4 Methodology in Social Psychology: Turning Ideas into Actions

A. S. R. MANSTEAD and G. R. SEMIN

Introduction

Procedures for gathering information in any discipline are known as methods. Methods provide a means of translating a researcher’s ideas into actions. The researcher’s ideas will generally 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 intelligent people make a decision that is stupid and could have been shown to be so at the time the decision was taken?’ (cf. Janis, 1972). A researcher interested in this question might have a hunch or a theory to explain this phenomenon. For example, it might be thought that the poor decision arises from the fact that the group has a powerful leader who expresses a preference early in the decision-making process and thereby stifles systematic evaluation of superior options. Assessing the correctness of this hunch would necessitate the collection of information about styles of leadership in groups making poor decisions. Methods are the procedures the researcher would follow in gathering that information, and methodology is a term used to refer to all aspects of the implementation of methods.

The specific method adopted by a researcher in conducting a particular investigation depends a good deal on the type of question he or she is trying to answer. An analogy may be helpful here. Imagine an individual who wants to cross a river. Various options are available: swimming, rowing or sailing a boat, floating a raft, building a bridge, hang-gliding, flying and so on. The task of crossing the river is analogous to the research question being addressed, and the various means of crossing the river are analogous to the research methods available. The means of crossing chosen by the individual will depend on factors such as the number of people who wish to cross the river, the frequency with which they want to make crossings, the prevailing weather and current
conditions, and so on. Similarly, the type of research method chosen by the investigator will depend on the ultimate goal of the research. If the research goal is to describe some phenomenon (e.g. whether it exists, how and where it is found, and so on), this implies a different type of method to the one suggested by the research goal of testing one or more explanations for or hypotheses about a phenomenon.

Research: descriptive, correlational and experimental

It will be useful at this point to distinguish between three different types of research: descriptive, correlational and experimental.

Descriptive research This is intended to provide the researcher with an accurate description of the phenomenon in question (‘Does A occur?’). For example, the researcher may want to know (as did Milgram, 1963) whether the average adult would obey orders from an authority figure to administer painful and potentially lethal electric shocks to a fellow human. The researcher would begin by observing and recording the proportion of adults that obeys an authority’s orders. This simply describes the phenomenon. Social-psychological research rarely stops at this point. The researcher typically wants to know why people behave as they do. If one finds, as did Milgram, that 65 percent of a sample of adults are fully obedient to orders to administer shocks, an obvious question is ‘Why?’

Correlational research The correlational approach takes us part of the way to answering this question. The goal here is to describe the extent to which variations in some behaviour, such as obedience, are related systematically to variations in some other factor (‘Is A related to B?’). For example, do those who obey the orders tend to be particular types of person (men rather than women, introverts rather than extroverts, and so on)? In posing such questions, the researcher is looking for relationships, or correlations, in the information that he or she collects. Discovering such a relationship can be helpful in working out why a phenomenon occurs, but correlational information is rarely decisive in this respect. To understand why this is so, let us take the case of a correlational finding from Milgram’s (1965) study of obedience. It was found that persons who were more obedient tended also to report experiencing more tension during their participation in the study. How should we interpret this correlation? Is the tension a sign of fear of the consequences of disobedience, which would suggest that obedience is ‘caused’ by individuals’ fears about what might happen to them if they were disobedient? Alternatively, might the tension simply reflect concern about the possible harm befalling the ‘victim’? In the first case the relationship between obedience (A) and tension (B) is explained as ‘B leads to A’; in the second case, the same relationship is interpreted as ‘A leads to B’. In the absence of further information, either interpretation is plausible. This is why it is almost always impossible to infer causality from correlational research.
Experimental research This is explicitly designed to yield causal information. The goal of an experiment is to see what happens to a phenomenon, such as obedience, when the researcher deliberately modifies some feature of the environment in which the phenomenon occurs ("If I change B, will there be resulting changes in A?"). By controlling the variation in B, the researcher can arrive at stronger conclusions about causality if it is found that A and B are related. Instead of simply knowing that more A is associated with more 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 the manipulated variations in B cause the observed variations in A. We shall have quite a lot more to say about experimental research below.

The descriptive, correlational and experimental research types are very general kinds of research method, and are by no means specific to psychology or social psychology. Our main purpose in this chapter is to introduce the reader more specifically to the research methods most often employed in social-psychological research. The principal aim is to enable the reader to evaluate social-psychological research; a secondary aim is to provide some preliminary guidance for the conduct of research. To facilitate the process of describing and discussing research methods, we shall consider separately two facets of research methodology. First, we shall describe various research strategies; by research strategy we mean the broad orientation one adopts in addressing a question. To pursue our river-crossing analogy, the decision to adopt a research strategy is akin to the decision to swim, float, sail, straddle or fly across the river. Next, we shall describe some of the most popular data collection techniques; these are the specific procedures the researcher follows in gathering information. Deciding which technique to adopt is akin to deciding how exactly to construct the raft, boat, bridge or whatever, once a general strategy has been chosen. The selection of a particular technique will be determined partly by the researcher's objectives and partly by the available resources, just as the decision to construct a solid bridge rather than rope bridge depends partly on how many people are expected to use the bridge (and how often), and partly on the availability of skill and raw materials.

Selection of Research Strategies

The strategies available for social-psychological research differ in terms of the degree of control and the amount of precision they offer, the realism of the situations in which data are collected, and so on. In this section, we shall provide an overview of the major research strategies and briefly describe their distinctive attributes, following a scheme devised by Runkel and McGrath (1972, pp. 81ff).

Systematic observations of phenomena in real-life settings are field studies (Bickman and Henchey, 1972). If some property of the real-life setting is
deliberately modified, as when an interracial housing policy is implemented on
a housing estate and levels of prejudice are compared with those on an
equivalent housing estate where no such policy is introduced, the strategy is
called a field experiment. If, however, the ecology of the life setting is
contrived, as in a university laboratory, and specific properties of this ecology
are deliberately modified, this strategy is known as a laboratory experiment
(Aronson, Brewer and Carlsmith, 1985). There is also a 'middle road' between
field and laboratory experiments in which the investigator devises a situation
that simulates key features of a naturally occurring setting (such as an
emergency situation) in a laboratory, and this type of strategy is known as an
experimental simulation (Abelson, 1968). The major difference between a
laboratory experiment and an experimental simulation is that in the former the
investigator attempts to create a setting which, although contrived, incor-
porates features believed to be common to a large class of settings. In the
experimental simulation, by contrast, the investigator attempts to mimic the
essential properties of one type of naturally occurring situation.

A different strategy for gathering information is to survey public opinion,
either by interview or by questionnaire. Here the investigator is not concerned
about the setting in which the data are collected, because this is assumed to be
irrelevant. On the other hand, the investigator is typically concerned with the
extent to which the respondents are representative of a population (such as all
adults living in a particular community, region or country). This type of
research strategy is known as a sample survey, and is well known to us in the
form of opinion polls (Schuman and Kalton, 1985).

Another strategy where the setting in which the data are collected is of no
great concern to the investigator is referred to by Runkel and McGrath as a
judgement task (Rosenthal, 1982). Here, as in the sample survey, the investi-
gator is concerned with the subject's report or judgement about some issue or
problem. Judgement tasks differ from sample surveys in that the investigator
exercises more control over the issues or problems being judged. In a sample
survey of attitudes towards different political candidates, for example, the
respondent might simply be asked to indicate which candidate he or she likes
most, would be likely to vote for, and so on. In a judgement task, by contrast,
the respondent might be asked to identify in what way two of three candidates
are alike, and in what way they differ from the third candidate. In the first case
the investigator's concern is to describe which candidate is most preferred by
some population (such as all adults eligible to vote in a particular constituency);
in the second case, the investigator's concern is to uncover the dimensions
along which respondents distinguish the candidates.

Runkel and McGrath identify two further research strategies, which they
term 'formal theory' and 'computer simulation'. These strategies differ from
the others we have considered in that they do not involve observation or
measurement of actual behaviour; that is they are non-empirical strategies.
In using formal theory the investigator proceeds by building a symbolic system
of interconnected statements, assumptions and postulates, and uses deduc-
tive logic to derive some consequences which should map on to empirical
observations. There are plenty of examples of the use of formal theory in social psychology; one of the better known is Zajonc’s (1965) drive theory of social facilitation effects. Social facilitation is the impact of the presence of other persons on task performance. Before 1965 researchers were puzzled by the fact that in some experiments this impact was found to be beneficial, but in others it was found to be detrimental. Indeed, research on social facilitation had virtually ground to a halt by the mid 1960s, bogged down in this morass of apparently inconsistent research findings. The utility of theory is demonstrated by the fact that Zajonc’s theoretical account for these findings stimulated an upsurge of research interest in this topic (see chapter 14). However, formal theory is rarely, if ever employed by social psychologists as an independent research strategy, because theory is only useful to the extent that it connects with empirical observations.

The other non-empirical research strategy – computer simulation – has become increasingly popular, particularly in cognitive science and cognitive psychology. The use of computer simulation is much less common in social psychology, although some cognitively oriented social psychologists (e.g. Abelson, 1968; Schank and Abelson, 1977) have used this strategy. Runkel and McGrath point out that the relationship between formal theory and computer simulation is similar to that between laboratory experiments and experimental simulations, in that computer simulations are in effect formal theories about concrete systems, stated in propositional form. Instead of assessing the applicability of the theory by examining its fit with empirical evidence, the computer simulation attempts to model the properties and dynamics of cognitive or behavioural systems.

A diagrammatic summary of these eight research strategies is shown in figure 4.1. Rather than describing each strategy superficially, we shall confine ourselves to a detailed consideration of experimental methods, on the grounds that experimentation has been the dominant methodology in social psychology for the past three decades.

Before examining experimental methods, however, it is worth noting some points of similarity and dissimilarity between the eight strategies. First, note that the circle shown in figure 4.1 is divided in half, both vertically and horizontally. Beginning with the vertical division, it can be seen that Runkel and McGrath consider the four strategies in the left-hand half of the figure to be ones that deal with universal behaviour systems, and the four strategies in the right-hand half of the figure to deal with particular behaviour systems. What is meant by this is that strategies of the first type attempt to study behaviour without reference to the particular context in which it is observed or measured; strategies of the latter type, by contrast, are ones in which the investigator studies behaviour in a specific and concrete context, with a view to understanding something about that particular behaviour setting.

Turning now to the horizontal division, it can be seen that the four strategies in the upper half of the figure are classed as obtrusive research operations and those in the lower half as unobtrusive research operations. What is meant by this is that strategies of the first type intrude into the normal course of events,
by manipulating and controlling aspects of the environment in which the subject behaves. Strategies of the latter type, by contrast, are either minimally obtrusive or completely unobtrusive, in that the research process does not disturb the natural course of events.

This brief overview of the alternative strategies available to the social psychologist should provide a general idea of the way in which the different strategies relate to one another, and of their relative advantages and disadvantages. In closing this section, we should point out that the strategies should be regarded as complementary. A good example of that complementarity comes from the field of group decision-making, where laboratory experiments (e.g. Fraser, Gouge and Billig, 1971) have revealed a systematic tendency for groups of persons to arrive at a more extreme or polarized decision on some issue than the average of the individual decisions made by those same group members (see chapter 15 for further information). In spite of the obvious real-world applicability of this finding, there has been surprisingly little field research on this issue. In one of the few published field studies, Semin and Glendon (1973) capitalized on the decision-making procedures used by a real managerial team, which happened to correspond very closely to the procedures followed in laboratory experiments on group polarization. Thus they were able to conduct a field experiment without interfering unduly with the normal decision-making procedure of this managerial group. They found no differences between the average of the individual decisions made by members of the managerial team
and the subsequent decisions made by the group as a whole. In attempting to explain the absence of the phenomenon found so consistently in laboratory experiments, Semin and Glendon pinpointed various factors that are characteristics of natural decision-making groups but are absent in experimental groups studied in the laboratory. This pinpointing of factors permits the researcher to modify the conditions of laboratory experiments until they too yield no evidence of polarization, and thereby enhances the understanding of the general phenomenon. Thus laboratory and field experimentation used in conjunction with each other can provide a more comprehensive answer to a research question than either strategy used alone.

**Experimental methods**

As noted above, experimentation has been the dominant research method in social psychology. One of the major advantages ascribed to experimentation is that it is the ideal method for providing unambiguous evidence about causal relationships among variables, and is therefore without parallel as a method for testing theories that predict such relationships. Standard guides to research in social psychology (e.g. Carlsmith, Aronson and Ellsworth, 1976; Aronson, Brewer and Carlsmith, 1985) tend to treat experimentation as the preferred research method. In fact there are grounds for questioning the extent to which experimental studies provide unambiguous evidence about causation, as we shall see. However, first we shall describe the principal features of the experimental approach to social-psychological research.

To assist the process of description, we shall use one experiment as an illustrative example. The work in question is the well-known study of obedience conducted by Milgram (1965), already referred to at the beginning of this chapter (also see chapter 15 for a fuller discussion of this study).

**The experimental scenario** The experimental scenario is the package in which the study is presented. In a field experiment the scenario should, of course, be one that occurs naturally, without contrivance on the experimenter's part. In laboratory settings, however, it is essential to devise 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 'leap out' at the subject. There is a sense in which the typical laboratory experiment is like staging a play, with the exception that the subject's lines are more or less unscripted. In the case of Milgram's study, the scenario presented to subjects was that of an investigation of the effects of punishment on learning. The subject was allocated, apparently at random, the role of 'teacher', while an accomplice of the experimenter, posing as another subject, took the role of 'learner'. The learner's ostensible task was to memorize a list of word pairs. The teacher's task was to read out the first word of each pair, 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). The experimental scenario was set up with a view to convincing the subject that the shocks were genuine (which they were not), and that the learner was indeed a fellow subject who was actually receiving the shocks. Thus what was actually a study of the extent to which subjects would obey the experimenter’s instruction to deliver steadily increasing electric shocks was presented as a study of the effects of punishment on learning.

The independent variable The independent variable is the variable that is deliberately manipulated by the experimenter. All other aspects of the experimental scenario are held constant, and the independent variable is changed in some respect with a view to assessing the consequences of this manipulation. Each change in the independent variable produces a new ‘condition’ of the experiment: one change yields two conditions, two changes yield three conditions, and so on. For example, in Milgram’s study a key independent variable was the proximity of the ‘learner’ to the ‘teacher’. In one condition, learner and teacher were in separate rooms, and the teacher could not hear or see the learner’s reactions to the shocks; in a second condition, the teacher could hear the learner, but still could not see him; in the third condition, the teacher could both see and hear the learner’s reactions; in the 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, see plate 4.1). All other aspects of the experimental setting were held constant, so that variations in the teacher-subjects’ behaviour in these four different conditions should have been attributable only to the change in proximity between teacher and learner. The adequacy 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 variable 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 notion of proximity. What is being manipulated, quite clearly, is physical proximity (rather than, say, psychological proximity); as long as this is what the experimenter intends to manipulate, all well and good. We should also consider whether the changes between the four conditions are sufficiently large to have an effect. In this particular case, it is difficult to see how the proximity variable could have been manipulated more powerfully, but 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 any impact.

Another important point in this context is that subjects should be randomly allocated to different experimental conditions. Failure to adhere to this stipulation interferes with one’s ability to draw causal inferences from the results. For example, in the four conditions of Milgram’s study described above, it was found that the number of shocks teacher-subjects were prepared to administer steadily declined as the proximity between teacher and learner
increased. This appears to show that obedience to the experimenter’s instructions diminished as the learner’s suffering became more salient to the teacher. Such an inference could not be drawn if there were grounds for thinking that the type of subject recruited for the four conditions differed in some systematic way. If, for example, there was any possibility that subjects in the high-proximity conditions were less ‘punitive’ in their attitudes or personalities than subjects in other conditions, this would be a plausible alternative explanation for the findings.
The dependent variable  Assessing the impact of an independent variable requires the experimenter to measure some feature of the subject’s behaviour or internal state. This measured variable is known as the dependent variable, so called because systematic changes in this measured variable should depend upon the impact of the independent variable. In the Milgram study, the dependent variable was the number of shocks in a 30-step sequence that the teacher was prepared to deliver by throwing switches corresponding to each shock level. A key question to ask of any dependent variable is the extent to which it is a good measure of the underlying theoretical variable. For example, is the willingness to deliver what appear to be increasingly strong shocks to another person a good measure of ‘destructive obedience’? In addition to this question of the ‘fit’ between a theoretical variable and the measured or dependent variable, the most important issue involved in designing dependent variables is what type of measure to use. This is a matter that will be discussed in some detail later in this chapter.

The follow-up  The follow-up is the phase of data collection that takes place after the independent variable has been introduced and the dependent variable has been measured. Typically, the follow-up takes the form of a post-experimental enquiry (usually in the form of an interview or questionnaire), followed by debriefing. The prime objectives of the post-experimental enquiry are (1) to check the effectiveness of the experimental manipulations, by measuring subjects’ perceptions of those features of the experimental scenario that are relevant to that manipulation; and (2) to establish whether the subject has suffered any ill-effects as a result of his or her participation in the experiment. In Milgram’s study particular attention was paid to the latter point, in view of the stressfulness of the experimental procedure and the serious way in which subjects were deceived. An example of a post-experimental enquiry designed to check the effectiveness of experimental manipulations is that used by Parkinson and Manstead (1981). Their experiment was intended to assess the effects of paying attention to or ignoring sounds that purportedly represented the subject’s heartbeat; one of the questions in the post-experimental questionnaire asked subjects to rate the amount of attention they had paid to these sounds. For the manipulation to be deemed effective, subjects in the ‘attend’ condition should have made higher ratings than subjects in the ‘ignore’ condition – which they did. Debriefing the subject refers to the process of informing the subject as fully as possible about the nature and purpose of the experiment, and the role their particular participation played in the study as a whole. Although this process is a good idea in any context, it is particularly important wherever the subject has been deceived about the purpose of the experiment and/or about some aspects of the experimental procedure. In Milgram’s study, for example, care was taken to assure subjects that the ‘shocks’ they had administered were in fact bogus, and that the learner had not been harmed in any way. Ideally, the debriefing process should leave subjects happy with their role in the experiment and with as much self-respect as they had on entering the laboratory.
Experimental designs

We have already seen that it is important (1) that experimenters keep all theoretically irrelevant features of the experimental setting constant across conditions, manipulating just the key independent variable, and (2) that subjects are allocated randomly to the different conditions of an experiment. Failure to achieve these goals hinders the researcher’s ability to draw the inference that observed differences in the dependent variable across conditions result from changes in the independent variable. We shall now examine more closely the question of designing experiments in such a way that alternative inferences are ruled out.

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

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

To take a concrete 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’ recall of the taught material (O). What conclusions can be drawn from such a design? Strictly speaking, the answer is none; the point is that there is nothing with which O can be compared, so the researcher cannot infer whether the observed recall is good, poor or indifferent.

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

\[
\begin{array}{c}
R \\
X \\
O_1 \\
\hline \text{time} \\
R \\
O_2
\end{array}
\]

Compared with the one-shot design, there are two important modifications. First, there are two conditions. In one the subjects are exposed to the manipulation (this is usually referred to as the experimental condition, and subjects in this condition are known as the experimental group), and possible effects of the manipulation are measured. In the other no manipulation is introduced (this is usually referred to as the control condition, and subjects in this condition are known as the control group, but these subjects are also measured 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, namely the observation made in the control condition (O2). The second important modification is that in this design subjects 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 subjects that were present before X was implemented. It follows that if O1 and O2 differ markedly, it is reasonable to infer that this difference is caused by X.

Although the post-test only control group design is one of the more commonly used experimental designs in social psychology, there are several other more sophisticated and complex designs, 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 and Campbell, 1979 for a full discussion). The prime object of experimental design, then, is to enhance the validity of the researcher's inference that differences in the dependent variable result from changes in the independent variable.

Finally, it is worth noting that although using a control group is a basic step that enables the researcher to infer that a given manipulation has a measurable effect, in practice researchers sometimes dispense with the use of a control group in which subjects are not exposed to any manipulation, and instead use two or more conditions which vary with respect to the size or strength of the manipulation to which subjects are exposed. An example of this is the Milgram experiment referred to above, in which the proximity of the learner to the teacher was progressively increased across four experimental conditions. Comparing the amount of obedience observed in these conditions permitted Milgram to draw inferences about the effect of proximity on obedience. The underlying point, then, is that the use of sound experimental designs enables the researcher to compare observations made under different conditions, and thereby to draw inferences about causal relationships.

**Threats to validity in experimental research**

Good experimental research maximizes each of three types of validity: internal validity, construct validity and external validity.

**Internal validity** This is promoted by the use of a sound experimental design. Internal validity refers to the validity of the conclusion that an observed relationship between independent and dependent variables reflects a causal relationship. 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 (Cook and Campbell, 1979 list 13!). Among these is the possibility that the groups being compared differ with respect to more than the independent variable of interest.

For example, assume for one moment that in the experiment described previously Milgram had used a different experimenter for each of the four
conditions, such that experimenter 1 ran all subjects in one condition, experimenter 2 ran all subjects in the next condition, and so on. Although it might seem sensible to divide the work of running the conditions among different experimenters, 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 each would be conducted by a different experimenter. Thus the differing amounts of obedience observed in the four conditions might reflect the causal influence of the physical proximity independent variable, or the influence of the different experimenters (or, indeed, some combination of these two factors). The problem is that the physical proximity variable would be confounded with a second variable, namely experimenter identity. It is impossible to disentangle the effects of confounded variables.

Construct validity Even when we are confident that the relationship between X and O is a causal one, in the sense that internal validity is high, we need to consider carefully the nature of the constructs involved in this relationship. Construct validity refers to the validity of the assumption that independent or dependent variables adequately capture the variables (or ‘constructs’) they are supposed to represent.

With regard to the construct validity of independent variables, the issue is whether the experimental manipulation really operationalizes the intended theoretical construct. For example, in a well-known experiment Aronson and Mills (1959) found that subjects who underwent a severe initiation in order to join what turned out to be a tedious discussion group subsequently reported greater liking for that group than did subjects who underwent a milder initiation. This was interpreted as evidence in support of a prediction derived from dissonance theory (see chapter 8). According to dissonance theory, the knowledge that one has suffered in order to attain a goal is inconsistent with the knowledge that the goal turns out to be worthless, thereby generating cognitive dissonance. To reduce this uncomfortable state of dissonance, it is argued, the individual re-evaluates the goal more positively. Gerard and Mathewson (1966) pointed out that Aronson and Carlsmith’s findings are in fact open to a variety of interpretations, all of which accept that the initiation manipulation used by Aronson and Carlsmith was responsible for the observed differences in liking for the discussion group, but assert that this effect resulted from something other than the differing amounts of dissonance supposedly experienced by the two groups of subjects. Accordingly, Gerard and Mathewson conducted a modified replication of the original experiment, effectively ruling out these alternative interpretations.

Even if the researcher has cause to feel satisfied with the construct validity of the independent variable, there remains the question of whether the measured variables actually assess what they were intended to assess. As we shall see below, devising a measure to capture the essence of a social-psychological construct is by no means straightforward. There are three main types of threat to the construct validity of dependent variables in social-psychological experi-
mentation: social desirability, demand characteristics and experimenter expectancy.

Social desirability is a term used to describe the fact that subjects are usually keen to be seen in a positive light, and may therefore be loath to provide honest reports of fears, anxieties, feelings of hostility or prejudice, or any other quality which they think would be regarded negatively. Equally, subjects may 'censor' some of their behaviours so as to avoid being evaluated negatively. To the extent that a researcher's measures are contaminated by social desirability effects, they will obviously be failing to assess the theoretical construct of interest. The most obvious means of reducing social desirability effects is to make the measurement process as unobtrusive as possible, on the premise that if subjects do not know what it is that is being measured, they will be unable to modify their behaviour.

Demand characteristics are cues in the experimental setting which convey to the subject the nature of the experimenter's hypothesis. The point here is that individuals who know that they are being studied will often be curious about what the experimenter is looking at and what types of responses are expected. Subjects may then attempt to provide the expected responses in order to please the experimenter. 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) has conducted a great deal of research into demand characteristics, and has suggested various methods of pinpointing the role they play in any given experimental situation. For example, he advocates the use of in-depth post-experimental enquiry in the form of an interview, preferably conducted by someone other than the experimenter, the object of which is to elicit from the subject what he or she believed to be the aim of the experiment, and the extent to which this belief affected behaviour in the experiment. Clearly, researchers should do all they can to minimize the operation of demand characteristics, for example by using unobtrusive measures, or by telling subjects 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 which leads subjects 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. However, an unconvincing cover story can create more problems than it solves, raising doubts in the mind of the subject that otherwise may not have arisen.

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 subjects 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. The extent to which experimenter expectancy can influence a phenomenon may be assessed by using several experimenters and manipulating their expectations about the
experimental outcome. An obvious strategy for reducing these effects is to keep experimenters ‘blind’ to the hypothesis under test; other possibilities include minimizing the interaction between experimenter and subject, and automating the experiment as far as possible. The goal in each case is to reduce the opportunity for the experimenter to communicate his or her expectancies.

**External validity** Even if the experimenter manages to circumvent all the above threats to internal and external validity, an important question concerning validity remains: to what extent can the causal relationship between X and O be generalized beyond the particular circumstances of the experiment? **External validity** refers to the generalizability of an observed relationship beyond the specific 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 subjects volunteer their participation, and to establish external validity it 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 and Rosnow, 1975 for a review; and Cowles and Davis, 1987 for a recent study). The general conclusion is that there are systematic personality differences between volunteers and non-volunteers. More importantly, in studies such as the one reported by Horowitz (1969) it has been found that the effects of some manipulations used in attitude change research were actually opposite for 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 only 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.

**Data Collection Techniques**

Whichever research strategy is adopted by an investigator, he or she will need to measure one or more variables. 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 conceptual variable (for example, aggression or attraction) into a measurable variable (for example, willingness to harm someone, or willingness to help someone). The researcher’s initial goal, then, is to specify what it is that he or she wants to record in order to represent the conceptual variable in a meaningful way. For example, is willingness to deliver a painful shock to another person, expressed behaviourally, a representative index of aggression as conceptualized by the investigator, or would it be better to adopt another index, such as the number of verbal insults directed at the person? In social-psychological research the
investigator typically chooses to record a variable using either observational measures or self-report 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 that are of interest to social psychologists are detectable without the assistance of sophisticated equipment and are enacted in public settings, which makes them suitable for observation. Although observational methods vary in kind, as we shall see, from the relatively 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 which 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, for the obvious reason that the observer participates in the activities of the group that is being observed.

Rather 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 eight feet apart, after which their gaze is averted. This hypothesis was first advanced by Goffman (1963), 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 still photographs were then 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 may raise tricky ethical questions, it does overcome a problem peculiar to any research that uses humans as subjects, namely the tendency for the measurement process itself to have an impact on subjects' behaviour. It is well established that the simple knowledge that one is being observed can influence behaviour enacted in front of observers. The best known instance of such an effect is a study of worker productivity conducted at the Hawthorne plant of the Western Electric Company (Roethlisberger and Dickson, 1939), where it was found that merely observing workers raised their motivation and thereby increased productivity. Although this was not the first instance of researchers being aware that observation can itself influence the behaviour being observed, 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. An entertaining and very useful sourcebook of methods of unobtrusive measurement has been compiled by Webb et al. (1981).

The most formal type of observational method is one in which the researcher uses a category system for scoring social behaviour. A well-known example of such a system is Bales's (1950b) interaction process analysis (IPA), developed to study interaction in small social groups. IPA consists of the 12 categories shown in figure 4.2. The observer's task in using this system is to concentrate on the verbal interaction taking place between members of a group, and to place individual statements or 'thought units' into one of the 12 categories, noting at the same time who made the statement and to whom it was directed. Such a system should be simple enough for codings to be made in real time, general enough to be applicable to most types of group, and yet specific enough to tap important facets of verbal interaction. The IPA system is a fairly successful one, judged by these criteria, but some of its limitations are apparent if one remembers that non-verbal behaviour (widely acknowledged to be an important feature of interaction – see chapter 9) is almost totally ignored.

Observational methods of data collection have two prime advantages over the self-report methods we shall consider below: first, they can often be made unobtrusively; second, even where the subject knows that his or her behaviour is being observed, enacting the behaviour is typically quite engrossing, with the result that subjects have less opportunity to modify their behaviour than they would when completing a questionnaire. Nevertheless, there are some types of behaviour that are either impossible to observe directly (because they took place in the past) or difficult to observe directly (because they are normally enacted in private). 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 subject’s beliefs, attitudes, behaviour or whatever are put directly to the subject. His or her responses constitute 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 usually recorded by the subject in the form of written responses. Finally, as noted above, some of the variables that are of most significance to social psychologists are not directly observable. For these several 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. As we shall see, however, self-report measures are not without problems.

There are two principal methods of collecting self-report data: the questionnaire and the interview. In the questionnaire method, the subjects are handed a set of questions, along with instructions on how to record their answers. In the interview method, questions are put to the subject by an interviewer, who then records the subject'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 with the respondent and ensure that the latter fully comprehends a question before answering. On the other hand, interviewing is a costly procedure 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 subjects with minimal expense, and the comparative anonymity of the process might be 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. In practice, social psychologists often manage to get round this problem by administering their questionnaires to subjects who are in some sense 'captive', in that they have already volunteered to participate in the study, and by having them complete the questionnaire in a lecture theatre or laboratory rather than letting them take it home.

Questionnaires are undoubtedly the most widely used form of data collection in social-psychological research. Some idea of the richness and variety of data collected exclusively by means of questionnaires can be gained by consulting the study reported by Folkman and Lazarus (1985). These investigators used questionnaire techniques to study how people appraised a stressful event (a university examination), what emotions they experienced as the event approached and passed, and how they coped with the stress induced by the event. It is difficult to envisage how Folkman and Lazarus could have conducted such a study without using questionnaires. It is certainly possible to measure some psycho-physiological indices of stress, such as heart rate, before, during and after exposure to a noxious stimulus such as an electric shock; but one cannot assume that the short-term stress induced by shock in a laboratory is comparable with the longer-term stress induced by 'natural' events such as
examinations, ill health, divorce or bereavement. Furthermore, the individual’s appraisals, emotions and coping strategies could not be assessed satisfactorily without the use of self-report measures.

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**, in that they would evoke the same response from a given individual if he or she were tested more than once under similar circumstances, and **valid**, in that they measure exactly what the researcher intends them to measure. Although there are many potential sources of unreliability in the construction of questionnaires, the most serious threat to question reliability is **ambiguity**: if a question is ambiguous, a given respondent might well interpret it differently on different occasions and therefore give different answers. 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. Because it is difficult to envisage all the potential pitfalls in questionnaire construction, there is no substitute for pilot work in which prototypes of the final questionnaire are administered to groups of subjects whose answers and comments provide a basis for revision. Constructing an entirely fresh questionnaire can therefore be a time-consuming and painstaking process. Fortunately, there are collections of previously developed and pre-tested questionnaires, such as those edited by Shaw and Wright (1967) and Robinson and Shaver (1969). It is worth checking these sources before setting out to construct an original questionnaire. If no suitable questionnaire already exists, the researcher should consult texts on questionnaire design such as Oppenheim (1966) and Payne (1951) before devising a fresh questionnaire.

As we have seen, self-report measures have several advantages; what are the drawbacks? Chief among these is the fact that it is not possible to collect self-report data completely unobtrusively: subjects are always 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. To the extent that these motivations bias the subject’s responses, the self-report measure will provide a distorted reflection of his or her beliefs, behaviours, or whatever. There is no simple solution to this problem, although there are some steps that can be taken which together should reduce the scale of the problem. First, it is worth emphasizing to subjects 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 subjects’ motivation to respond truthfully by treating them as research accomplices rather than ‘guinea-pigs’.

**Choosing a measure**

As we have seen, both 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, there are two points that should be borne in mind when judging the appropriateness of a dependent measure. First, the two types of measure can be used in conjunction with each other in many types of research. Second, the two types of measure differ in terms of the type of information they yield. Let us consider each of these points more closely.

Assume that you wish to study interpersonal attraction. Under laboratory or field conditions you introduce two people, previously unknown to each other, and ask them to get to know each other in the course of a 15 minute discussion. If you want to measure how much these two like each other at the end of the session, you could simply depend on self-report measures, such as responses to questions about how much each person liked the other, would be prepared to work with the other, and so on. You could also use observational measures: unobtrusively video recording the interaction would permit you to measure various aspects of behaviour, both verbal (e.g. the extent to which the two persons discovered mutual interests or shared attitudes) and non-verbal (e.g. the amount of smiling or direct looking at the other person).

Consider the advantages of using both types of measure. First, the observational data provide one type of check on the validity of the self-report data, and vice versa. Just as questionnaire data can be distorted by the respondents' motivations, so too can observers' perceptions be distorted by the nature of the coding system they are using. If both kinds of data point to the same conclusion, this would enhance confidence in their validity. A second, potentially more important, advantage is that while self-reported attraction can be said to be an outcome of the interaction, observational measures provide an insight into the processes that might mediate that outcome. Researchers would typically be interested in finding out why people did or did not like each other; examining the behaviours that occurred during the interaction might shed some light on this.

In summary, using more than one type of measure is often helpful to the researcher. If observational and self-report measures of the same conceptual variable point to the same conclusion, this enhances confidence in that conclusion. Furthermore, self-report measures often assess the outcome of a process; by using observational measures as well, the researcher may gain insight into the process responsible for that outcome.

Problems with Experimentation

It is widely assume that the experimental method provides the 'royal road' to causal inference (cf. Aronson, Brewer and Carlsmith, 1985). In fact causal inference from the results of experiments is more problematic than many commentators allow.

One problem concerns what Gergen (1978) has called the cultural embeddedness of social events, by which he means that 'few stimulus events considered
independently have the capacity to elicit predictable social behavior' (p. 509). It follows that, even in the most tightly controlled laboratory experimental demonstration that the manipulation of independent variable X has a causal impact on dependent variable O, the circumstances in which X was manipulated may play a key role in producing the observed effects on O. The inference that 'X causes O' may therefore only be true under particular circumstances. As Gergen puts it: 'What passes for knowledge within the discipline may thus rest on an immense number of unstated assumptions and obscured conditions' (1978, p. 511).

A related problem, also articulated by Gergen, is that although the experimental method purportedly 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. Gergen argues; 'In dealing with human beings in a social setting it is virtually impossible to manipulate any variable in isolation of all the others. Even the most elemental variations in an independent variable have the capacity to elicit a host of intervening reactions' (1978, p. 515). 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 an alternative mediating process. Social psychology abounds with such debates between rival accounts for findings (see Greenwald, 1975b; Ostrom, 1977; Tetlock and Levi, 1982; Tetlock and Manstead, 1985), and some have come to the view that experimentation is not a suitable means of settling such between-theory disputes.

Yet another inferential problem confronting the experimental researcher in social psychology also stems from the fact that social behaviour is culturally embedded. In every culture there are norms that define the boundaries of appropriate social behaviour in particular settings, with the result that most individuals behave similarly in such settings. Such behaviour is best regarded as the product of that culture's conventions or rules, rather than intraindividual psychological processes. Difficulties arise when researchers examine social behaviour in experimental settings. These settings are not free from the operation of cultural norms; indeed, there are grounds for thinking that laboratory experiments may promote the occurrence of behaviours that are guided by norms (Semin and Manstead, 1979). Inferential difficulties arise when behaviour in such settings is interpreted exclusively in terms of hypothetical internal processes. For example, it might be argued that cultural norms prescribe that one does not question the instructions of someone running a scientific experiment, and that when one is asked to deliver an increasingly strong series of shocks to another person, apparently in the interests of scientific research, one should do so. That people are willing to do so, even when the shocks are strong enough to produce fatal results, is by no means uninteresting; but whether it reveals something about the psychological processes mediating obedience to authority is another matter. In short, it is important to avoid the temptation to formulate causal laws about psychological processes where there are grounds for thinking that the phenomena
being ‘explained’ have their origins in cultural convention (cf. Brandstaetter, 1982; Semin, 1986; Smledslund, 1985).

One final and related problem worth mentioning in this context 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, we have to take account of the fact that the relationship between researcher and the object of the research is radically different in the two types of science. The 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. This contrasts sharply with the objects of investigation in social sciences: being people, these ‘objects’ do of course attribute meaning and significance to their actions. Social psychology cannot therefore be neatly distinguished from what it studies; lay persons and social psychologists alike are concerned with understanding and interpreting their social environments. Lay persons are able to acquire social-psychological knowledge and use it to modify their actions in a way that atoms, elements and particles cannot. As Giddens (1982) puts it: ‘The fact that the “findings” of the social sciences can be taken up by those to whose behaviour they refer is not a phenomenon that can, or should be, marginalized, but it is integral to their very nature. . . . Human beings . . . are not merely inert objects of knowledge, but agents able to – and prone to – incorporate theory and research within their own action’ (pp. 14–16). One implication of this is that social-psychological theories should not be regarded as embodying ‘laws’ that will hold good across time: if learning about a social-psychological theory can lead individuals to modify the very behaviour that the theory purports to explain, it is clear that the theory has only limited temporal validity. Gergen (1973, 1978) has been the most persuasive advocate of this sobering view, although his arguments have been challenged by Schlenker (1974) and by Semin and Manstead (1983, chapter 5).

What are the implications of these problems for the status of experimentation in social-psychological research? It should be noted that even stern critics of the experimental approach do not advocate the abandonment of experimentation. For example, Gergen acknowledges that experiments will 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 does make 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, Brewer and Carlsmith (1985) sticking to the view that experimentation provides a firm basis on which to build knowledge, and critics such as Gergen questioning this assumption.

Summary and Conclusion

Methods are procedures followed by researchers in gathering information that helps them to answer research questions. Methodology is the term used to refer to all aspects of the implementation of these methods.

The type of method used by a given researcher will depend to a large extent on the kind of question he or she is addressing. We distinguished between three basic types of research – descriptive, correlational and experimental – and we noted that social-psychological research is typically either correlational or experimental.

In describing methods in more detail, we drew a distinction between research strategies and data collection techniques. Eight research strategies were described: field studies, field experiments, laboratory experiments, experimental simulations, sample surveys, judgement studies, formal theorizing and computer simulations. The principal ways in which these strategies differ are in terms of (1) their sensitivity to the context in which the data are collected; and (2) their obtrusiveness in implementation.

Experimentation (i.e. field experiments, laboratory experiments, experimental simulations) was singled out for more detailed discussion, in view of its dominance as a research strategy in social psychology during the last three decades. The main features of experimentation were identified as: the experimental scenario; the independent variable; the dependent variable; and the follow-up.

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 subjects are randomly allocated to one of two conditions, only one of which involves being exposed to the manipulation.

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.

We identified two principal methods of collecting data in social-psychological research: observational measurement and self-report measurement. Observational measures have the advantage of being less susceptible to social desirability effects, and can be made completely unobtrusive. On the other hand, they cannot directly tap covert cognitive phenomena such as causal
attributions (see chapter 6). Here the researcher must rely on self-report measures, although the advantages of using both types of measure in conjunction should not be overlooked.

Finally, we noted that some social psychologists have questioned the utility of conventional methods, and of laboratory experiments in particular. The cultural embeddedness of social behaviour, the fact that social behaviour is determined by multiple factors, the difficulty of discriminating between normative and psychological causation, and the ability of humans to modify their behaviour in the light of social-psychological theories, were singled out as grounds for questioning the notion that experimentation will result in cumulative knowledge of the laws governing social behaviour.

<table>
<thead>
<tr>
<th>Glossary Terms</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer simulation</td>
<td>Hypothesis</td>
</tr>
<tr>
<td>Confounding</td>
<td>Independent variable</td>
</tr>
<tr>
<td>Construct validity</td>
<td>Interaction process analysis (IPA)</td>
</tr>
<tr>
<td>Control group</td>
<td>Internal validity</td>
</tr>
<tr>
<td>Cover story</td>
<td>Judgement task</td>
</tr>
<tr>
<td>Debriefing</td>
<td>Laboratory experiment</td>
</tr>
<tr>
<td>Demand characteristics</td>
<td>Obtrusive measures</td>
</tr>
<tr>
<td>Dependent variable</td>
<td>One-shot case study</td>
</tr>
<tr>
<td>Experiment</td>
<td>Participant case study</td>
</tr>
<tr>
<td>Experimental scenario</td>
<td>Post-experimental enquiry</td>
</tr>
<tr>
<td>Experimental simulation</td>
<td>Post-test only control group design</td>
</tr>
<tr>
<td>Experimenter effects</td>
<td>Random allocation</td>
</tr>
<tr>
<td>External validity</td>
<td>Reliability</td>
</tr>
<tr>
<td>Field experiment</td>
<td>Sample survey</td>
</tr>
<tr>
<td>Field studies</td>
<td>Unobtrusive measures</td>
</tr>
<tr>
<td>Formal theory</td>
<td>Validity</td>
</tr>
<tr>
<td>Hawthorne effect</td>
<td></td>
</tr>
</tbody>
</table>

Further Reading


