Research Methods in Social Psychology
Antony S.R. Manstead
This chapter provides an overview of research methods in social psychology, from the development of theory to the collection of data. After describing three quantitative research strategies (survey research, experiments and quasi-experiments), the chapter briefly discusses qualitative approaches, focusing on discourse analysis. There follows a description of the key elements of experiments and of threats to validity in experimental research, and a discussion of problems with experimental research in social psychology. The final section of the chapter contains a description of three methods of data collection (observation, self-report and implicit measures).

Introduction

How do social psychologists develop their theories?
How do social psychologists go about testing their theories?

Methods provide a means of translating a researcher’s ideas into actions. These ideas usually revolve around one or more questions about a phenomenon. An example of such a question in social psychology would be: ‘How can a group of capable people make a decision that is stupid and could moreover have been shown to be so at the time the decision was taken?’ (see Chapter 11). A researcher might have a hunch about how to explain this phenomenon. For example, the poor decision might have arisen from the fact that the group had a powerful leader who expressed a preference early in the decision-making process and thereby stifled proper evaluation of superior options. To assess the correctness of this hunch the researcher would have to collect information about styles of leadership in groups making poor decisions. Research methods are the procedures the researcher would follow in gathering such information, and methodology is a term used to refer to all aspects of the implementation of methods.

Although this chapter is primarily concerned with the methods used by social psychologists to test the validity of their ideas, it is worth considering where these ideas originate. In the typical case, the researcher begins with a theory about the phenomenon under investigation. Where does such a theory come from? An obvious source is observation of real-life events. Consider Janis’s (1982) theory concerning the poor quality of decision-making that is apparent even in groups of competent and experienced persons. This theory arose from his reading of accounts of how the United States government took the decision to invade Cuba in 1961 (see Research close-up 2.1, pp. 23–24). A second important element of theory building in social psychology is existing theory. The fact that Janis was already conversant with theory and research on group processes and social influence in groups provided him with ideas that he could use to explain defective decision-making by groups.
Another version of this process of theory building begins with a set of apparently conflicting findings from previous research. An example is Zajonc’s (1965) attempt to reconcile conflicting findings in previous studies of the effects on individual task performance of being observed by others (see Chapter 11). Some researchers had found that being observed by others had beneficial effects on task performance, but others had found that it resulted in poorer performance. To reconcile these findings, Zajonc drew on principles derived from learning theory. Once again, the theorist began with a phenomenon that required an explanation, and drew on existing theoretical concepts and processes to make sense of that phenomenon.

In what sense does a theory ‘explain’ a phenomenon such as the defective decision-making of high-calibre groups, or the divergent effects of being observed on task performance? Social psychological theories usually consist of a set of concepts and statements about the relationships among these concepts. For example, Janis’s (1982) theory consists of one set of concepts representing the antecedent conditions of poor group decision-making, another set representing the symptoms of groupthink, a third set representing symptoms of poor decision-making, and a final set representing the process linking antecedent conditions to the symptoms of groupthink and poor decision-making (see Figure 2.1). One of the antecedent conditions is a ‘cohesive group’, a group whose members are psychologically dependent on the group. Because they are dependent on their group membership, they are more likely to conform to what they believe to be the consensual position in the group. An example symptom of groupthink is the presence of ‘mind guards’, a term Janis used to describe group members who take it upon themselves to protect the group from information that questions the correctness or morality of the emerging decision. An example symptom of defective decision-making is failure to examine the risks of the preferred decision. The mediating process specified by Janis is a ‘concurrence-seeking tendency’, a powerful preference for agreement with fellow group members. Thus antecedent conditions are linked to symptoms via a mediating process (see p. 34, below).

Three concepts need to be introduced at this point. **Construct** is the term used to refer to abstract concepts in a theory. In Janis’s theory concepts such as group cohesiveness and concurrence-seeking tendency are theoretical constructs. **Variable** is a term used to refer to a measurable representation of a construct. To represent the construct of group cohesiveness, for example, we might assess one or more of the following: how long the group has been in existence; the extent to which group members nominate each other as personal friends; and how much group members say they value their membership of the group. So there are various ways in which the researcher can represent the

![Figure 2.1](image.png)

**Figure 2.1** Antecedent conditions, mediating process and symptoms of groupthink in Janis’s (1982) theoretical model (based on Figure 10.1 in Janis, 1982).
construct of cohesiveness as a variable. Operationalization refers to the way in which a construct is turned into a measurable variable. If group cohesiveness is measured in terms of how much group members value their membership of the group, this is a different operationalization of cohesiveness than if it is measured in terms of the extent to which group members nominate each other as personal friends.

We can derive predictions from a theory. In the case of Janis’s theory, a prediction that we can logically derive from the theory is that groups that are more cohesive should be more prone to making poor-quality decisions than groups that are less cohesive. Armed with such a prediction (or hypothesis), the researcher tries to find evidence to support the prediction (see Research close-up 2.1).

To the extent that the evidence is consistent with the prediction, confidence in the theory from which the prediction was derived is enhanced. Correspondingly, if the evidence is inconsistent with the prediction, confidence in the underlying theory is weakened. So methods are the means by which researchers put their ideas to the test.

**SUMMARY**

Methods are the tools researchers use to test their theoretical ideas. These ideas can come from a variety of sources, but two that are quite common in social psychology are observations of real-life events and conflicts in previous research findings. A theory consists of a set of constructs linked together in a system, and specifies when particular phenomena should occur.

---

**Archival analyses of ‘groupthink’**


**Introduction**

Janis’s research on groupthink provides an excellent example of ‘archival research’, a research strategy that is not described or discussed elsewhere in the present chapter. In archival research the data come from archives, that is, from stored records of facts. ‘Archival data may include such items as personal documents (letters or diaries), creative products (poems, paintings, essays), biographies or autobiographies, and histories or governmental records’ (Simonton, 1981, p. 218). Janis (1972) decided to study in detail archival material relating to four major US foreign policy fiascoes: the Bay of Pigs invasion of Cuba in 1961; the decision to escalate the Korean War in 1950; the failure to be prepared for the attack on Pearl Harbour in 1941; and the decision to escalate the Vietnam War in 1964. Janis argues that in the case of each of these disastrous decisions, information was potentially or actually available to the policy-making groups that should have led them to different decisions.

**Method**

Janis’s research took the form of careful scouring of all the documentary sources of information on the circumstances in which each of these decisions can be regarded as forming a consistent social psychological pattern, the essence of which is shown in Figure 2.1. Janis (1982) published a second edition of his book in which he applied the notion of groupthink to the Watergate incident that ultimately led to US President Richard Nixon’s resignation in 1974.

**Later research**

Tetlock (1979) conducted a more quantitative analysis of archival materials. He applied standardized procedures for analysing the content of public statements made by key decision-makers involved in the ‘groupthink’ and ‘non-groupthink’ decisions examined by Janis (1972). Tetlock was particularly interested in assessing the extent to which public statements made by key decision-makers reflected ‘a tendency to process policy-relevant information in simplistic and biased ways’ (p. 1317), and the extent to which these statements reflected ‘a tendency to evaluate one’s own group highly positively and to evaluate one’s . . . opponents highly negatively’ (p. 1317). To assess these two aspects of groupthink, Tetlock identified six key decision-makers who were intimately involved in five different foreign policy decisions, two of which were classified by Janis as ‘non-groupthink’, while the other three were classified by Janis as ‘groupthink’ decisions. He then randomly selected 12 paragraph-sized passages from the public statements made by each decision-maker at the time of each crisis for content analysis. He found that the public statements of decision-makers in groupthink crises were characterized by significantly lower levels of ‘integrative complexity’ – a measure of complexity of
RESEARCH STRATEGIES

What are the principal research strategies available to the social psychologist?
What are the strengths and weaknesses of each strategy?

Researchers who want to test their ideas and predictions have different research strategies available to them. It is worth pointing out that although some research strategies will be better suited than others to studying a given phenomenon, each and every strategy, however sophisticated its implementation, has its limitations. It is for this reason that one of the great pioneers of research methodology in the social sciences, Donald Campbell (see Pioneer box, opposite), argued for triangulation. By this he meant that the use of multiple methods to study a given issue would provide a better basis for drawing conclusions than would any single method. The term triangulation comes from navigation: an accurate way to determine the position of a fixed point is by calculating the angles to it from two fixed points that are a known distance apart. Because each method has its own strengths and weaknesses, the use of different methods will help the strengths of one method to compensate for the weaknesses of another, and vice versa.

Many research strategies are available to the social psychologist. Here we will consider three quantitative strategies before briefly considering qualitative research.

Survey research

One strategy for gathering research evidence is to survey public opinion and/or behaviour, by interview or by questionnaire. This strategy is known as survey research (Schwarz, Groves & Schuman, 1998) and is well known in the form of opinion polls. The main
objective is to describe the characteristics of one or more groups of people. Such descriptions can range from the simple (e.g., describing the percentage of persons eligible to vote in a particular constituency who say that they intend to vote for a particular political candidate) to the more complex (e.g., describing the personal and social characteristics associated with illegal use of drugs among school-age children and teenagers). Note that the first type of description is ‘pure’ description, while the second describes relationships between variables – such as those between drug use, on the one hand, and age, sex, socioeconomic status and educational achievement, on the other.

The survey researcher’s primary concern is with the extent to which the respondents are representative of a population (such as all adults living in a particular community, region or country). One way of addressing this issue would be to interview or collect completed questionnaires from the entire population in question (as is done in a census). If you are able to describe the entire population, the findings are by definition ‘representative’ of that population. In most cases, however, collecting data from all members of a population is simply not practicable. Then the researcher has to choose which members of that population to survey. The process of selecting a subset of members is known as sampling.

Two main types of sampling are used in survey research: probabilistic and non-probabilistic. The most basic form of probabilistic sampling is the simple random sample. A simple random sample is one which satisfies two conditions: first, each member of the population has an equal chance of being selected; second, the selection of every possible combination of the desired number of members is equally likely. To explain the second condition, imagine that the population size is 10 (consisting of persons labelled A to J) and the sample size is 2. There are 45 possible combinations of 2 members of the population (A + B, A + C, A + D and so on to I + J). In simple random sampling each of these 45 possible combinations of 2 members has to be equally likely. In practice researchers achieve this by allocating numbers to each member of the population and using computer-generated random numbers to select a sample of the required size (see www.randomizer.org/). So the first randomly generated number defines the first member of the population to be sampled, and so on, until the sample is full.

Because probability sampling is expensive and time-consuming, non-probability sampling is frequently used. The most common form of non-probability sample is the quota sample. Here the objective is to select a sample that reflects basic attributes of the population. Such attributes might be age and sex. If you know the age and sex composition of the population concerned, you then ensure that the age and sex composition of the sample reflects that of the population. The term ‘quota’ refers to the number of persons of a given type (e.g., females between the ages of 55 and 60) who have to be interviewed. The major advantage of quota sampling is that the interviewer can approach potential respondents until the quotas are filled, without needing to recruit a specifically identified respondent.

**Experiments and quasi-experiments**

Experimental research is designed to yield causal information. The goal of an experiment is to see what happens to a phenomenon when the researcher deliberately modifies some feature of the environment in which the phenomenon occurs (‘If I change variable B, will there be resulting changes in variable A?’). By controlling the variation in B, the researcher who finds that there are changes in A can draw causal conclusions. Instead of just knowing that more of variable A is associated with more of variable B, the experimental researcher discovers whether A increases when B is increased, decreases when B is reduced, remains stable when B is left unchanged, and so on. Such a pattern of results would suggest that changes in B cause the changes in A.

The experimental method is a theme with many variations. Two common variations are the quasi-experiment and the true randomized experiment. They differ with respect to the realism of the setting in which the data are collected and the degree of control that the researcher has over that setting. A quasi-experiment is typically conducted in a
CHAPTER 2 RESEARCH METHODS IN SOCIAL PSYCHOLOGY

RESEARCH CLOSE-UP 2.2

A field experiment to study helping behaviour


Introduction

The researchers were interested in testing the idea that one reason why bystanders do not come to the assistance of others, even when these others clearly need help, is that helping is costly. The particular ‘cost’ they studied in their research was time. To come to a stranger’s assistance often involves a departure from your original plan. Such a departure can throw you off your schedule. The researchers also wanted to examine whether reminding people of the parable of the Good Samaritan, in which a passer-by does come to the assistance of a stranger in need of help, would influence willingness to give help. They tested these notions in a field experiment (see also Chapter 9, this volume).

Method

The participants in their study were male seminary students (i.e., trainee priests) who believed that they were taking part in a study on ‘religious education and vocations’. Each participant began the study in one building and was then asked to proceed to a second building to complete the study. Before leaving the first building, the participant was led to believe one of three things about the speed with which he should go to the other building: that there was no special hurry, that there was an intermediate degree of hurry, or that he was late for the second part of the study and should hurry up. This was the manipulation of the first variable, time pressure. In the second part of the study, the participant expected to do one of two things: either talk about the parable of the Good Samaritan or talk about job prospects for seminary students. This constituted the second manipulation: either having or not having the parable of the Good Samaritan made psychologically salient.

On his way to the other building, the participant passed through an alley in which a person (the ‘victim’, but actually an accomplice of the experimenters) was sitting slumped in doorway, head down, eyes closed. As the participant passed the victim, the latter coughed twice and groaned. The dependent variable in this field experiment was the extent to which the participant did anything to help this person apparently in distress. The extent of the participant’s helping behaviour was observed and coded.

Results

Helping was significantly influenced by the time pressure manipulation. Those in the ‘no hurry’ condition were more helpful than those in the ‘intermediate hurry’ condition, who in turn were more helpful than those in the ‘hurry’ condition. There was also a tendency for being reminded about the parable to have an influence. Those who were reminded were more helpful than those who were not.

Discussion

Even those who have chosen to be trained in a vocation in which helping others is supposed to play a central role were affected by the time pressure variable. When they were in a hurry, even those trainee priests who thought that they were on their way to a discussion of the parable of the Good Samaritan were less likely to offer help to a stranger in need than were their counterparts who were in less of a hurry. From a methodological perspective, the neat thing about this experiment is that it was conducted in a natural, everyday setting. Participants were randomly allocated to one of the six conditions of the experiment, so any differences found between these six conditions resulted in principle from the experimental manipulations, so internal validity was high (i.e., the researchers could be confident that changes in the independent variable caused changes in the dependent variable). But the fact that the setting of the experiment was such an everyday one means that this study also scores quite highly on realism. It is a good example of a field experiment.

Plate 2.2 Would you be more likely to help someone in need after hearing a sermon on the parable of the Good Samaritan?
natural, everyday setting, one over which the researcher does not have complete control. The true randomized experiment, by contrast, is one in which the researcher has complete control over key features of the setting; however, this often involves a loss of realism. It is worth emphasizing that it is possible to conduct true experiments in field settings, in which case they are referred to as field experiments, which attempt to combine the control of a laboratory experiment with the realism of a quasi-experiment. An example of such a field experiment is given in Research close-up 2.2.

To grasp the key difference between a quasi-experiment and a true experiment, we need to consider further what is meant by the term experiment. Experiments are studies in which the researcher examines the effects of one class of variables (independent, or manipulated, variables) on another class of variables (dependent, or measured, variables). In a true randomized experiment the researcher has control over the independent variable and over who is exposed to this variable. Most importantly, the researcher is able to allocate research participants randomly to different conditions of the experiment (random allocation). In a quasi-experiment the researcher usually cannot control who is exposed to the independent variable. In a typical quasi-experiment, pre-existing groups of people are either exposed or not exposed to the independent variable. Examples of each method may help to bring out the points of difference.

Social psychologists interested in aggression have studied whether exposure to violent film and television material has an impact on the subsequent behaviour of the viewer (see Chapter 8). This can be done using true randomized experiments or quasi-experiments. An example of a true experiment on this issue is the study reported by Liebert and Baron (1972). Male and female children in each of two age groups were randomly allocated to one of two experimental conditions, one in which they viewed an excerpt from a violent television programme and another in which they viewed an exciting athletics race. Later both groups of children were ostensibly given the opportunity to hurt another child. Those who had seen the violent material were more likely to use this opportunity than were those who had seen the non-violent material. Because children had been allocated to the violent and non-violent conditions randomly, the observed difference can be attributed to the difference in type of material seen, rather than any difference in the type of children who saw the material.

An example of a quasi-experimental study of the same issue is the study reported by Black and Bevan (1992). They asked people to complete a questionnaire measure of tendency to engage in aggressive behaviour under one of four conditions: while waiting in line outside a cinema to see a violent movie; while waiting in line to see a non-violent movie; having just seen a violent movie; and having just seen a non-violent movie. As can be seen in Figure 2.2, the researchers found that those waiting to see the violent film had higher aggression scores than those waiting to see the non-violent film; and also that those who had just seen the violent film scored higher than those waiting to see the violent film, although there was no difference in aggression scores between those who had just seen a non-violent movie and those waiting to see a non-violent movie. These findings are consistent with the notion that viewing a violent movie increases the tendency to aggress, but the fact that participants were not allocated at random to the different conditions makes it impossible to rule out alternative explanations. For example, it may be that violent movies only increase aggressive tendencies among those who are attracted to view such movies in the first place.

Often the only way in which to conduct an experimental study of a social phenomenon is via a quasi-experiment. Ethical and practical considerations frequently make it impossible to allocate people randomly to different experimental conditions. If, like
Stroebe, Stroebe and Domittner (1988), you wish to study the effects of bereavement, for example, you obviously cannot randomly allocate research participants to a ‘bereaved’ and ‘non-bereaved’ condition. The same applies in many other fields of research. Thus the choice of research strategy is often a compromise between what is optimal and what is practicable. Fortunately, the sophistication of some quasi-experimental designs is such that it is possible to draw conclusions about causality with some confidence (Judd & Kenny, 1981a; West, Biesanz & Pitts, 2000).

Qualitative approaches

Traditionally, social psychological research has been quantitative and the overwhelming majority of the research discussed in this book is quantitative in nature. That is, it seeks to operationalize constructs in ways that make them quantifiable, and thereby allow the researcher to describe a variable, or the relationship between two or more variables, in quantitative terms. By contrast, research in other social science disciplines, such as social anthropology, is typically qualitative in nature, meaning that constructs and relationships between constructs are described and discussed using ordinary language. A fundamental assumption shared by qualitative researchers is that we should try to understand the meanings of social behaviours and social experiences from the perspectives of the participants concerned, and that to do this properly we need to pay due regard to the contexts in which these behaviours or experiences naturally occur (see Henwood, 1996).

Although it is common practice for social psychologists to use qualitative methods (such as participant observation or semi-structured interviewing, both described later) in the early stages of their research (for example, to develop and refine the questions to be posed in a questionnaire), there are also social psychologists who exclusively use qualitative methods. They do so in the belief that a quantitative approach provides at best a partial and at worst a distorted view of the phenomenon they want to study. In the context of the present chapter we can do little more than sketch the range of qualitative research methods that can be deployed in social psychological research. Henwood (1996) provides a good overview of the different possibilities within qualitative enquiry. Here we limit ourselves to a brief description of one prominent example of qualitative enquiry, namely discourse analysis. Although this sounds like a single method, it is a term used to describe a family of methods for analysing talk and texts. Discourse analysis starts from the observation that everyday talk is orderly and can be systematically studied through the

Plate 2.4 Are those who choose to see violent films more aggressive?

Figure 2.2 Self-reported tendency to aggress, as a function of type of film, and whether or not the respondent was waiting to see the film or had just seen the film (based on data reported by Black & Bevan, 1992).
transcription and analysis of audiovisual recordings. By recording and analysing everyday interaction and discourse, it is argued, we can gain a better insight into the ways that people conduct their lives and account for themselves.

The discourse analyst seeks to show how, for example, racist or sexist attitudes arise not because of the beliefs or biases of the individual who expresses them, but rather as evaluations that emerge in the context of particular social interactions. Rather than being relatively fixed products of individual cognitive systems, such evaluations arise in the context of conversations and vary according to the particular cultural setting. An example of the use of discourse analysis is the study reported by Wetherell, Stiven and Potter (1987). These researchers were interested in male and female university students’ views about employment opportunities for women. They reasoned that analysing how a group of 17 students talk about these issues would reveal the practical ideologies that are used to reproduce gender inequalities. The students were interviewed in a semi-structured way and their responses were transcribed and analysed. A benefit of this approach is that it enabled the researchers to identify contradictions in the way ordinary people talk about issues like gender inequality. Rather than having a single attitude, the students tended to endorse different positions at different points during the interview, and some of these positions were inconsistent with each other.

This sort of qualitative approach is not represented in the present volume, where the emphasis is on the strengths of a realist, quantifiable social psychology. This is not to say that qualitative methods play no role in the research that is represented in this book. It is more that, as noted above, qualitative methods are used in the early stages of such research, rather than being the sole research method. The role played by qualitative research methods in social psychology largely reflects differences in philosophical belief about the causation of social behaviour. For realist social psychologists, social behaviour has causes, and the goal of research is to identify those causes. This scenario was devised with a view to devising a scenario for which there is a convincing and well-integrated rationale, because the situation should strike participants as realistic and involving, and the experimental manipulations and the measurement process should not ‘jump out’ at the participant. In Milgram’s study, participants were told that the study was an investigation of the effects of punishment on learning. The participant was given, apparently at random, the role of ‘teacher’, while an accomplice of the experimenter posing as another participant (known as a confederate) took the role of ‘learner’. The learner’s task was to memorize a list of word pairs. The teacher’s task was to read out the first word of each pair, to see whether the learner could correctly remember the second word, and to administer a graded series of punishments, in the form of electric shocks of increasing severity, if the learner failed to recall the correct word (which he had been instructed to do from time to time). This scenario was devised with a view to convincing the participant that the shocks were genuine (which they were not), and that the learner was actually receiving the shocks.
The independent variable is the one that is deliberately manipulated by the experimenter. All other aspects of the scenario are held constant, and the independent variable is changed systematically. Each change produces a new ‘condition’ of the experiment: one change yields two conditions, two changes yield three conditions, and so on. In Milgram’s research a key independent variable was the proximity of the ‘learner’ to the ‘teacher’. In one condition, learner and teacher were in separate rooms; in a second condition, the teacher could hear the learner but could not see him; in a third condition, the teacher could both see and hear the learner’s reactions; in a fourth condition, the teacher had to hold the learner’s hand down on a metal plate in order for the shock to be delivered (the touch-proximity condition). All other aspects of the experimental setting were held constant, so that variations in the teachers’ behaviour in these four conditions were attributable to the change in proximity between teacher and learner.

The success of an experiment often hinges on the effectiveness of manipulations of the independent variable. By effectiveness we mean (1) the extent to which changes in the independent variable capture the essential qualities of the construct that is theoretically expected to have a causal influence on behaviour, and (2) the size of the changes that are introduced. For example, in Milgram’s study, we should consider how well the four proximity conditions capture the construct of proximity. What is being manipulated, clearly, is physical proximity. Then there is the question of whether the changes between the four conditions are sufficiently large to produce an effect. In this case it is hard to see how the proximity variable could have been manipulated more powerfully; an investigator who adopts weaker manipulations runs the risk of failing to find the predicted effects simply because the variations across levels of the independent variable are too subtle to have an impact. It has become standard practice in social psychological experiments to include among the measured variables one or more measures of the effectiveness of the manipulation: these are known as manipulation checks.

Assessing whether an independent variable has had an effect requires the measurement of the participant’s behaviour or internal state. This measured variable is known as the dependent variable, so called because systematic changes in this measured variable depend upon the impact of the independent variable. In Milgram’s study, the dependent variable was the intensity of shocks in a 30-step sequence that the teacher was prepared to deliver. The results of Milgram’s experiments are often expressed in terms of the percentage of participants who gave the maximum shock level (corresponding to 450 volts). The results of the Milgram (1965) study are shown in these terms in Figure 2.3. A key question to ask of any dependent variable is the extent to which it is a good measure of the underlying theoretical construct. In addition to this question of the ‘fit’ between a theoretical construct and the measured or dependent variable, the most important issue involved in designing dependent variables is what type of measure to use. We will discuss this in more detail below.

Laboratory experiments often involve deception, in the sense that the participant is misled about some aspect of the research. The extent of this deception can range from withholding information about the purpose of the research to misleading participants into thinking that the research is concerned with something other than its real purpose. The main reason for using deception is that participants would act differently if they were aware of the true objective of the study. If Milgram’s participants had known that his was a study of obedience, we can be sure that the rate of disobedience would have been higher: the participants would have wanted to demonstrate their ability to resist orders to harm a fellow human. Attitudes to the use of deception in social psychological research have changed during the past 40 years: misleading participants about the nature of an experiment is now viewed more negatively. The reason for this change is partly moral (i.e., where possible one should avoid deceiving someone else, whether or not in the context of an experiment) and partly practical (if participants are routinely misled about research, they will enter any future participation in the expectation that they are going to be misled, which may influence their behaviour). Striking an appropriate balance between being completely honest with participants and wanting to study them free of the influence of their knowledge of the nature of the experiment is difficult. Most universities in Europe, North America and Australasia have some form of ethics committee that monitors research involving human participants, and national bodies such as the American Psychological Association (APA) and the British Psychological Society (BPS) have published guidelines concerning research using human participants that have to be followed by their members.
One way to address ethical issues entailed in using deception is by careful use of **debriefing**. This takes place at the end of the experimental session and involves informing the participant as fully as possible about the nature and purpose of the experiment, and the reason for any deception. In Milgram’s study, for example, care was taken to assure participants that the ‘shocks’ they had administered were in fact bogus, and that the learner had not been harmed in any way; the reason for the deception was also carefully explained. Ideally, the debriefing process should leave participants understanding the purpose of the research, satisfied with their role in the experiment, and with as much self-respect as they had before participating in the study.

### Experimental designs

**Why is it important to have a control condition in an experiment? What is an interaction effect?**

As we have seen, it is important that participants are allocated randomly to the different conditions of an experiment. Failure to achieve this goal hinders the researcher’s ability to conclude that observed differences between conditions in the dependent variable result from changes in the independent variable. We shall now examine more closely the question of designing experiments in order to rule out alternative inferences as far as possible.

First, consider a study that may appear to be an experiment but cannot properly be described as experimental. This is the **one-shot case study**. Following Cook and Campbell (1979), we shall use the symbols X to stand for a manipulation (i.e., of the independent variable) and O to stand for observation (i.e., the dependent variable). In these terms the one-shot design looks like this:

\[
\begin{array}{c}
X \\
\text{time}
\end{array}
\]

To take an example, imagine that an educational researcher wanted to know the effect of a new teaching method on learning. The researcher takes a class of students, introduces the new method (X) and measures the students’ comprehension of the taught material (O). What conclusions can be drawn from such a design? Strictly speaking, none, for there is nothing with which O can be compared, so the researcher cannot infer whether the observed comprehension is good, poor or indifferent.

A simple extension of the one-shot design provides the **minimum requirements** for a true experimental study and is known as the **post-test only control group design**. Let R stand for random assignment of participants to conditions, and X and O stand for manipulation and observation, as before. This design looks like this:

\[
\begin{array}{c}
\text{Experimental group} \\
\text{Control group}
\end{array}
\]

\[
\begin{array}{c}
X \\
O_1 \\
O_2
\end{array}
\]

\[
\text{time}
\]

Here there are two conditions. In the experimental condition participants are exposed to the manipulation (participants in this condition are known as the **experimental group**), and possible effects of the manipulation are measured. In the control condition there is no manipulation (here the participants are known as the **control group**), but these participants are also assessed on the same dependent variable and at the same time point as the experimental group. Now the observation made in the experimental condition (O1) can be compared with something: the observation made in the control condition (O2). So the researcher might compare one group of students who have been exposed to the new teaching method with another group who continued to receive the normal method, with respect to their comprehension of the course material. An important point is that participants are randomly allocated to the two conditions, ruling out the possibility that differences between O1 and O2 are due to differences between the two groups of participants that were present before X was implemented. It follows that if O1 and O2 differ markedly it is reasonable to infer that X causes this difference.

There are several other more sophisticated and complex **factorial experiments**, each representing a more complete attempt to rule out the possibility that observed differences between conditions result from something other than the manipulation of the independent variable (see Cook & Campbell, 1979). A common design in social psychological experiments is the **factorial experiment**, in which two or more independent variables are manipulated within the same study. The **experimental group** refers to participants who are exposed to a condition of the experiment, i.e., the variable(s) used in experimental research. Measures of the dependent variable derived from these participants are compared with those derived from participants who are exposed to the independent variable (i.e., the experimental group), providing a basis for inferring whether the dependent variable determines scores on the dependent variable.
Arguments in favour of making the university examination system tougher. The second variable was involvement, i.e., whether the participants, who were students, thought that the university would introduce the tougher exam system next year, such that it would affect them personally (high involvement), or in the next decade, such that it would not affect them personally (low involvement). According to the elaboration likelihood model, argument quality should have a stronger impact on attitudes when participants are involved with the message topic than when they are not. Figure 2.4 shows some of the key findings from Petty et al.’s (1981) study. It can be seen that the effect of argument quality on attitudes was much greater when involvement was high than when it was low, just as the theory predicts. Because the predicted effect is an interaction, testing this prediction requires a factorial design.

**Figure 2.4** Interaction between argument quality and involvement, showing that argument quality had a much stronger effect on attitudes when involvement was high (based on data reported by Petty, Cacioppo & Goldman, 1981).

Threats to validity in experimental research

What is the difference between internal and external validity? What is meant by the term ‘confounding’ in the context of experimental research?

In a research context, validity refers to the extent to which one is justified in drawing inferences from one’s findings. Experimental research attempts to maximize each of three types of validity: internal validity, construct validity and external validity.

**Internal validity** refers to the validity of the conclusion that an observed relationship between independent and dependent variables reflects a causal relationship, and is promoted by the use of a sound experimental design. We have already seen that the use of a control group greatly enhances internal validity, but even if one uses a control group there remain many potential threats to internal validity (Brewer, 2000; Cook & Campbell, 1979). Prime among these is the possibility that the groups being compared differ with respect to more than the independent variable of interest.

For example, let’s assume that in Milgram’s obedience research a different experimenter had been used for each of the four conditions described earlier, such that experimenter 1 ran all participants in one condition, experimenter 2 ran all participants in another condition, and so on. It might seem sensible to divide the work among different experimenters, but to do so in this way poses a major threat to the internal validity of the experiment. This is because the four conditions would no longer differ solely in terms of the physical proximity of the ‘victim’; they would also differ in that different experimenters conducted them. Thus the differing amounts of obedience observed in the four conditions might reflect the impact of the physical proximity variable, or the influence of the different experimenters (or, indeed, some combination of these two factors). The problem is that the physical

The separate effects of each independent variable are known as **main effects**. If the combined effect of two independent variables differs from the sum of their two main effects, this is known as an **interaction effect**.

A factorial design contains all possible combinations of the independent variables. In the design shown above, each independent variable has two levels, resulting in four conditions. The main benefit of a factorial design is that it allows the researcher to examine the separate and combined effects of two or more independent variables. The separate effects of each independent variable are known as **main effects**. If the combined effect of two independent variables differs from the sum of their two main effects, this is known as an **interaction effect**.

To illustrate an interaction effect, let us consider a study of the effects of persuasive communications on attitude change, reported by Petty, Cacioppo and Goldman (1981). To test Petty and Cacioppo’s (1986a) elaboration likelihood model of persuasion (see Chapter 7), these researchers manipulated two variables. The first was argument quality, i.e., whether the persuasive communication the participants read consisted of strong or weak arguments.
confounding a variable that incorporates two or more potentially separable components is a confounded variable. When an independent variable is confounded, the researcher's ability to draw unambiguous causal inferences is seriously constrained.

construct validity the validity of the assumption that independent and dependent variables adequately capture the abstract variables (constructs) they are supposed to represent.

Demand characteristics are cues that convey the experimenter’s hypothesis to the participant. Individuals know that they are being studied will often have hunches about what the experimenter is expecting to find. They may then attempt to provide the expected responses. When behaviour is enacted with the intention of fulfilling the experimenter’s hypotheses, it is said to be a response to the demand characteristics of the experiment. Orne (1962, 1969) conducted much research on demand characteristics and suggested ways of pinpointing the role they play in any given experimental situation. For example, he advocated the use of post-experimental enquiry, in the form of an interview, preferably conducted by someone other than the experimenter, the object being to elicit from participants what they believed to be the aim of the experiment and the extent to which this affected their behaviour. Clearly, researchers should do all they can to minimize the operation of demand characteristics, for example by using unobtrusive measures, that is, measures that are so unobtrusive that participants are unaware of the fact that they are being taken, or by telling participants that the purpose of the experiment cannot be revealed until the end of the study and that in the meantime it is important that they do not attempt to guess the hypothesis. A cover story that leads participants to believe that the purpose of the study is something other than the real purpose is a widely used means of lessening the impact of demand characteristics.

Experimenter expectancy refers to the experimenter’s own hypothesis or expectations about the outcome of the research. This expectancy can unintentionally influence the experimenter’s behaviour towards participants in such a way as to enhance the likelihood that they will confirm his or her hypothesis. Rosenthal (1966) called this type of influence the experimenter expectancy effect. The processes mediating experimenter expectancy effects are complex, but non-verbal communication is centrally involved. An obvious strategy for reducing these effects is to keep experimenters ‘blind’ to the hypothesis under test, or at least blind to the condition to which a given participant has been allocated; other possibilities include minimizing the interaction between experimenter and participant, and automating the experiment as far as possible. Indeed, in much current social psychological research, the entire experiment, including all instructions to the participants, is presented via a computer. This obviously limits the opportunity for experimenters to communicate their expectancies.

Even if the experimenter manages to avoid all these threats to internal and construct validity, an important question remains: to what extent can the causal relationship between X and O be generalized beyond the circumstances of the experiment? External validity refers to the generalizability of a finding beyond the circumstances in which it was observed by the researcher. One important feature of the experimental circumstances, of course, is the type of person who participates in the experiment. In many cases participants volunteer their participation, and to establish external validity is important to consider whether results obtained using volunteers can be generalized to other populations. There is a good deal of research on differences between volunteers and non-volunteers in psychological studies (see Rosenthal & Rosnow, 1975). The general conclusion is that...
there are systematic personality differences between volunteers and non-volunteers. Such findings are explained in terms of volunteers’ supposedly greater sensitivity to and willingness to comply with demand characteristics. The external validity of studies based on volunteers’ behaviour is therefore open to question, and the solution to this problem is to use a ‘captive’ population, preferably in a field setting.

Another criticism of social (and indeed other) psychological experiments is that the participants are often university students. Sears (1986) examined research articles published in major social psychology journals in 1985 and found that 74 per cent were conducted with student participants. Although students are certainly unrepresentative of the general population, being younger, more intelligent and more highly educated than the average citizen, this in itself is not a threat to the validity of the research. This is because the goal of much social psychological research is to understand the process(es) underlying a phenomenon (such as attitude change or stereotyping), rather than to describe the general population (a goal for which survey research is much better suited). In any case, there is often little reason to suppose that the processes underlying a phenomenon such as attitude change or stereotyping differ in some fundamental way between students and non-students.

**Social psychological experiments on the Internet**

**What are the advantages and disadvantages of web-based experiments?**

A relatively new development in psychological research is the use of the Internet to recruit and conduct experiments (Internet experiments). People are invited to participate in the research by visiting a website where the server runs the whole study, from allocating participants to an experimental condition to debriefing them about the nature and purpose of the study once they have completed the experimental task. Birnbaum (2000) noted that the number of experiments listed on sites such as the one maintained by the American Psychological Society (psych.hanover.edu/Research/exponnet.html) has grown very rapidly, by around 100 per cent per year, and that many of these studies are social psychological.

What are the primary advantages and disadvantages of such web-based experiments? A major advantage is the ease with which quite large amounts of data can be collected in a relatively short time. Other advantages are that participants are recruited from different countries, from different age groups and – to the extent that access to the Internet becomes more widespread – from different socioeconomic backgrounds. Obvious disadvantages are that the researcher loses a degree of control over the running of the experiment. Participants complete the study in different physical settings, at different times of the day and night, and probably with differing levels of motivation and seriousness. There are also issues to do with the representativeness of those who choose to participate in an Internet study (they tend to be white, from the USA or from Europe, and to be relatively young – but not as young as those who take part in laboratory experiments) and with the effect of linguistic competence on the reliability and validity of responses (most studies posted on the web are in English, and although the majority of respondents tend to be from the USA or other English-speaking countries, some are not).

Despite the potential problems associated with running experiments on the web, the evidence suggests that Internet studies yield results that parallel those of conventional experiments (see Krantz & Dalal, 2000). It is clear that this way of conducting experiments is going to continue to expand very rapidly. Before embarking on such research it is important to consult sources such as Nosek, Banaji and Greenwald (2002) and Reips (2002), who offer advice about how best to avoid the potential pitfalls.

**Problems with experimentation**

*What are the main criticisms that have been levelled at the use of experiments in social psychology?*

*What is meant by the term ‘mediation’ in the context of psychological research?*

It is widely assumed that the experimental method provides the ‘royal road’ to causal inference (Aronson et al., 1998). In fact causal inference from the results of experiments is more problematic than some commentators allow. One problem concerns what Gergen (1978) has called the cultural embeddedness of social events, by which he means that a laboratory experimental demonstration that independent variable X has an impact on dependent variable O needs to be qualified by adding that the circumstances in which X was manipulated may have played a role in producing the observed effects on O. Smith and Bond (1998) review many social psychological experiments, including the Milgram obedience experiment, that have been conducted in different countries. It is not unusual for these experiments to produce different findings as a function of the cultural setting.

A related problem noted by Gergen is that although the experimental method supposedly allows us to trace the causal sequence from antecedent conditions to the behaviour of interest, its capacity to do so depends on the assumption that external events are related in a one-to-one fashion with particular states or processes in the individual. The result is that what one experimenter believes to be a demonstration of the effect of X on O via the mediating process Z, another will prefer to explain in terms of a different process. Social psychology abounds with such debates between rival accounts for findings (for examples, see Tetlock & Levi, 1982; Tetlock & Manstead, 1985), and some have come to the view that experimentation is not a suitable way to settle such between-theory disputes.

The heart of the problem identified by Gergen is that phenomena of interest to social psychologists often entail chains of events. If we strip this issue down to its bare essentials, we can ask whether variable X influences variable O directly, or whether the relation between X and O is mediated by another variable, Z. By conducting an experiment we may establish that there is a causal relation between X and O; but had we also measured Z, we might
have found that the relation between X and Z is also very high, as is the relation between Z and O. Indeed, we might find that once the X–Z and Z–O relationships are statistically taken into account, the originally established relationship between X and O disappears. This is the type of situation in which one can infer that the relationship between X and O is mediated by Z (Baron & Kenny, 1986).

Indeed, one strategy that helps to overcome the problem of alternative explanations identified by Gergen is to design experiments that include the assessment of possible mediating variables. In modern social psychological research, researchers often attempt to measure such variables and then to conduct mediational analysis, for which there are well-established procedures (see Judd & Kenny, 1981b; Kenny, Kashy & Bolger, 1998).

A final problem worth mentioning is that although the ostensible goal of social psychological experimentation is the accumulation of scientific knowledge, in the form of laws or principles of social behaviour that are valid across time, there is some reason to doubt whether experimentation (or, indeed, any other method) is capable of generating evidence that could be the basis of such laws. To understand why this is the case in social sciences but not in natural sciences, bear in mind that the relationship between researcher and the object of study is radically different in these two types of science. Testing of theories in the natural sciences is concerned with the analysis and explanation of the object world, a world that does not engage in the construction and interpretation of the meaning of its own activity. The objects of investigation in social sciences are people, who do of course attribute meaning and significance to their own actions. Social psychology cannot therefore be neatly separated from what it studies. Laypersons are able to acquire social psychological knowledge and use it to modify their actions in a way that atoms, elements and particles cannot.

One implication of this is that even well-supported social psychological theories should not be regarded as embodying ‘laws’ that hold good across time: if learning about a theory leads people to modify the behaviour that the theory tries to explain, the theory has limited temporal validity. Gergen (1973, 1978) has been the leading advocate of this sobering view, although others, including Schlenker (1974), have challenged his arguments. It is also worth noting that some of the problems of accumulation of knowledge in social psychology can be addressed through the use of meta-analysis. This is a technique for statistically integrating the results of independent studies of the same phenomenon in order to establish whether findings are reliable across a number of independent investigations (see Cooper, 1990; Hedges & Olkin, 1985; Johnson & Eagly, 2000). The increasing use of meta-analysis in social psychology (where relevant, one is cited in every chapter of this book) has shown, without doubt, that many social psychological claims have, in fact, been confirmed over multiple experiments, often conducted over many decades. This accumulation of evidence does not support Gergen’s claim.

What are the implications of these problems for the status of experimentation in social psychological research? Even some of the harshest critics of the experimental approach do not advocate the abandonment of experimentation. For example, Gergen acknowledged that experiments would continue to play an important role in the explication of the relationship between biological processes (such as physiological arousal) and social behaviour; that studies such as the Milgram experiment are useful for raising consciousness about the insidious nature of social influence; that experiments can increase the impact of theories by providing vivid demonstrations of conditions under which a theory makes successful predictions; and that experimentation can be useful to evaluate social reforms, such as the effectiveness of measures designed to conserve energy. Thus the debate about the utility of experimentation revolves around the types of inference that can reasonably be made on the basis of experimental evidence, with ‘traditionalists’ such as Aronson et al. (1998) sticking to the view that experimentation provides a firm basis on which to build knowledge, and critics such as Gergen questioning this assumption. Given that over 30 years have now elapsed since Gergen’s critique, and experimental social psychology continues to grow and flourish, we can conclude in any case that experiments have not been abandoned.

**SUMMARY**

We examined different aspects of the use of experimentation in social psychology. We began by describing the principal features of the social psychological experiment, before going on to discuss some common experimental designs. We then considered the main threats to validity in experimental research, such as demand characteristics and experimenter expectancy effects, before going on to describe how researchers are making increasing use of the Internet to conduct experiments. Finally, we considered some possible problems with the use of experiments in social psychological research.

**DATA COLLECTION TECHNIQUES**

What are the principal data collection techniques used in social psychological research?

**What are the strengths and weaknesses of each of these techniques?**

Assuming that an investigator is conducting quantitative research, he or she will need to measure one or more variables, regardless of which research strategy has been adopted. In correlational designs
the researcher has to measure each of the variables that are expected to correlate. In experimental designs the researcher needs to measure the dependent variable. In either case, the investigator is confronted with the task of translating a theoretical construct (for example, aggression) into a measurable variable (for example, willingness to harm someone). Any psychological measure should be both reliable and valid. reliability refers to the stability of the measure. If you measure an adult’s height, the measurement will be highly stable from one day to the next and will also be independent of who is doing the measuring.

A reliable measure is one that is not dependent on the time of measurement or on the person taking the measurement. A measure can be highly reliable and yet be low in validity. To pursue the height example, let us imagine that what you really want to measure is a person’s weight. In the absence of a proper weighing scale you decide to measure height instead, because you do have a tape-measure. Of course, height and weight are correlated with each other, so height may be a better estimate of weight than simple guesswork. But clearly, height is not especially valid as a measure of weight. So validity in this context refers to the extent to which the measured variable really captures the underlying construct.

In social psychological research the investigator typically chooses to measure a variable using one or more of the following: observational measures, self-report measures or (a more recent development) implicit measures.

Observational measures

If the object of one’s research is to collect information about social behaviour, an obvious means of doing so is by observation. Many behaviours of interest to social psychologists are detectable without sophisticated equipment and take place in public settings, which makes them suitable for observation. Although observational methods vary in kind from the informal and unstructured to the highly formal and structured, the object in each case is the same: to abstract from the complex flux of social behaviour those actions that are of potential significance to the research question, and to record each instance of such actions over some period (Weick, 1985).

Sometimes the nature of the research setting or topic dictates that observation is conducted in a relatively informal and unstructured manner, with the researcher posing as a member of the group being observed. A classic example of research employing this method is Festinger, Riecken and Schachter’s (1956) study of the consequences of blatant disconfirmation of strongly held beliefs. The investigators identified a religious sect that predicted that the northern hemisphere would be destroyed by flood on a certain date. By joining that sect, members of the research team were able to observe what happened when the predicted events failed to materialize. Under such circumstances, observation clearly has to be covert and informal: if other sect members suspected that the researchers were not bona fide believers, the opportunity for observation would be removed. This type of observation is known as participant observation, and typically yields qualitative data.

More formal methods of observation can be used when it is possible to record actions relevant to the research question without disrupting the occurrence of the behaviour. An example is Carey’s (1978) series of studies investigating the hypothesis that when one pedestrian approaches another on the street, a rule of ‘civil inattention’ applies, whereby each looks at the other up to the point where they are approximately 8 feet apart, after which their gaze is averted. Goffman (1963) first advanced this hypothesis on the basis of informal observation. Carey’s purpose was to verify, using more formal methods, the existence of this rule, and to establish parameters such as the distance between pedestrians when gaze is first averted. He covertly photographed pairs of pedestrians as they approached and passed each other on a street, taking the photographs from upper storeys of buildings overlooking the street. The resulting photographs were coded for variables such as distance between the pair, whether their heads and eyelids were level or lowered, and whether gaze direction was towards or away from the approaching person.

The two examples cited above have in common the fact that the targets of the researchers’ observations were unaware that they were being observed. Although such failure to inform persons of their involuntary participation in a research project raises ethical questions, it does overcome a problem peculiar to any research that uses humans as participants, namely the tendency for the measurement process itself to have an impact on participants’ behaviour, a phenomenon known as reactivity. It is well established that the knowledge that one is being observed can influence behaviour. A well-known instance of such an effect is a study of worker productivity conducted at the Hawthorne plant of the Western Electric Company (Roethlisberger & Dickson, 1939), where it was found that merely observing workers raised their motivation and thereby increased productivity. Instances of such influence have come to be known as Hawthorne effects. Awareness of this problem has led many researchers to develop unobtrusive methods of observing and measuring behaviour. Webb, Campbell, Schwartz and Sechrest (2000) compiled a useful sourcebook of methods of unobtrusive measurement.

The most formal type of observational method is one in which the researcher uses a predetermined category system for scoring social behaviour. A well-known example of such a system is Bales’s (1950) interaction process analysis (IPA),
developed to study interaction in small social groups. Here the verbal exchanges between group members are coded in terms of 12 predetermined categories (e.g., ‘requests information’; see Chapter 12, this volume). The scores of group members can then be used to determine (among other things) who is the leader of the group (see Bales & Slater, 1955). Further examples of observational coding schemes can be found in Bakeman (2000).

Observational methods of data collection have two main advantages over the self-report methods we shall consider below: first, they can often be made unobtrusively; second, even where the participant knows that his or her behaviour is being observed, enacting the behaviour is typically quite engaging, with the result that participants have less opportunity to modify their behaviour than they would when completing a questionnaire. Nevertheless, there are some types of behaviour that are either difficult to observe directly (because they are normally enacted in private) or impossible to observe directly (because they took place in the past). Moreover, social psychologists are often interested in measuring people’s perceptions, cognitions or evaluations, none of which can be directly assessed simply through observation. For these reasons, researchers often make use of self-report measures.

**Self-report measures**

The essential feature of data collection using self-report measures is that questions about the participant’s beliefs, attitudes and behaviour are put directly to the participant. The responses are self-report data. Self-report measurement is usually quicker, cheaper and easier to use than observational measurement. The researcher does not have to contrive a laboratory setting or find a natural setting in which to observe a behavioural response; furthermore, there is typically no need to train observers or to use recording equipment, for self-reports are normally recorded by the participant. Finally, as noted above, some of the variables that are of most significance to social psychologists are not directly observable. For these reasons, self-report measurement is very common in social psychological research, and it is not unusual for studies to depend exclusively on self-report data.

There are two principal methods of collecting self-report data: the questionnaire and the interview. In the questionnaire method, participants are handed a set of questions, along with instructions on how to record their answers. In the interview method, questions are put to the participant by an interviewer, who then records the participant’s responses. Interviewing is particularly useful when there is reason to believe that the questions might be difficult to understand without clarification. A tactful and sensitive interviewer should be able to establish rapport and ensure that the respondent fully comprehends a question before answering. Another advantage of interviewing is that interviews can be ‘semi-structured’, meaning that although the interviewer has a set series of topics to be covered in the interview, he or she is able to vary the specific questions that are asked so that they are relevant to the unfolding discussion. However, interviewing is costly in terms of time and money, and a poorly trained interviewer can easily bias the respondent’s answers by hinting at a desired or socially acceptable response. Questionnaires are especially useful for gathering data from large numbers of participants with minimal expense, and the comparative anonymity of the process is preferable when the questions touch on sensitive issues. On the other hand, many people who are given questionnaires fail to complete and/or return them. Response rates for questionnaires sent by mail to randomly selected names and addresses vary between 10 and 50 per cent. Because there is always the danger that non-respondents differ systematically from respondents in some respect, low response rates are undesirable.

Devising a good questionnaire or interview schedule is a harder task than one might imagine. As with any psychological measure, the goal is to produce questions that are reliable and valid. Although there are many potential sources of unreliability in the construction of questionnaires, the most serious threat to reliability is ambiguity: if a question is ambiguous, different respondents may interpret it differently and therefore provide answers to what is in effect a different question. The most serious threat to question validity is failure on the part of the investigator to have specific objectives for each question: the hazier the intent of the researcher in posing a particular question, the greater are the chances that it will fail to elicit information relevant to his or her objectives. However, there are other sources of unreliability and invalidity that cannot easily be controlled. A simple rule-of-thumb is never to assume that answers to a single question will reliably or validly measure a construct. If two or more items are used to measure that construct, the factors that decrease reliability and validity of responses to any single question should cancel each other out, so a measure based on the average of the responses to the different items will be a more reliable measure of the underlying construct.

Because it is difficult to envisage all the potential pitfalls in questionnaire construction, there is no substitute for pilot work in which drafts of the final questionnaire are administered to participants whose answers and comments provide a basis for revision. Constructing an entirely new questionnaire can therefore be a time-consuming and painstaking process. Fortunately, there are collections of previously developed and pre-tested questionnaires, such as the one edited by Robinson, Shaver and Wrightsman (1991). It is worth checking such a source before setting out to construct an original questionnaire. If no suitable questionnaire already exists, the researcher should consult a text on questionnaire design such as the one by Oppenheim (1992) before devising a fresh questionnaire.

Self-report measures have several advantages. What are their drawbacks? Obviously, it is not possible to collect self-report data completely unobtrusively: participants are aware that they are under investigation, and may modify their responses as a result of this awareness. In particular, there is ample opportunity for the respondent’s answers to be influenced by motivational factors, such as social desirability. There is no simple solution to this difficulty, although there are steps that can be taken which reduce the scale of the problem. First, it is worth emphasizing to participants whenever possible that their responses are anonymous. Second, it is worth stressing the point that there are no right or wrong answers. Third, it is often possible to increase participants’ motivation to respond truthfully by treating them as research accomplices rather than ‘guinea-pigs’.
Implicit measures

A recent development in social psychological research methods has been the increasing use of techniques for measuring perceptions, cognitions and evaluations that do not rely on the usual type of self-report measure, thereby avoiding the disadvantages of the latter. These techniques are often referred to as implicit measures (Greenwald & Banaji, 1995). The use of such measures has quite a long history in social psychology: Campbell (1950) published a classic paper on the indirect assessment of attitudes more than half a century ago. What is different about the modern use of implicit measures is that they usually take advantage of computer technology. Here computers are used not only for the presentation of experimental materials but also (and more importantly) for the precise measurement of various aspects of the participants’ responses to these materials. An example of an implicit measure is the use of response latencies (i.e., how long it takes a participant to answer a particular question). Such measures can provide fresh insights into cognitive structures and processes. For instance, in the study reported by Gaertner and McLaughlin (1983), the automatic operation of stereotypes was assessed by the speed (response latency) with which participants made judgements about pairs of words. The participants’ task was to say ‘yes’ if there was an association between each pair of words. White participants responded significantly faster to white-positive word pairs (e.g., white–smart) than black-positive word pairs (e.g., black–smart), thereby suggesting that they engaged in automatic stereotyping of racial groups. Examples of the use of implicit measures to assess attitudes can be found in Chapter 6.

A major advantage of implicit measures is that they are not reactive. That is, implicit measures are not subject to biases such as social desirability and demand characteristics, because they tap processes that operate outside awareness. However, it is by no means certain that such measures have high validity. How does one know, for example, whether a fast reaction time reflects automatic stereotyping as opposed to individual differences in lexical knowledge? To address questions such as this, one ideally needs to have other measures (e.g., observational) that provide evidence that converges with the evidence provided by implicit measures. In principle such evidence helps to establish the convergent validity of both types of measure. Convergent validity is established when different operationalizations of the same construct produce the same results. However, the argument that implicit measures tap processes in a way that is less subject to the influence of self-presentational concerns than are other measures (especially self-report) obviously raises some tricky issues with regard to cross-validating one measure by means of another.

Another key advantage of implicit measures is that they can assess constructs and processes that may be outside the awareness of the individual. If people are not aware of having certain thoughts or feelings, they will by definition be unable to report them, even if they are highly motivated to be honest. The study of ‘automatic’ processes has become a central theme in social cognition research (see Bargh & Chartrand, 2000). Given that one of the attributes of an automatic process is that the individual is unaware of it, studying such a process requires the use of implicit measurement.

Choosing a measure

All three types of measure considered here have certain advantages and disadvantages. Although there are no hard-and-fast rules for choosing one type of measure rather than the other, two points should be borne in mind when judging the appropriateness of a measure. First, the three types of measure – observational, self-report and implicit – can be used in conjunction with each other in many types of research. Second, the three types of measure differ in terms of the type of information they yield. If observational, self-report and implicit measures of the same conceptual variable point to the same conclusion, this clearly enhances confidence in that conclusion. Furthermore, self-report measures often assess the outcome of a process; by using observational and implicit measures as well, the researcher can gain insight into the processes responsible for that outcome. A special quality of implicit measures is that they enable researchers to capture aspects of the individual’s thoughts, feelings and behaviour that are outside awareness and therefore not susceptible to feigning.
SUMMARY

We examined the main data collection techniques available to the social psychological researcher. Three such techniques were identified: observational, self-report and implicit measurement. We noted that each technique has its own advantages and disadvantages, and that there is often a case for using more than one type of measure in a piece of research.

SUMMARY AND CONCLUSIONS

- Research methods are the procedures a researcher uses to gather information, and methodology is a term used to refer to all aspects of the implementation of methods.
- The information gathered using research methods is used to test the researcher’s theoretical predictions. These predictions are derived from a theory. The theory is often generated through observation of real-life events or by trying to make sense of puzzling findings from previous research.
- We drew a distinction between research strategies and data collection techniques. We described three quantitative research strategies: survey research, quasi-experiments and true randomized experiments. Two key ways in which these strategies differ are in terms of (1) the degree to which one is able to generalize to a population and (2) the degree to which one can draw inferences about causality.
- We briefly discussed qualitative research methods, noting that these are often used by researchers who believe that quantitative methods are unsuited to studying the phenomenon under investigation. Discourse analysis was identified as a popular qualitative approach. Discourse analysis emphasizes the importance of how social phenomena are constructed through discourse.
- Experimentation was singled out for more detailed discussion because of its prominence as a research strategy in social psychology during the last six decades. The main features of experimentation were identified as: the experimental scenario; the independent variable; the dependent variable; the manipulation check; and debriefing.
- A true experimental design is one that enables the researcher to infer that changes in the independent variable produce changes in the dependent variable. Such a design must therefore incorporate more than one condition, allowing the researcher to compare observations made under different conditions.
- The minimal true experimental design is the post-test only control group design, in which participants are randomly allocated to one of two conditions, only one of which involves being exposed to the manipulation. Several more complex designs are available, and of these the factorial design is very commonly used, mainly because of its ability to test predictions concerning interaction effects.
- Drawing strong inferences from social psychological research depends on three types of validity: internal, construct and external. We identified confounding as a threat to internal validity; social desirability effects, demand characteristics and experimenter effects as threats to construct validity; and volunteer/non-volunteer differences as a threat to external validity.
- The Internet has provided social (and other) psychologists with a new arena in which to conduct experiments, enabling them to reach larger and more diverse groups of participants. The evidence to date suggests that despite the potential problems of web-based experiments, their results tend to parallel those obtained using conventional methods.
- We noted that some social psychologists have questioned the utility of experiments. The cultural embeddedness of social behaviour, the fact that social behaviour is determined by multiple factors, and the ability of humans to modify their behaviour in the light of social psychological theories were identified as grounds for questioning the assumption that experimentation generates cumulative knowledge of the laws governing social behaviour.
- We identified three principal methods of collecting data in social psychological research: observational measurement, self-report measurement and implicit measures. Observational and implicit measures have the advantage of being less susceptible to social desirability effects, and can be made completely unobtrusive. However, observational measures are obviously limited to phenomena that can be observed and are not suited to the assessment of covert cognitive phenomena such as attitudes, causal attributions and stereotypes (see Chapters 3, 6, 7 and 14, this volume).
- To study these more covert phenomena researchers have traditionally relied on self-report measures, although there has been an increasing tendency to make use of implicit measures, the goal of which is to reveal phenomena that may either be outside the awareness of the individual or are likely to be misreported in conventional self-report measures due to social desirability concerns.
- There are obvious advantages in using these different types of measure in conjunction with each other.
Suggestions for further reading


