certainly had an adverse effect on telephone surveys in terms of response rates and the general difficulty of getting access to respondents through conventional landlines. Households that rely exclusively on mobile phones represent a particular difficulty.
- Telephone interviewers cannot engage in observation. This means that they are not in a position to respond to signs of puzzlement or unease on the faces of respondents when they are asked a question. In a personal interview, the interviewer may respond to such signs by restating the question or attempting to clarify the meaning of the question, though this has to be handled in a standardized way as far as possible. A further issue relating to the inability of the interviewer to observe is that, sometimes, interviewers may be asked to collect subsidiary information in connection with their visits (for example, whether a house is dilapidated). Such information cannot be collected when telephone interviews are employed.
- It is frequently the case that specific individuals in households or firms are the targets of an interview. In other words, simply anybody will not do. This requirement is likely to arise from the specifications of the population to be sampled, which means that people in a certain role or position or with particular characteristics are to be interviewed. It is probably more difficult to ascertain by telephone interview whether the correct person is replying.
- The telephone interviewer cannot readily employ visual aids such as show cards (see below) from which respondents might be asked to select their replies or to use diagrams or photographs.
- There is some evidence to suggest that the quality of data derived from telephone interviews is inferior to that of comparable face-to-face interviews. A series of experiments reported by Holbrook et al. (2003) on the mode of survey administration in the USA using long questionnaires found that respondents interviewed by telephone were more likely to: express no opinion or 'don't know' (see Chapter 11 for more on this issue); to answer in the same way to a series of linked questions; to express socially desirable answers; to be apprehensive about the interview; and to be more likely to be dissatisfied with the time taken by the interviews (even though they were invariably shorter than in the face-to-face mode). Also, telephone interviewees tended to be less engaged with the interview process. While these results should be viewed with caution, since studies like these are bound to be affected by such factors as the use of a large questionnaire on a national sample, they do provide interesting food for thought.

Computer-assisted interviewing

In recent years, increasing use has been made of computers in the interviewing process, especially in commercial survey research of the kind conducted by market research and opinion polling organizations. There are two main formats for computer-assisted interviewing: computer-assisted personal interviewing (CAPI) and computer-assisted telephone interviewing (CATI). A very large percentage of telephone interviews is conducted with the aid of personal computers. Among commercial survey organizations, almost all telephone interviewing is of the CATI kind nowadays, and this kind of interview has become one of the most popular formats for such firms. The main reasons for the growing use of CAPI has been that the increased portability and affordability of 'laptop' computers, and the growth in the number and quality of software packages that provide a platform for devising interview schedules, provide greater opportunity for them to be used in connection with face-to-face interviews. CAPI and CATI have not infiltrated academic survey research to anything like the same degree that they have commercial survey research, although that picture is likely to change considerably because of the many advantages they possess. Indeed, the survey element of the mixed methods study Cultural Capital and Social Exclusion (CCSE), referred to in Research in focus 2.9, was administered by CAPI. In any case, many of the large datasets that are used for secondary analysis (see Chapter 14 for examples) derive from computer-assisted interviewing studies undertaken by commercial or large social research organizations.

With computer-assisted interviewing, the questions that comprise an interview schedule appear on the screen. As interviewers ask each question, they 'key in' the appropriate reply using the keyboard (for open questions) or using a mouse (for closed questions) and proceed to the next question. Moreover, this process has the great advantage that, when *filter questions* (see Tips and skills 'Instructions for interviewers in the use of a filter question') are asked, so that certain answers may be skipped as a result of a person's reply, the computer can be programmed to 'jump' to the next relevant question. This removes the possibility of interviewers inadvertently asking inappropriate questions or failing to ask ones that should be asked. As such, computer-assisted interviewing enhances the degree of control over the interview process and can therefore improve standardization of the asking and recording of questions. However, there is very little evidence to suggest that the quality of data deriving from computer-assisted interviews is demonstrably superior to comparable paper and pencil interviews (Couper and Hansen 2002). If the interviewer is out in the field all day, he or she can either take a disk with the saved data to the research office or send the data down a telephone line with the aid of a modem. It is possible that technophobic respondents may be a bit alarmed by their use, but, by and large, the use of computer-assisted interviewing seems destined to grow.

For their part, there is evidence that professional interviewers generally like computer-assisted interviewing, often feeling that it improves the image of their occupation, though there are many who are concerned about the problems that might arise from technical difficulties and the inconvenience of correcting errors with a computer as opposed to with a pen. One issue that sometimes disconcerts interviewers is the fact that they can see only part of the schedule at any one time (Couper and Hansen 2002). One potential problem with CAPI and CATI is 'miskeying', where the interviewer clicks on the wrong reply. Whether this is more likely to occur than when the interviewer is using pen and paper is unknown. In the CCSE study, as noted in Research in focus 2.9, qualitative interviews were conducted with some of the survey respondents. In part this was done so that participants in the semi-structured interview phase could be asked about some of the answers they had given in the survey interview. As a result, the researchers found that sometimes the participant had been recorded as giving a particular answer that was in fact incorrect. An example is a respondent who had been recorded as indicating in the survey interview as preferring to eat out in Italian restaurants when in fact it should have been Indian ones (Silva and Wright 2008). As the researchers note, it is impossible to know how this error occurred, but miskeying is one possible reason.

The discussion in the previous section of telephone interviewing and of CATI in this section presumes that the medium is a landline. However, with the huge growth in the use of mobile phones (cellular or cell phones) there is the prospect that these will have a role in future years. Since lists of mobile-phone users are unlikely to be available in the way that telephone directories are, random digit dialing (RDD) is most likely to be employed by researchers seeking to interview by mobile phone. Zuwallack (2009) reports the findings of some CATI projects conducted by mobile phone in the USA on health-related issues. The researchers found that a lot of people hung up when contacted but that those respondents who persisted formed a useful complement to conventional landline telephone surveys because many of them had characteristics often under-represented in such surveys,

such as young adults and minorities. Of particular interest is that a large percentage of respondents lived in households without a landline, suggesting that, if the number of mobile-only households increases, mobile-phone surveys may become increasingly significant. Zuwallack also reports that the mobile-phone survey is more expensive than the equivalent landline CATI survey.

One further point to register in connection with computer-assisted interviewing is that the section has not included **Internet surveys**. The reason for this is that such surveys are more properly considered as using self-completion questionnaires rather than structured interviewing (see Figure 8.2). With such surveys, there is no interviewer in the sense of a person who verbally asks questions. Internet surveys are covered in Chapter 28.

Conducting interviews

Issues concerning the conduct of interviews are examined here in a very general way. In addition to the matters considered here, there is clearly the important issue of how to word the interview questions themselves. This area will be explored in Chapter 11, since many of the rules of question-asking relate to self-completion questionnaire techniques like postal questionnaires as well as to structured interviews. One further general point to make here is that the advice concerning the conduct of interviews provided in this chapter relates to structured interviews. The framework for carrying out the kinds of interviewing conducted in qualitative research (such as unstructured and semi-structured interviewing and focus groups) will be handled in later chapters.

Know the schedule

Before interviewing anybody, an interviewer should be fully conversant with the schedule. Even if you are the only person conducting interviews, make sure you know it inside out. Interviewing can be stressful for interviewers, and it is possible that under duress standard interview procedures like filter questions (see Tips and skills 'Instructions for interviewers in the use of a filter question') can cause interviewers to get flustered and miss questions out or ask the wrong questions. If two or more interviewers are involved, they need to be fully trained to know what is required of them and to know their way around the schedule. Training is especially important in order to reduce the likelihood of interviewer variability in the asking of questions, which is a source of error.

Introducing the research

Prospective respondents have to be provided with a credible rationale for the research in which they are being asked to participate and for giving up their valuable time. This aspect of conducting interview research is of particular significance at a time when response rates to survey research appear to be declining, though, as noted in Chapter 8, the evidence on this issue is the focus of some disagreement. The introductory rationale may be either spoken by the interviewer or written down. In many cases, respondents may be presented with both modes. It comes in spoken form in such situations as when interviewers make contact with respondents on the street or when they 'cold call' respondents in their homes in person or by telephone. A written rationale will be required to alert respondents that someone will be contacting them in person or on the telephone to request an interview. Respondents will frequently encounter both forms—for example, when they are sent a letter and then when they ask the interviewer who turns up to interview them what the research is all about. It is important for the two accounts to be consistent, as this could be a test!

Introductions to research should typically contain the bits of information outlined in Tips and skills 'Topics and issues to include in an introductory statement'. Since interviewers represent the interface between the research and the respondent, they have an important role in maximizing the response rate for the survey. In addition the following points should be borne in mind.

- Interviewers should be prepared to keep calling back if interviewees are out or unavailable. This will require taking into account people's likely work and leisure habits—for example, there is no point in calling at home on people who work during the day. In addition, people living alone may be reluctant to answer the door when it is dark because of fear of crime.
- Be self-assured. You may get a better response if you presume that people will agree to be interviewed rather than that they will refuse.

- Reassure people that you are not a salesperson. Because of the tactics of certain organizations whose representatives say they are doing market or social research, many people have become very suspicious of people saying they would just like to ask you a few questions.
- Dress in a way that will be acceptable to a wide spectrum of people.
- Make it clear that you will be happy to find a time to suit the respondent.

Rapport

It is frequently suggested that it is important for the interviewer to achieve *rapport* with the respondent. This means that very quickly a relationship must be established that encourages the respondent to want (or at least be prepared) to participate in and persist with the interview. Unless an element of rapport can be established, some respondents may initially agree to be interviewed but then decide to terminate their participation because of the length of time the interview is taking or perhaps because of the nature of the questions being asked. While this injunction essentially invites the interviewer to be friendly with respondents and to put them at ease, it is important that this quality is not stretched too far. Too much rapport may result in the interview going on too long and the respondent suddenly deciding that too much time is being spent on the activity. Also, the mood of friendliness may result in the respondent answering questions in a way that is designed to please the interviewer. The achievement of rapport between interviewer and respondent is therefore a delicate balancing act. Moreover, it is probably somewhat easier to achieve in the context of the face-to-face interview than in the telephone interview, since in the latter the interviewer is unable to offer obvious visual cues of friendliness such as smiling or maintaining good eye contact, which are also frequently regarded as conducive to gaining and maintaining rapport.

Tips and skills

Topics and issues to include in an introductory statement

There are several issues to include in an introductory statement to a prospective interviewee. The following list comprises the principal considerations.

- Make clear the identity of the person who is contacting the respondent.
- Identify the auspices under which the research is being conducted—for example, a university, a market research agency.
- Mention any research funder, or, if you are a student doing an undergraduate or postgraduate dissertation or doing research for a thesis, make this clear.
- Indicate what the research is about in broad terms and why it is important, and give an indication of the kind of information to be collected.
- Indicate why the respondent has been selected—for example, selected by a random process.
- Make it clear that participation is voluntary.
- Reassure the respondent that he or she will not be identified or be identifiable in any way. This can usually be achieved by pointing out that data are anonymized when they are entered into the computer and that analysis will be conducted at an aggregate level.
- Provide reassurance about the confidentiality of any information provided.
- Provide the respondent with the opportunity to ask any questions—for example, provide a contact telephone number if the introduction is in the form of a written statement, or, if in person, simply ask if the respondent has any questions.

These suggestions are also relevant to the covering letter that accompanies postal questionnaires, except that researchers using this method need to remember to include a stamped-addressed envelope!

Asking questions

It was earlier suggested that one of the aims of the structured interview is to ensure that each respondent is asked exactly the same questions. Recall that in Thinking deeply 9.1 it was pointed out that variation in the ways a question is asked is a potential source of error in survey research. The structured interview is meant to reduce the likelihood of this occurring, but it cannot guarantee that this will not occur, because there is always the possibility that interviewers will embellish or otherwise change a question when it is asked. There is considerable evidence that this occurs, even among centres of social research that have a solid reputation for being rigorous in following correct methodological protocol (Bradburn and Sudman 1979). The problem with such variation in the asking of questions was outlined above: it is likely to engender variation in replies that does not reflect 'true' variation—in other words, error. Consequently, it is important for interviewers to appreciate the importance of keeping exactly to the wording of the questions they are charged with asking.

You might say: 'does it really matter?' In other words, surely small variations to wording cannot make a significant difference to people's replies? While the impact of variation in wording obviously differs from context to context and is in any case difficult to quantify exactly, experiments in question-wording suggest that even small variations in wording can exert an impact on replies (Schuman and Presser 1981). Three experiments in England conducted by Social and Community Planning Research concluded that a considerable number of interview questions is affected by interviewer variability. The researchers estimated that, for about two-thirds of the questions that were considered, interviewers contributed to less than 2 per cent of the total variation in each question (M. Collins 1997). On the face of it, this is a small amount of error, but the researchers regarded it as a cause for concern.

The key point to emerge, then, is the importance of getting across to interviewers the importance of asking questions as they are written. There are many reasons why interviewers may vary question-wording, such as reluctance to ask certain questions, perhaps because of embarrassment (M. Collins 1997), but the general admonition to keep to the wording of the question needs to be constantly reinforced when interviewers are being trained. It also needs to be borne in mind for your own research.

Recording answers

An identical warning for identical reasons can be registered in connection with the recording of answers by interviewers, who should write down respondents' replies as exactly as possible. Not to do so can result in interviewers distorting respondents' answers and hence introducing error. Such errors are less likely to occur when the interviewer has merely to allocate respondents' replies to a category, as in a closed question. This process can require a certain amount of interpretation on the part of the interviewer, but the error that is introduced is far less than when answers to open questions are being written down (Fowler and Mangione 1990).

Clear instructions

In addition to instructions about the asking of questions and the recording of answers, interviewers need instructions about their progress through an interview schedule. An example of the kind of context in which this is likely to occur is in relation to *filter questions*. Filter questions require the interviewer to ask questions of some respondents but not others. For example, the question:

> For which political party did you vote at the last general election?

presumes that the respondent did in fact vote. This option can be reflected in the fixed-choice answers that are provided, so that one of these is a 'did not vote' alternative. However, a better solution is not to presume anything about voting behaviour but to ask respondents whether they voted in the last general election and then to filter out those who did not vote. The foregoing question about the political party voted for can then be asked of those who did in fact vote. Similarly, in a study of meals, there is no point in asking vegetarians lots of questions about eating meat. It will probably work out best to filter vegetarians out and then possibly ask them a separate series of questions. Tips and skills 'Instructions for interviewers in the use of a filter question' provides a simple example in connection with an imaginary study of alcohol consumption. The chief point to register about this example is that it requires clear instructions for the interviewer. If such instructions are not provided, there is the risk that either respondents will be asked inappropriate questions (which can be irritating for them) or

the interviewer will inadvertently fail to ask a question (which results in missing information).

Question order

In addition to interviewers being warned about the importance of not varying the asking of questions and the recording of answers, they should be alerted to the importance of keeping to the order of asking questions. For one thing, varying the question order can result in certain questions being accidentally omitted, because the interviewer may forget to ask those that have been leap-frogged during the interview. Also, variation in question order may have an impact on replies: if some respondents have been previously asked a question that they should have been asked whereas others have not, a source of variability in the asking of questions will have been introduced and therefore a potential source of error.

Tips and skills

Instructions for interviewers in the use of a filter question

Each of the following questions includes an instruction to the interviewer about how to proceed.

1. Have you consumed any alcoholic drinks in the last twelve months?
Yes ____
No ____
(if **No** proceed to question 4)

2. (*To be asked if interviewee replied* **Yes** *to question 1*)
Which of the following alcoholic drinks do you consume most frequently?
(Ask respondent to choose the category that he or she drinks most frequently and tick one category only.)
Beer ____
Spirits ____
Wine ____
Liquors ____
Other ____ specify _____

3. How frequently do you consume alcoholic drinks?
(Ask interviewee to choose the category that comes closest to his or her current practice.)
Daily ____
Most days ____
Once or twice a week ____
Once or twice a month ____
A few times a year ____
Once or twice a year ____

4. (*To be asked if interviewee replied* **No** *to question 1*)
Have you ever consumed alcoholic drinks?
Yes ____
No ____

Quite a lot of research has been carried out on the general subject of question order, but few if any consistent effects on people's responses that derive from asking questions at different points in a questionnaire or interview schedule have been unveiled. Different effects have been demonstrated on various occasions.

A study in the USA found that people were less likely to say that their taxes were too high when they had been previously asked whether government spending ought to be increased in a number of areas (Schuman and Presser 1981: 32). Apparently, some people perceived an inconsistency between wanting more spending and lower taxes, and adjusted their answers accordingly. Research on crime victimization in the USA suggests that earlier questions may affect the salience of later issues (Schuman and Presser 1981: 45). Respondents were asked whether they had been victims of crime in the preceding twelve months. Some respondents had been previously asked a series of questions about their attitudes to crime, whereas others had not. Those who had been asked about their attitudes reported considerably more crime than those who had not been asked.

Mayhew (2000) provides an interesting anecdote on question order in relation to the British Crime Survey. Each wave of the BCS has included the question:

> Taking everything into account, would you say the police in this area do a good job or a poor job?

In 1988 this question appeared twice by mistake for some respondents! For all respondents it appeared early on, but for around half it also appeared later on in the context of questions on contact with the police. Of those given the question twice, 66 per cent gave the same rating, but 22 per cent gave a more positive rating to the police and just 13 per cent gave a less favourable one. Mayhew suggests that, as the interview wore on, respondents became more sensitized to crime-related issues and more sympathetic to the pressures on the police.

However, it is difficult to draw general lessons from such research, at least in part because experiments in question order do not always reveal clear-cut effects of varying the order in which questions are asked, even in cases where effects might legitimately have been expected. There are two general lessons.

1. Within a survey, question order should not be varied (unless, of course, question order is the subject of the study!).
2. Researchers should be sensitive to the possible implications of the effect of early questions on answers to subsequent questions.

The following rules about question order are sometimes proposed.

- Early questions should be directly related to the topic of the research, about which the respondent has been informed. This removes the possibility that the respondent will be wondering at an early stage in the interview why he or she is being asked apparently irrelevant questions. This injunction means that personal questions about age, social background, and so on should *not* be asked at the beginning of an interview.
- As far as possible, questions that are more likely to be salient to respondents should be asked early in the interview schedule, so that their interest and attention are more likely to be secured. This suggestion may conflict with the previous one, in that questions specifically on the research topic may not be obviously salient to respondents, but it implies that as far as possible questions relating to the research topic that are more likely to grab their attention should be asked at or close to the start of the interview.
- Potentially embarrassing questions or ones that may be a source of anxiety should be left till later. In fact, research should be designed to ensure that, as far as possible, respondents are not discomfited, but it has to be acknowledged that with certain topics this effect may be unavoidable.
- With a long schedule or questionnaire, questions should be grouped into sections, since this allows a better flow than skipping from one topic to another.
- Within each group of questions, general questions should precede specific ones. Tips and skills 'A sequence of questions on the topic of identity cards' provides an illustration of such a sequence, which follows the recommendations of Gallup (1947, cited in Foddy 1993: 61–2). The example is concerned to demonstrate how the approach might operate in connection with identity cards, which have been an area of discussion and some controversy in the UK in recent years. The question order sequence is designed with a number of features in mind. It is designed to establish people's levels of knowledge of identity cards before asking questions about it and to distinguish those who feel strongly about it from those who do not. According to Foddy (1993), the second question is always open ended, so that respondents' frames of references can be established with respect to the topic at hand. However, it seems likely that, if sufficient pilot research has been carried out, a closed question could be envisaged, a point that applies equally to question 4.

Tips and skills
A sequence of questions on the topic of identity cards

1. Have you heard of identity cards?
   Yes ____ No ____
2. What are your views about identity cards?
3. Do you favour or not favour identity cards?
   Favour ____ Not favour ____
4. Why do you favour (not favour) identity cards?
5. How strongly do you feel about this?
   Very strongly ____
   Fairly strongly ____
   Not at all strongly ____

- A further aspect of the rule that general questions should precede specific ones is that it has been argued that, when a specific question comes before a general one, the aspect of the general question that is covered by the specific one is discounted in the minds of respondents because they feel they have already covered it. Thus, if a question about how people feel about the amount they are paid precedes a general question about job satisfaction, there are grounds for thinking that respondents will discount the issue of pay when responding about job satisfaction.
- It is sometimes recommended that questions dealing with opinions and attitudes should precede questions to do with behaviour and knowledge. This is because it is felt that behaviour and knowledge questions are less affected by question order than questions that tap opinions and attitudes.
- During the course of an interview, it sometimes happens that a respondent provides an answer to a question that is to be asked later in the interview. Because of the possibility of a question order effect, when the interviewer arrives at the question that appears already to have been answered, it should be repeated.

However, question order effects remain one of the more frustrating areas of structured interview and questionnaire design, because of the inconsistent evidence that is found and because it is difficult to formulate generalizations or rules from the evidence that does point to their operation. An interesting discussion about question order took place some years ago in connection with the study of social class and is discussed in Thinking deeply 9.2.

Thinking deeply 9.2
A debate about question order

An interesting case of the issue of question order becoming a focus of controversy is provided by the research on social class by Marshall et al. (1988), which is referred to in more detail in Research in focus 7.4 and 9.2. In a critique of the research, Saunders (1989) argues that it reveals what he calls 'socialist preconceptions', implying that values overtly intruded into the research (see Figure 2.3). Saunders argues that one way in which this was revealed was the sheer weight of questions about social class prior to respondents being asked about the groups to which they saw themselves as belonging. Saunders (1989: 4) writes:

> A glance at their questionnaire reveals that respondents were bombarded with questions about class right from the start of the interview. Following no fewer than 28 detailed questions about the class system, respondents were then asked if they thought they belonged to any social class and whether there was, by any chance, any other grouping they identified with apart from their class. Not surprisingly, most agreed that they did belong to one class or another . . . and that they could not think of any other identity. . . . Armed with their 'findings', the authors then conclude that we are all class-oriented after all and that other identities are far less important.

Two of the book's authors replied with a spirited rebuttal. They replied that

> the question about class was preceded by 30 substantive items. Six of these have no obvious relationship to the issue of social identities; for example, they elicit perceptions of Britain's economic performance. . . . No less than 17 . . . were specifically designed to make interviewees see the world in terms other than those of social class. They invited people to think of themselves as consumers . . . as voters . . . as members of ethnic or gender groupings; as employees . . . in short, as everything but members of an identifiable social class. Interviewees were also asked whether there were 'any important conflicts in Britain today' before the word social class was ever mentioned. Then, and only then, were they quizzed about their perception of Britain as a specifically class society. (Marshall and Rose 1989: 5)

This is an interesting debate because it raises the issue of the role of values and bias in social research and also because it relates to the issue of question order while demonstrating the difficulty of being definitive about the issue. For example, while Marshall and Rose's reply is convincing, it might be that it is the *relative number* of questions about social class preceding questions of identity that may have been influential and to which Saunders alludes. Nonetheless, the debate usefully demonstrates the difficulty of producing conclusive evidence about question order effects.

Probing

Probing is a highly problematic area for researchers employing a structured interview method. It frequently happens in interviews that respondents need help with their answers. One obvious case is where it is evident that they do not understand the question—they may either ask for further information or it is clear from what they say that they are struggling to understand the question or to provide an adequate answer. The second kind of situation the interviewer faces is when the respondent does not provide a sufficiently complete answer and has to be probed for more information. The problem in either situation is obvious: the interviewer's intervention may influence the respondent, and the nature of interviewers' interventions may differ. A potential source of variability in respondents' replies that does not reflect 'true' variation is introduced—that is, error.

Some general tactics with regard to probes are as follows.

- If further information is required, usually in the context of an open question, standardized probes can be employed, such as 'Could you say a little more about that?' or 'Are there any other reasons why you think that?' or simply 'mmmm . . . ?'.
- If the problem is that when presented with a closed question the respondent replies in a way that does not allow the interviewer to select one of the pre-designed answers, the interviewer should repeat the fixed-choice alternatives and make it apparent that the answer needs to be chosen from the ones that have been provided.
- If the interviewer needs to know about something that requires quantification, such as the number of visits to building societies in the last four weeks or the number of building societies in which the respondent has accounts, but the respondent resists this by answering in general terms ('quite often' or 'I usually go to the building society every week'), the interviewer needs to persist with securing a number from the respondent. This will usually entail repeating the question. The interviewer should not try to second guess a figure on the basis of the respondent's reply and then suggest that figure to him or her, since the latter may be unwilling to demur from the interviewer's suggested figure.

However, from the point of view of standardizing the asking of questions in surveys using structured interviewing, probing should be kept to a minimum (assuming it cannot be eliminated) because it introduces error. This occurs because it is impossible for interviewers to probe in a consistent manner and because interviewer effects are more likely to occur, whereby characteristics of the interviewer have an impact on the respondent's replies (Groves et al. 2004: 281–2).

Prompting

Prompting occurs when the interviewer suggests a possible answer to a question to the respondent. The key prerequisite here is that all respondents receive the same prompts. All closed questions entail standardized prompting, because the respondent is provided with a list of possible answers from which to choose. An unacceptable approach to prompting would be to ask an open question and to suggest possible answers only to some respondents, such as those who appear to be struggling to think of an appropriate reply.

During the course of a face-to-face interview, there are several circumstances in which it will be better for the interviewer to use 'show cards' rather than rely on reading out a series of fixed-choice alternatives. Show cards (sometimes called 'flash cards') display all the answers from which the respondent is to choose and are handed to the respondent at different points of the interview. Three kinds of context in which it might be preferable to employ show cards rather than to read out the entire set of possible answers are as follows.

- There may be a very long list of possible answers. For example, respondents may be asked which daily newspaper they each read most frequently. To read out a list of newspapers would be tedious, and it is probably better to hand the respondent a list of newspapers from which to choose.
- Sometimes, during the course of interviews, respondents are presented with a group of questions to which the same possible answers are attached. An example of this strategy is Likert scaling, an approach to attitude measurement that was discussed in Key concept 7.2. The components of a Likert scale are often referred to as *items* rather than as *questions*, since strictly speaking respondents are not being asked questions but are presented with statements to which they are asked to indicate their levels of agreement. See Research in focus 7.2 and 7.5 for examples. It would be excruciatingly dull to read out all five or seven possible answers ten times. Also, it may be expecting too much of respondents to read out the answers once and then require them to keep the possible answers in their heads for the entire batch of questions to which they apply. A show card that can be used for the entire batch and to which respondents can constantly refer is an obvious solution. As was mentioned in Key concept 7.2, most Likert scales of this kind comprise five levels of agreement/disagreement, and it is this more conventional approach that is illustrated in the show card in Tips and skills 'A show card'.

Tips and skills
A show card

Card 6

Strongly agree

Agree

Undecided

Disagree

Strongly disagree

- Some people are not keen to divulge personal details such as their age or their income. One way of neutralizing the impact of such questioning is to present respondents with age or income bands with a letter or number attached to each band. They can then be asked to say which letter/number applies to

them (see Tips and skills 'Another show card'). This procedure will obviously not be appropriate if the research requires *exact* ages or incomes. It may be extendable to sensitive areas such as number of sexual partners or sexual practices for the same kinds of reason.

Tips and skills
Another show card

Card 11

1. Below 20
2. 20–29
3. 30–39
4. 40–49
5. 50–59
6. 60–69
7. 70 and over

Student experience
The need for structure in a survey interview

Joe Thompson's survey research on students and their views of accommodation and facilities at his university was part of a team project. After he and other members of his team had piloted the interview schedule, they decided that it was not sufficiently structured. They felt that they needed to impose more structure and decided to use show cards (he refers to them by their other common name 'cue cards').

> The group therefore used opportunistic sampling to test if the questionnaire would be successful when applied in a social setting, having to give the questionnaire to one person over the week. The following week the group discussed the issues they had encountered when carrying out the pilot questionnaire, raising amongst others the concern of not having a standard interview procedure, which would mean that certain biases could affect the results. Therefore the group decided they would use cue cards when giving the options in answer to the question, so as to avoid leading questions, etc. After these changes were implemented, the final version of the questionnaire was produced.

To read more about Joe's research experiences, go to the Online Resource Centre that accompanies this book at:
www.oxfordtextbooks.co.uk/orc/brymansrm4e/

Leaving the interview

Do not forget common courtesies like thanking respondents for giving up their time. But the period immediately after the interview is one in which some care is necessary, in that sometimes respondents try to engage the interviewer in a discussion about the purpose of the interview. Interviewers should resist elaboration beyond their standard statement, because respondents may communicate what they are told to others, which may bias the findings.

Training and supervision

On several occasions, reference has been made to the need for interviewers to be trained. The standard texts on

survey research and on interviewing practice tend to be replete with advice on how best to train interviewers. Such advice is typically directed at contexts in which a researcher hires an interviewer to conduct a large amount or even all the interviews. It also has considerable importance in research in which several interviewers (who may be either collaborators or hired interviewers) are involved in a study, since the risk of interviewer variability in the asking of questions needs to be avoided.

For many readers of this book who are planning to do research, such situations are unlikely to be relevant because they will be 'lone' researchers. You may be doing an undergraduate dissertation, or an exercise for a research methods course, or you may be a postgraduate conducting research for a Master's dissertation or for a thesis. Most people in such a situation will not have the luxury of being able to hire a researcher to do any interviewing (though you may be able to find someone to help you a little). When interviewing on your own, you must in a sense train yourself to follow the procedures and advice provided above. This is a very different situation from a large research institute or market research agency, which relies on an army of hired interviewers who carry out the interviews. Whenever people other than the lead researcher are involved in interviewing, they will need training and supervision in the following areas:

- contacting prospective respondents and providing an introduction to the study;
- reading out questions as written and following instructions in the interview schedule (for example, in connection with filter questions);
- using appropriate styles of probing;
- recording exactly what is said;
- maintaining an interview style that does not bias respondents' answers.

Fowler (1993) cites evidence that suggests that training of less than one full day rarely creates good interviewers.

Supervision of interviewers in relation to these issues can be achieved by:

- checking individual interviewers' response rates;
- tape-recording at least a sample of interviews;
- examining completed schedules to determine whether any questions are being left out or if they are being completed properly;
- call-backs on a sample of respondents (usually around 10 per cent) to determine whether they were interviewed and to ask about interviewers' conduct.

Research in focus 9.2 provides an example of some of the ingredients of research involving multiple interviewers.

Research in focus 9.2
An example of research involving multiple interviewers

This example is taken from the study by Marshall et al. (1988), a team of sociologists from the University of Essex, of social class in modern Britain (see Research in focus 7.4). The interviewing was carried out by a leading independent social research institute, Social and Community Planning Research (SCPR). The research aimed to achieve a sample of 2,000 respondents (1,770 was the number actually achieved; see Research in focus 8.1 for details of the sampling procedure).

> One hundred and twenty-three interviewers were employed on the survey. Six full-time briefing sessions were held, all of which were attended by a member of the Essex team, and interviewers were also given a full set of written instructions. The first three interviews conducted by each interviewer were subjected to an immediate thorough checking in order that critical comments, where appropriate, could be conveyed. During the course of fieldwork the work of interviewers was subject to personal recall. Ten per cent of issued addresses were re-issued for recall . . . In addition, 36 interviewers were accompanied in the field by supervisors . . . (Marshall et al. 1988: 291)

Problems with structured interviewing

While the structured interview is a commonly used method of social research, certain problems associated with it have been identified over the years. These problems are not necessarily unique to the structured interview, in that they can sometimes be attributed to kindred methods, such as the self-completion questionnaire in survey research or even semi-structured interviewing in qualitative research. However, it is common for the structured interview to be seen as a focus for the identification of certain limitations, which are briefly examined below.

Characteristics of interviewers

There is evidence that interviewers' attributes can have an impact on respondents' replies, but, unfortunately, the literature on this issue does not lend itself to definitive generalizations. In large part, this ambiguity in the broader implications of experiments relating to the effects of interviewer characteristics is due to several problems, such as: the problem of disentangling the effects of interviewers' different attributes from each other (race, gender, socio-economic status); the interaction between the characteristics of interviewers and the characteristics of respondents; and the interaction between any effects observed and the topic of the interview. Nonetheless, there is undoubtedly some evidence that effects due to characteristics of interviewers can be discerned.

The ethnicity of interviewers is one area that has attracted some attention. Schuman and Presser (1981) cite a study that asked respondents to nominate two or three of their favourite actors or entertainers. Respondents were much more likely to mention black actors or entertainers when interviewed by black interviewers than when interviewed by white ones. Schuman and Converse (1971) interviewed 619 black Detroiters shortly after Martin Luther King's assassination in 1969. The researchers found significant differences in the answers given between black and white interviewers in around one-quarter of the questions asked.

Although this proportion is quite disturbing, the fact that the majority of questions appear to have been largely unaffected does not give rise to a great deal of confidence that a consistent biasing factor is being uncovered. Similarly inconclusive findings tend to occur in relation to experiments with other sets of characteristics of interviewers. These remarks are not meant to play down the potential significance of interviewers' characteristics for measurement error, but to draw attention to the limitations of drawing conclusive inferences about the evidence. All that needs to be registered at this juncture is that almost certainly the characteristics of interviewers do have an impact on respondents' replies but that the extent and nature of the impact are not clear and are likely to vary from context to context.

Response sets

Some writers have suggested that the structured interview is particularly prone to the operation among respondents of what Webb et al. (1966: 19) call 'response sets', which they define as 'irrelevant but lawful sources of variance'. This form of response bias is especially relevant to multiple-indicator measures (see Chapter 7), where respondents reply to a battery of related questions or items, of the kind found in a Likert scale (see Key concept 7.2). The idea of a response set implies that people respond to the series of questions in a consistent way but one that is irrelevant to the concept being measured. Two of the most prominent types of response set are known as the 'acquiescence' (also known as the 'yeasaying' and 'naysaying' effect) and the **social desirability** effect.

Acquiescence

Acquiescence refers to a tendency for some people consistently to agree or disagree with a set of questions or items. Imagine a respondent who replied with agreement to all the items in Research in focus 7.2. The problem is that agreement with some of the items implies a low level of commitment to work (items 1–4), whereas agreement with others implies a high level of commitment to work (items 5–10). One of the reasons why researchers who employ this kind of multiple-item measure use wordings that imply opposite stances (that is, some items implying a high level of commitment and others implying a low level of commitment to work) is to weed out those respondents who appear to be replying within the framework of an acquiescence response set.

Social desirability bias

The social desirability effect refers to evidence that some respondents' answers to questions are related to their

perception of the social desirability of those answers. An answer that is perceived to be socially desirable is more likely to be endorsed than one that is not. This phenomenon has been demonstrated in studies on mental health using psychiatric inventories. These inventories are meant to be concerned not with chronic mental illness but with minor neuroses and anxieties. Research in New York by Dohrenwend (1966) noted that Puerto Ricans tended to score much higher on the inventory that he administered than other ethnic groups. He found that this tendency was not due to a higher level of mental illness in this ethnic group, but to the effect of social desirability in respondents' answers. Puerto Ricans were much less likely than the other ethnic groups to perceive the items in the inventory as undesirable. This meant that what the researcher had found was a link not between ethnicity and mental health, but between ethnicity and perceptions of the social desirability of mental health inventory items. Later research suggested that variation in the social desirability of mental illness symptoms was related to the perceived prevalence of those symptoms among the respondent's friends and acquaintances (Phillips 1973). The presence of social desirability effects has been demonstrated in other settings (e.g. Arnold and Feldman 1981).

In so far as these forms of response error go undetected, they represent sources of error in the measurement of concepts. However, while some writers have proposed outright condemnation of social research on the basis of evidence of response sets (e.g. Phillips 1973), it is important not to get carried away with such findings. We cannot be sure how prevalent these effects are, and to some extent awareness of them has led to measures to limit their impact on data (for example, by weeding out cases obviously affected by them or by instructing interviewers to limit the possible impact of the social desirability effect by not becoming overly friendly with respondents and by not being judgemental about their replies).

The problem of meaning

A critique of survey interview data and findings gleaned from similar techniques was developed by social scientists influenced by phenomenological and other interpretivist ideas of the kinds touched on in Chapter 2 (Cicourel 1964, 1982; Filmer et al. 1972; Briggs 1986; Mishler 1986). This critique revolves around what is often referred to in a shorthand way as the 'problem of meaning'. The kernel of the argument is that when humans communicate they do so in a way that not only draws on commonly held meanings but also simultaneously creates meanings. 'Meaning' in this sense is something that is worked at and achieved—it is not simply pre-given. Allusions to the problem of meaning in structured interviewing draw attention to the notion that survey researchers presume that interviewer and respondent share the same meanings of terms employed in the interview questions and answers. In fact, the problem of meaning implies that the possibility that interviewer and respondent may not be sharing the same meaning systems and hence imply different things in their use of words is simply sidestepped in structured interview research. The problem of meaning is resolved by ignoring it.

The feminist critique

The feminist critique of structured interviewing is difficult to disentangle from the critique launched against quantitative research in general, which was briefly outlined in Chapter 2. However, for many feminist social researchers the structured interview symbolizes more readily than other methods the limitations of quantitative research, partly because of its prevalence but also partly because of its nature. By 'its nature' is meant the fact that the structured interview epitomizes the asymmetrical relationship between researcher and subject that is seen as an ingredient of quantitative research: the researcher extracts information from the research subject and gives nothing in return. For example, standard textbook advice of the kind provided in this chapter implies that rapport is useful to the interviewer but he or she should guard against becoming too familiar. This means that questions asked by respondents (for example, about the research or about the topic of the research) should be politely but firmly rebuffed on the grounds that too much familiarity should be avoided and because the respondents' subsequent answers may be biased.

This is perfectly valid and appropriate advice from the vantage point of the canons of structured interviewing with its quest for standardization and for valid and reliable data. However, from the perspective of feminism, when women interview women a wedge is hammered between them that, in conjunction with the implication of a hierarchical relationship between the interviewer and respondent, is incompatible with its values. An impression of exploitation is created, but exploitation of women is precisely what feminist social science seeks to fight against. Oakley (1981) found in her research on childbirth that she was frequently asked questions by

her respondents. It was these questions that typified the problems of being a feminist interviewing women.

> The dilemma of a feminist interviewer interviewing women could be summarised by considering the practical application of some of the strategies recommended in the textbooks for meeting interviewees' questions. For example, these advise that such questions as 'Which hole does the baby come out of?' 'Does an epidural ever paralyse women?' and 'Why is it dangerous to leave a small baby alone in the house?' should be met with such responses from the interviewer as 'I guess I haven't thought enough about it to give a good answer right now', or 'a head-shaking gesture which suggests "that's a hard one"' (Goode and Hatt [1952: 198]). (Oakley 1981: 48)

Such advice still appears in textbooks concerned with survey research. For example, Groves et al. (2004: 283) supply the following advice:

1. Interviewers should refrain from expressing views or opinions on the topics covered by the survey instrument.
2. Interviewers should refrain from presenting any personal information that might provide a bias for inferring what their preferences or values might be that are relevant to the content of the interview.
3. Although a little informal chatting about neutral topics, such as the weather or pets, may help to free up communication, for the most part, interviewers should focus on the task.

This is in fact good advice from the point of view of reducing error that might arise from the interviewer influencing or biasing the interviewee's replies. As such, it is likely to reduce error arising from the influence of the interviewer.

Oakley's point is that to act according to such canons of textbook practice would be irresponsible for a feminist in such a situation. It was this kind of critique of structured interviewing and indeed of quantitative research in general that ushered in a period in which a great many feminist social researchers found qualitative research more compatible with their goals and norms. In terms of interviewing, this trend resulted in a preference for forms of interviewing such as unstructured and semi-structured interviewing and focus groups. These will be the focus of later chapters. However, as noted in Chapter 2, there has been some softening of attitudes towards the role of quantitative research among feminist researchers. For example, Walby and Myhill (2001) have shown how surveys of violence against women that are dedicated to uncovering such violence (rather than general crime surveys like the BCS) reveal higher levels than are often thought to occur. By paying greater attention to issues like greater privacy in the interview and special training in sensitive interviewing, dedicated surveys in some countries have proved highly instructive about the causes and incidence of violence against women. Such research, which is based on structured interviewing, would not seem to be inconsistent with the goals of most feminist researchers and indeed may be of considerable significance for many women. Nonetheless, there is still a tendency for qualitative research to remain the preferred research strategy for many feminist researchers.

Key points

- The structured interview is a research instrument that is used to standardize the asking and often the recording of answers in order to keep interviewer-related error to a minimum.
- The structured interview can be administered in person or over the telephone.
- It is important to keep to the wording and order of questions when conducting survey research by structured interview.
- While there is some evidence that interviewers' characteristics can influence respondents' replies, the findings of experiments on this issue are somewhat equivocal.
- Response sets can be damaging to data derived from structured interviews and steps need to be taken to identify respondents exhibiting them.
- The structured interview symbolizes the characteristics of quantitative research that feminist researchers find distasteful: in particular, the lack of reciprocity and the taint of exploitation.

Questions for review

The structured interview

- Why is it important in interviewing for survey research to keep interviewer variability to a minimum?
- How successful is the structured interview in reducing interviewer variability?
- Why might a survey researcher prefer to use a structured rather than an unstructured interview approach for gathering data?
- Why do structured interview schedules typically include mainly closed questions?

Interview contexts

- Are there any circumstances in which it might be preferable to conduct structured interviews with more than one interviewer present?
- 'Given the lower cost of telephone interviewing compared to face-to-face interviews, the former is generally preferable.' Discuss.

Conducting interviews

- Prepare an opening statement for a study of manual workers in a firm, to which access has already been achieved.
- To what extent is rapport an important ingredient of structured interviewing?
- How strong is the evidence that question order can significantly affect answers?
- How strong is the evidence that interviewers' characteristics can significantly affect answers?
- What is the difference between probing and prompting? How important are they and what dangers are lurking with their use?

Problems with structured interviewing

- What are response sets and why are they potentially important?
- What are the main issues that lie behind the critique of structured interviewing by feminist researchers?

Online Resource Centre

www.oxfordtextbooks.co.uk/orc/brymansrm4e/

Visit the Online Resource Centre that accompanies this book to enrich your understanding of structured interviewing. Consult web links, test yourself using multiple choice questions, and gain further guidance and inspiration from the Student Researcher's Toolkit.