Breakwell-3389-Ch-01.qxd 2/7/2006 8:48 PM Page 1

PART 1

THE BASES OF RESEARCH

1
Theory, Method and Research Design

Glynis M. Breakwell

AIMS

The purpose of this chapter is to introduce the importance of theory building and the difficulties associated with theory testing. In order to do this, the scientific method and its limitations are described. The relevant, and often debated, distinctions between positivist and constructionist approaches to theory and research are summarized. Different types of data elicitation are outlined and their relationship to data analysis considered. The varieties of research designs that can be used for monitoring change are examined and the role of manipulation in research designs is explained. The chapter concludes by discussing the significance of integrating findings from different types of methods.

Key terms

- anomalies
- cross-sectional design
- deduction
- Duhem–Quine thesis
- falsify
- functional explanation
- hypotheses
- induction
- longitudinal design
- manipulation checks
- normal science
- Occam's razor
- paradigm shift
- positivist
- process explanation
- relational rules
- revolution
- sequential design
- stipulative statement
- theory
- qualitative treatment
- quantitative treatment

1.1 THEORY BUILDING AND THEORY TESTING

1.1.1 The importance of theories

We do psychological research to make sense of the thinking, feeling and actions of people. We do it to try to understand what is happening. To be more precise, we do research to find out what has happened, how it happened, and, if possible, why it happened. We use 'happened' rather than 'happening' because by the time we have recorded something it is inevitably in the past. But there is more: once we have some idea of the kinds of things that may happen and their relationship – in other words, when we have a **theory** – we can use that theory to predict what will happen in the future. If we also understand why things happen the way they do, we may even be able to improve the future by intervening in the world. Knowledge is said to be power, and knowledge is stored elegantly and systematically in the form of theory. Consequently, good theory is both powerful and of practical relevance.

theory

Research is not inevitably tied to formal theory building or theory testing. Some researchers self-consciously eschew the construction of theory for philosophical reasons. They use their research to describe in detail specific happenings without any attempt to use these as instances to illustrate or test some general underlying explanatory framework. Other researchers ignore theory because they do their research for purely practical reasons. They think they only need to know what has happened in order to decide what they (or their clients) should do next.

Yet it could be argued that even those researchers who have no time for formal theories are in fact working with implicit theories. Any set of ideas about the relationships between variables (or what are sometimes referred to as constructs or concepts) has the attributes of a theory. Implicitly, we build theories all the time. In fact, George Kelly (1955), when he developed the theory of personal constructs to explain personality and cognitive processes, based his argument on the metaphor of 'man the scientist'. Kelly suggested that we all behave as scientists in so far as we are inveterate constructors of theories. They help us to navigate in the world by letting us explain to ourselves what we think is happening and why it is happening. Informal theories of this type are particularly valuable because we invariably use them as the basis for predicting what will happen next.

So, for instance, you might observe events which lead you to conclude that men in their late teens or early twenties are more likely than other people to drive their cars aggressively and erratically, with their windows open, with heavy bass music blaring, on sunny days. From this you might generate an informal

theory that attributes their driving behaviour to their age or possibly some interaction of their age with the music and the sunshine. As a naïve scientist, you are not obliged to test your theory. You can go ahead to predict from it that young men playing audible music on sunny days are more likely to be a danger to you and adjust your own behaviour accordingly. Theory building of this sort has survival value – though when it goes wrong it also has the capacity seriously to endanger you. Such theories can mislead you, directing you to cues in the situation which are irrelevant. Your sunny day, music-loving bad driver theory is only really valuable if it turns out that in the rain young men are at least as good as any other driver. However, Kelly pointed out that the erroneous theory is usually just a staging post to a better version. As naïve scientists, we are mostly willing to refine our theories on the basis of new information which proves the earlier versions wrong. This is a characteristic of the scientific approach properly pursued.

The implicit theories which inform some research have the same sort of power to both aid and hinder survival. Even while they remain unmentioned, these implicit theories direct the attention of the researcher to focus on some things rather than others, to use certain research approaches rather than others, and to try this rather than that form of analysis. In most cases, it would be better if the researcher did articulate these implicit theories. By doing so, it is possible to analyse their logical weaknesses (e.g. inconsistencies) and their substantive weaknesses (e.g. omission of important variables). Some researchers resist making their implicit theories explicit because they do not see their task as theory building. Yet this is actually a poor excuse. It does not really preclude the need to specify what theoretical assumptions underlie your work. Since these assumptions will inevitably affect what you do, they should be described so that other researchers can judge how far your research activities and findings are influenced by them (e.g. Dobson & Rose, 1985). Researchers who do try to lay bare these underlying assumptions often find that the discipline needed to articulate them has the effect of leading them to a new understanding of their research problem. Essentially, whether you see yourself as a theory builder or not, it is always useful to examine the implicit theories which affect your research. Indeed, many approaches in psychology (e.g. those in Chapters 13, 15, 16, 17 and 18) now require the researcher to be explicit about their own position in relation to their research question and data, tasking them to be self-reflexive and to lay bare their own preconceptions and expectations. Thus even where researchers might be rejecting the application of a priori theoretical frameworks, they are seeking to expose their preconceptions. This is undoubtedly important in order to allow others to interpret the nature of data and explanations ultimately presented.

Box 1.1 The Basic Scientific Method

The basic steps in the scientific method might be summarized as:

1. Formulate the research problem clearly, simply and completely. It might be: what is the relationship between variable *X* and variable *Y*?
2 Develop an idea of what might be the form of the relationship between *X* and *Y* and outline it in general terms. It might be: *X* results in *Y*.
3. Specify an exact hypothesis about the relationship between *X* and *Y*. It might be: the occurrence of *X* always precedes the occurrence of *Y* and *X* never occurs without *Y* following it.
4 Set up a controlled test of the hypothesis, specifically attempting to engender conditions where the hypothesis can be shown to be wrong. It might entail describing all naturally occurring incidences of *X* and *Y* to determine whether they always co-occur. It might entail inducing the occurrence of *X* under a variety of constrained conditions and establishing whether *Y* will always occur.
5. If the test shows the hypothesis to be wrong, it should be abandoned or, more likely, amended.
6. If the test fails to disprove the hypothesis it may be accepted conditionally before designing further tests and refining the range of its applicability. The object is always to define the limits of the predictive power of the theoretical model.

Some of the limitations of this basic scientific approach to theory construction are given in Sections 1.1.4–1.1.6. Of particular importance is the critique of this approach from those researchers who would not accept that hypothesis testing is an appropriate approach.

Of course, the assertion that everyday thought parallels scientific thought is based on a set of assumptions about what constitutes the scientific approach. Box 1.1 outlines some basic elements of the traditional scientific method. There has been endless debate in psychology about whether it can consider itself a science. There is no final answer to this question: not all psychological research deploys the scientific method; not all psychologists would aspire to do so. The fundamental problem that faces psychology in the pursuit of scientific status is that the majority of the constructs that it finds interesting (e.g. intelligence, motivation, identity) lie at a level of analysis that means that they can only defined after many levels of extrapolation from any objectively measurable events.

1.1.2 The basic construction of a theory

The process of formal theory building was traditionally supposed by philosophers of science to proceed in an orderly manner from description, to taxonomy, and

thence to testable causal **hypotheses**. This would mean that the first task of the theorist is to describe the phenomena of interest thoroughly and systematically. The next task is to categorize phenomena, showing how specific instances are characterized by common attributes which make them capable of being treated in some sense as equivalent to each other. Such categorization is one way of ordering the plethora of data which are generated whenever descriptions are not prestructured. The categorization scheme can be labelled as a theoretical construct. Learning theory has categorized phenomena to generate two very salient constructs: stimuli and responses. To the behaviourist, all phenomena at any one moment can be categorized as either a stimulus or a response. By this act of definition, suddenly the multitudinous world is dichotomized, order is imposed and it is our task to explain the relationship between stimulus and response.

hypotheses

Once the taxonomy is complete, the theorist's next task entails stating how one category of phenomena is related to another. The description of a single set of relationships between phenomena does not become a theory unless general principles about the relationships of similar phenomena are formulated. The observation 'The woman kicked the dog after he bit her and he was never found to bite her again' is a description of a pattern of events: it is not a theory. But if one were to say, 'The woman punished the dog for biting and he never did it again', there remains only one further step towards generalization before a theory comes into being: 'Punishment of a behaviour leads to the diminution of that behaviour'. The result is a recognizable basic tenet of learning theory (to be much qualified later by statements about the frequency of punishment, the temporal relationship of the punishment to the behaviour, and the availability of alternative rewards for the behaviour, etc.). Basic theories are sets of what might be called **relational rules**. The relational rule specifies how variation in one theoretical construct is related to variation in one or more others.

relational rules

1.1.3 The nature of explanation: process and functional

An explanation can be of two sorts: the mechanistic or process variety or the functional variety. The **process explanation** accounts for a phenomenon in terms of phenomena which are its precursors. It is usually in the form: if A and B occur, then C will follow. In contrast, the **functional explanation** accounts for a phenomenon in terms of the consequences it has. It is usually in the form: A occurs in order that B will follow. The functional explanation assumes that the phenomenon to be explained is purposive, intentional or teleological (i.e. it occurs to achieve some goal).

Process explanation

functional explanation

Another way of talking about the distinction between the process and functional types of explanation is to say that the former is concerned with causes and the latter with reasons: with how as opposed to why. Traditional theories, for example in physics, tend to deal only with the first of these. However, psychological theories

use both types of explanation. Some theorists use both explanatory forms to account for a single psychological phenomenon. For instance, in studying altruism (helping or pro-social behaviour) researchers have found that people are less likely to offer someone help if they perceive that person to be in need because he or she has made too little effort, not used his or her own ability and not chosen to get out of the difficulty when it was possible to do so. One explanation of helping suggests that people see the need for assistance, then assess whether the individual is responsible for his or her own predicament; if he or she is responsible this leads to anger and no helping, but if he or she is deemed not to be responsible this leads to sympathy and helping. This is clearly a process explanation. Another explanation of helping suggests that people are unwilling to help an unfortunate whom they see to be the origin of his or her own fate because they wish to punish the miscreant for failures of effort or judgement. In this explanation the punishment (i.e. failure to help) serves the function in some way of exacting restitution and may warn others that such slack behaviour is unacceptable and not rewarded with help. It should be noted that these two explanations of the same phenomenon are not mutually exclusive. The functional explanation may serve to account for the anger, so central to the process explanation, which is aroused when the needy are shown not to have tried to shift for themselves.

1.1.4 Building more complex theories: induction and deduction

This mixing of mechanistic and functional explanations is common in psychological theories. It may emanate, in part, from the way psychological processes and thus psychological theories often traverse many levels of analysis. It is argued here (see also Breakwell, 1994; Rose & Dobson, 1985) that psychologists should be building theories which encompass processes at the intrapsychic (i.e. physiological, cognitive, affective and oretic) level, the interpersonal level and the societal level of analysis. These theories should be integrative, linking the hypotheses and models that explain specific psychological processes. But we are currently a long way away from the grand theory in psychology. We have low- or middle-range theories designed to explain relatively narrow bands of phenomena. Thus, for instance, we have theories of aggression distinct and separate from theories of altruism when common sense might think them to be in some way connected. While these low-level theories may offer a detailed process explanation of their target phenomena, they tend to rely upon what Israel (1972) called **stipulative statements** that concern assumptions about the nature of the individual, the nature of society and the nature of the relationship between the individual and society. These stipulative statements are often functionalist (e.g. a variety of social Darwinism as illustrated in the theory of altruism described above). This results in a strange mixture of explanatory types being moulded together in many psychological theories – made more strange by the fact that

stipulative statements

some significant element of the explanation remains unsaid. Those elements which lie at a different level of analysis from that of the main theory will reside at the margin, unexamined and untested.

Just as process and functional explanations are not so simple to keep apart, the distinction between theories built from **induction** and theories developed through **deduction** is not easy to maintain in practice. Induction entails inferring a general law from particular instances (such as the theory about young men drivers given above). Deduction entails drawing from the general an inference to the particular. In practice, theory building is a messy, iterative process. Relational rules that seem to be valid are usually crafted by successive approximation. This process of approximation will involve both deductive and inductive reasoning (Oldroyd, 1986). You may well, for instance, in developing a theory of how identity processes which concern self-esteem affect memory capacity, set off by cataloguing the range of examples of instances where memory capacity has been shown to be greater for self-relevant information and where it has been proven to be more accurate for information which is positive about the self. From this you might induce a generalization: the memory for self-evaluative information will be greater and more accurate if that information is positive than if it is negative. From this generalization, you might go on to deduce that memory for exam results will be better if they are the individual's own results and especially if they were good results.

induction

deduction

In summary, the process of induction allows us to produce theoretical generalizations which are based on evidence about a range of specific instances; one reason for doing research is to collect this evidence. The process of deduction allows us to derive specific predictions from those generalizations, and another reason for doing research is to test these predictions.

1.1.5 Theory testing

For a long time, it was thought that theory testing involves showing that a theory gives rise to accurate predictions about what will happen under a particular set of circumstances. However, this method is not really convincing, since it can never prove that the theory will always be right under every possible set of circumstances, no matter how many times it is tested. Instead, it has been suggested that what we should do in testing a theory is to try to prove it wrong (Popper, 1959). By showing where a theory is wrong, we show which bits need to be removed and in most cases we also show what needs to be substituted in their place. Research designed to test a theory will be organized so as to show whether a prediction deduced from that theory is wrong. This would **falsify** the theory. If we fail to disprove the prediction, the theory survives to face another test. Research can never prove a theory, it can merely accumulate examples of where the theory has not been disproved. The reason why a theory cannot be proven in absolute terms is that it must entail generalization and the empirical research can

falsify

only ever sample specific instances of that generality. A good theory is one which survives intact through many sincere and severe attempts at falsification.

One problem with this approach is that theories may survive because they are not strictly falsifiable. Some theories are unfalsifiable because they rest upon a tautology. For instance, some critics of learning theory would argue that one of its fundamental assertions cannot be falsified because the concept of a reinforcer is defined in a tautological way. Thus a reinforcer is defined as anything which acts to increase the frequency of a response. The theory then goes on to state that responses which are reinforced increase in frequency. The circularity in the argument is clear when the theory is reduced to its fundamentals in this way. Such a tautology means that the theory cannot be tested because a key concept cannot be operationally defined independently of other concepts within the theory.

Another problem is the Freudian theory of ego-defence mechanisms, which cannot be falsified for a different reason. In this case, the theory attempts to explain how the conscious mind protects itself from material which must remain in the unconscious or preconscious mind. Freud explained that such material is handled by a series of ego-defence mechanisms (sublimation, displacement, regression, fixation, etc.). What makes this aspect of psychoanalytic theory untestable is the fact Freud offers such an array of defence mechanisms that it is impossible to formulate a test of the operation of one which would not be potentially abrogated by the operation of another. For instance, one might set out to test the notion of sublimation which would say that an unacceptable unconscious impulse driven by the id should be translated into one which is socially acceptable before it could gain access to the conscious mind. The first problem the empiricist would have is in knowing that the impulse existed at all. The second would be that the impulse need not be dealt with by sublimation: it could be treated through reaction formation. This would mean that even if you established when the impulse was occurring and monitored no evidence of sublimation, you would not have falsified the theory because the impulse had been dealt with by another equally valid defence mechanism. Freud basically produced what could be called an overdetermined model: a theory which allows multiple determiners of outcomes in such a way that no single determiner can be empirically proven to be irrelevant.

1.1.6 Advanced theory structures

Other problems with falsifiability appear when we consider theories in advanced areas of research. Although induction can give us our initial ideas when we first begin to investigate a fresh research field, delving further into an already established area of research requires new skills entirely. Deductive skills will serve for testing simple theories, but in advanced areas of research the theories have been built into complex structures, with many levels and enormous ranges of applicability.

Some philosophers have divided the components of such theories into two types: the fundamental 'hard core' of assumptions and presuppositions that are

basic to the whole enterprise, and a collection of auxiliary hypotheses that derive from the hard core and make predictions about what will happen in particular situations (Lakatos, 1970). The whole complex is known as a research programme. Everyday research concerns tests of the auxiliary hypotheses, and the more the programme generates new hypotheses, and the more these find empirical support, the more 'progressive' the research programme is said to be.

A single experiment is, however, not sufficient to falsify a complex theory: instead, only a series of negative results (failed predictions), together with a paucity of new ideas stemming from within the programme, can suggest that the programme is 'degenerative'. Only after a period of such degeneration may the situation be considered so bad that it is necessary to revise or even reject the whole theory.

Of course, in practice it is difficult to judge when this latter course of action is necessary. This is because one of the characteristics of a complex theory structure is that it can be so easily modified by adding or subtracting new components. In fact a fundamental problem with these structures is that just about any negative or unexpected empirical result can be accommodated by adding *ad hoc* hypotheses to the theory. This is known as the **Duhem–Quine thesis**.

Duhem–Quine thesis

For example, the presupposition that children learn by copying others might lead to the hypothesis that children who view violent cartoons on television will subsequently show similar antisocial behaviour. Suppose then that observations of a group of 10-year-olds who regularly watch 'Tom and Jerry' cartoons reveal no difference in some measure of violent social behaviour, when compared with 10-year-olds who do not watch 'Tom and Jerry'. The theorist can then say: ah, well, the antisocial behaviour will only be revealed later, when the child enters adolescence and encounters more testing situations such as gang fights. Or the cartoons do have the predicted effect but only on children of a particular age, for example below 10 years old, because by age 10 the children have learned the difference between cartoons and reality. Or only cartoons involving violence between humans will lead to antisocial behaviour. And so on. In each of these cases, the theorist has added an additional hypothesis after the facts of the experiment have become known. These hypotheses act to save the core tenets of the theory from falsification (disproof), by making the theory a little more complicated. Although in principle we should not allow such *post hoc* theorizing, it is inevitable, given that life is complicated and a correct theory will therefore also need to be complex, if it is to specify in exact detail what will happen when and under what circumstances. However, the indefinitely large number of these possible *post hoc* hypotheses does make it very difficult to disprove a theory conclusively.

1.1.7 Testing a theory structure

Although psychology as a whole does not yet have a single grand unifying theory, many of the theories you encounter will nevertheless be so complex as to render simple interpretation and methods inadequate. This fact has several consequences for research methodology in those situations.

First, the choice of which observations it is worthwhile to make is informed by the background knowledge that has already been built up in that field. This requires you to know not only what observations have already been accrued, but also which theories and hypotheses are still under development and are worthy of further investigation.

Second, the meaning of any new observations you make depends on the background theory: you have to interpret your empirical findings in the light of the theory. For example, criminal behaviour can appear very different, depending whether your underlying theory of behaviour is genetic or behaviourist or whatever: are violent criminals just basically evil, or victims of their hormones, or morally corrupted by a corrupt society around them, or copying the example set by their violent parents, or suffering from brain damage?

You also have to know what the implications of the observation are for the whole structure of the theory: for example, is an unexpected observation inconsistent only with the particular hypothesis you are testing, or does it count against the underlying assumptions of the whole theory? In other words, does the failure of prediction arise because the fundamental tenets of the whole approach are wrong, or merely one of the auxiliary hypotheses – and if so which one? How can you tell? Falsifiability becomes problematic when there are so many interacting hypotheses, data bases and collateral assumptions that any one of them might be in error, so throwing your prediction out.

Take, for example, Bandura's (1997) theory of self-efficacy, which states that people vary in the extent to which they believe that they can achieve whatever they set out to do (i.e. they vary in self-efficacy). People who have high self-efficacy expectancies are healthier, more effective and generally more successful than those with low self-efficacy expectancies. From this theory we may derive a hypothesis that a person who is high in self-efficacy and who becomes ill will be more likely to take medication to cure the condition. However, this behaviour will only take place if the person has a belief that the medication will be efficacious. A method of testing this hypothesis is to monitor whether people with high self-efficacy who become ill are actually more likely to take medicine when advised to do so. Suppose, however, we find they do not. Is this because Bandura's theory is wrong, because our hypothesis is wrong, because the variables of self-efficacy and belief in the medication have been indexed inappropriately, or because our method of measuring behaviour was inadequate or inappropriate?

All predictions also have a *ceteris paribus* clause attached: they assume that no extraneous variables or factors will interfere with the observations or invalidate them in any way. It is a frequent occurrence for data not to be as expected, and perhaps the commonest way of explaining the anomaly is to postulate an extra variable that is affecting the outcome. In the above example, for instance, people might not take medicine voluntarily because they have an additional belief, such as that medication in general does not work, or is immoral on religious grounds,

or would cause too many side-effects. Controlled experiments that are designed to reveal any such extraneous variable are accordingly often performed on an *ad hoc* basis after the main body of observations has been made.

Third, the empirical methods used are themselves theory relative. It is not possible to divide science into theory and observation, as was claimed by **positivist** philosophers earlier in the last century. Although empirical science worships the idol of the 'neutral' observer who is an unbiased recorder of nature, in practice this standard is an unattainable ideal. Thus we all have some idea of what we might find when we make an observation, and in many cases we know what we want to find. The techniques of 'blind' experimentation have been developed to help cope with exactly this problem.

positivist

Additionally, the very measuring instruments we use have been developed under particular theoretical backgrounds; their construction depends upon a whole network of theoretical assumptions about the nature of the materials used and the way these interact with the subjects of the experiment. For example, psychoanalytic theory developed the Rorschach inkblot test as an instrument to facilitate behaviour sampling. Responses to this test have been used to make inferences about a person's 'personality' type. However, the results do not make sense under modern personality theory, which instead uses complex statistical analyses of responses to a much wider variety of simpler but more strictly defined tests (e.g. questionnaires). The kinds of personality categories recognized under psychoanalytic theory are not commensurate with those of alternative theories. Another example would be trying to assess memory ability by presenting nonsense syllable pairs on a memory drum to see how many associations can be remembered. This makes sense under behaviourism; but under Gestalt theory, or modern cognitive theory, the number of associations is just not interesting: instead it is the emergent, holistic and semantic organizational properties of memory that are relevant, and the kinds of tests that are considered appropriate are very different (e.g. structuring in free recall). Moreover, under some theories there is no such thing as a 'nonsense syllable': all stimuli are taken to have meaning.

The idea that the interpretation of the data generated always depends upon a network of background theory does not just apply to the 'apparatus' by which we interact with the subject; it is no less true of the quantitative numerical techniques we use to analyse the data (Lamiell, 1995). For example, the basic methods of statistics were developed under positivist principles, which are now considered obsolete. They assume that theory-neutral observations can be made by unprejudiced objective impartial observers. The process of using statistics thus creates a delusion of certainty that makes it all too easy for us to fall into the trap of letting the empirical results and the significance levels of the associated analyses tell us what to believe. You should not allow yourself to be dominated by the probability level in deciding what conclusions to draw. Quantitative results have to be interpreted in the light of a whole lot of background knowledge, theory and opinion.

Finally, an established theory will have accreted to itself the results of numerous empirical tests, some of which support the theory and some of which do not. If the supporting evidence comes from many different sources and is of many different types, the convergence on to the same conclusion from all these sources is regarded as making the theory stronger or more valid than if the evidence comes from repeated observations all of the same type because the latter might be caused by some artefact or error in the method. Predicting new observations is also generally regarded as carrying more weight than fitting retrospectively an old body of observations. False predictions of the theory are categorized as **anomalies**, and most advanced theories bear a number of these – although not all the supporters of the theory will necessarily admit their existence. This is because evidence is not always accepted at face value, if only because experiments are complicated and not perfectly reliable. It is only with years of hindsight that empirical outcomes can be classified as valid or invalid, in the light of which theory has turned out to be the correct one (Lakatos, 1970). In the interim, it is conventional to live with the anomalies, provided they are not too many in number or too crucial and convincing.

anomalies

Periods in which scientists beaver away collecting data under a generally accepted theory have been described as **normal science** (Kuhn, 1962). Complete rejection of such generally accepted theories does not usually occur unless there exists an alternative or competing theory (as well as many anomalies in the old theory). Until such arrives, researchers have no real alternative but to carry on using the old theory, despite all its faults. When another theory exists, it predicts a pattern of results that differs from the pattern predicted by the old theory. If the new pattern fits the actual data better, the new theory is likely to be adopted (allowing also for the other criteria for theory acceptance outlined below). Once such a **revolution** or **paradigm shift** occurs, the meaning of all the empirical results is reinterpreted: what previously appeared to be peculiar if not downright bizarre data may now be seen to be understandable, given the new theory. So previous anomalies are now consonant with expectation, in that they can be deductively related to the covering laws (relational rules) in the new theory. Any observations that remain unexplained become anomalies under the new theory; adoption of the new theory is intended to reduce the number of anomalies as much as possible. Positivist philosophers suggested one could simply count up the number of successful empirical predictions of a theory, subtract the number of anomalies, and pick the theory with the highest score. However, this assumed all observations carry equal weight and each is an isolated nugget of fact. Instead, observations fit together to form a pattern, and the links between them should form a logically coherent structure.

normal science

revolution
paradigm shift

The psychological process of reinterpreting and reorganizing the whole set of empirical data has been likened to the Gestalt-like change in perception and understanding that can occur in individual knowledge following an act of 'insight' (Kuhn, 1962). A novel synthesis of the data into a new pattern is usually

presented in the form of a review article and brings great credit to its originators. While you will almost always undertake a 'literature review' as part of any research project, in which you summarize the extant findings in the field, if you perceive a novel pattern by which more of the data can be accounted for by some new theory than by any existing theory, your contribution will be highly valued.

We have discussed theory testing as involving primarily empirical testing and observation, but given the existence of complex theory structures and the Duhem–Quine thesis, reliance on simple observational evidence is clearly not enough. The alternative is to use rationalist principles alongside empiricist ones: theories can be evaluated according to several non-empirical criteria, such as their parsimony, ease of communication, flexibility, fruitfulness, insightfulness, internal consistency, simplicity, elegance, breadth, and so on. Such principles are used implicitly by all scientists and are actually a central reason for the success of science. No one criterion alone is sufficient: a judicious and balanced combination of arguments should guide one's choice.

Consider simplicity, for example. A corollary of the Duhem–Quine thesis is that any given set of data can in principle be explained by an infinite number of theories. The thesis states that theories can be elaborated to any degree of complexity we wish. It is therefore possible to take almost any theory and by adding sufficient auxiliary hypotheses modify that theory so it can explain a particular set of data. In such cases, our choice of which theory to adopt is 'underdetermined' by the data, since the data do not point unequivocally to one theory and one theory only. The normal response to this problem is to choose the simplest theory, applying what is called **Occam's razor** (sometimes spelt **Ockham's** razor), which states that we should not multiply hypotheses needlessly. A problem with simplicity is that it assumes that the complexity of a theory can be measured in some objective way so that different theories can be compared. However, advanced theory structures differ so much qualitatively as well as in their quantitative 'complexity' that comparison by any single common yardstick may in practice be impossible. Thus theories are often incommensurable because their underlying purposes and assumptions are different; they have different criteria for their own success, since their aims and intended context of application are very different. Theory comparison and selection may then appear a matter of subjective judgement.

Occam's razor
Ockham's

1.1.8 The meaning of theories

The notion that a theory can be successful for a period of time, and can then be replaced by a totally different and incommensurable theory that is even more successful, leads to the question of why the first theory should have been successful at all, given that it was wrong. Research into scientific practice itself has shown that science is not so different from other spheres of life: it is a social activity, and the choice of which theory to believe, which data to accept as

correct, which professor or department to rely on, is as much a matter of attitude and opinion formation as is any other psychological belief. Personal and social factors cannot be excluded from science. At its most extreme, this school of opinion (known as the strong programme in the sociology of knowledge) denies that any theory describes an objective reality: it is just a matter of socially reached consensus among the scientific community as to what to believe (relativism).

Other researchers treat theories as useful ways of predicting what will happen under given circumstances, but without making any claim one way or the other as to whether the theories describe an actual 'reality' (this attitude is variously called pragmatism, operationalism or instrumentalism). This issue is still the subject of intense debate in the philosophy of science. Moderating voices accept the (empirically observed) fact that social factors do operate in science, but that objective empirical data about the world play their part too: the social factors are not the sole determinants of theory acceptance or rejection. Thus although interpersonal disputes play a crucial part at the cutting edge of research, where the truth is still uncertain, in the long run scientists are kept on the right path by some kind of objective reality operating via the empirical observations (e.g. Oldroyd, 1986; Hull, 1988; Kitcher, 1993; Klee, 1997).

Box 1.2 summarizes in bald terms the distinction between the approach of positivist-realists and the constructionist-relativist traditions.

The Duhem–Quine thesis and the strong programme in the sociology of knowledge have led some people to conclude that relativism is the norm. Thus

Box 1.2 Comparison of positivist and constructionist* approaches to research

POSITIVIST	CONSTRUCTIONIST
Facts can have an objective reality	Facts are subjective constructs
Data validity and reliability are sought	Reliability and validity are irrelevant concepts since the data are not judged in terms of any external notion of truth
Hypotheses should be explicit and pre-date data collection	Understanding is emergent and explanation can emerge after data are collected
Prediction is an objective Falsification of hypotheses is an objective	Description is an objective Usefulness of interpretation is an objective

*There are many labels used to describe these broad traditions and there are many subsets within each, all having somewhat different philosophies. The comparison here is in bald terms and at the most generic level.

although everybody starts out believing that there is an absolute truth about how the world is, and there is a single correct answer to every problem, which it is science's job to find, we soon realize that life is more complicated, that people in positions of authority can hold differing opinions (often diametrically opposed) and that no one can be 100% sure about anything. Most research students reach a point when they realize that observations are not pure nuggets of truth, and that they all depend upon a network of assumptions (about the nature of the measuring instrument, the theoretical and observational presuppositions from which the hypotheses were derived, and so on, as explained above). At this point it is important not to despair; life goes on and science does work. You have to realize that all beliefs have pros and cons, that whatever theory you adopt will have some evidence and arguments in favour of it and some against. You have to decide which theory to believe in, otherwise you cannot act. You will select what is in your opinion the best theory, given the currently available theories and evidence. However, your choice of theory should not be adhered to as a dogma that cannot be contradicted. You must adopt a flexible attitude: you must realize that your choice is provisional, and you must always be ready to change your belief in the light of new evidence and arguments.

Nowadays, there is increasing appreciation of the complexity of psychological systems and processes. Psychologists and biologists look for particular causal mechanisms, rather than for universal covering laws (e.g. Bechtel & Richardson, 1993). This avoids the problems of the grand theory structures outlined above. The idea is still to explain the phenomena that are observed, but to give understanding of how and why they arise, in terms of what causes them, not which universal law of nature they are deductively in accordance with. The idea of simple universal laws arose within seventeenth-century physics; more recently, however, we have come to realize that biological systems are too complex to analyse using the same methods, and entirely different principles are called for. Functional and mechanistic explanations (Section 1.1.3) must be given together. People and animals evolved and survived within a chaotically changing range of environments that have shaped and altered us over the aeons in ways that no simple law will describe. Our aims as psychologists therefore have to be to explain the particulars of mental life and behaviour we encounter in terms of the individual people we are observing, their constitutions and their immediate and past circumstances. To do so we have to use multiple methods, both rationalist and empiricist in nature. It is to this end that many of the methods described in the following chapters of this book are directed.

1.2 MATCHING METHODOLOGIES TO THEORY

Different types of theory have to be tested using different types of research method. The nature of the theory limits the range of research methods which can be meaningfully used to test it. For example, a theory explaining variation in

visual acuity is likely to need to measure acuity using some psychophysical technique (see Chapter 9). However, the extent of these limitations should not be overestimated. Most psychological theories can be tested using more than one method. In fact, it is advisable to try to test a theory using a variety of methods in order to prove that it is no artefact of the method which results in the theory being supported.

A piece of research can differ along a series of five independent dimensions:

1 type of data elicited;
2 technique of data elicitation;
3 type of design for monitoring change;
4 amount of manipulation used;
5 treatment of data as qualitative or quantitative.

1.2.1 Type of data elicited

In psychological research the data can vary in origin: they can be intrapersonal (e.g. genotypic information, cognitions, emotions), interindividual (e.g. friendship networks, communication patterns), or societal (e.g. institutional hierarchies, ideological systems).

1.2.2 Technique of data elicitation

Data can be elicited directly or indirectly from a target. Direct elicitation methods would include any stimulus to self-report (e.g. interviewing, self-completion questionnaires) or self-revelation through behaviour (e.g. role play, performance on tasks). Indirect elicitation methods would include techniques that rely upon the researcher observing behaviour (e.g. participant observation) or using informants about the target's behaviour, thought or feelings (e.g. archival records, witnesses).

Data elicitation can vary in terms of the amount of control exerted by the researcher upon a target. This control can be manifest in restrictions imposed upon the freedom of the target to give information (e.g. forced-choice options rather than open-ended responses to questions). It can be evident in the extent to which the target is manipulated (e.g. in experiments through the creation of artificial contexts or in surveys through the use of cover stories designed to mislead the target about the purpose of the study).

1.2.3 Type of design for monitoring change

A central task for psychological theories is to explain change. Researchers whose objective is to identify and explain change have a choice of three main classes of design for data collection: longitudinal, cross-sectional or sequential.

A **longitudinal design** involves data being collected from the same sample of individuals on at least two occasions. The interval between data collections and the number of data collections vary greatly: the research can be contained in a few days or spread over several decades. A longitudinal design allows researchers to establish changes in individuals over time as the sample ages or experiences some identifiable alteration in experience. In experimental parlance, a longitudinal design might be called a repeated measures or within-subject design (see Chapter 4).

longitudinal design

A **cross-sectional design** involves eliciting information at a single time from people in a number of different conditions that are expected to be significant to the change. Often this means studying people in different age cohorts because, particularly in theories of developmental psychology, age is deemed to be a major determinant of change. The term 'age cohort' refers to the total population of individuals born at approximately the same time, which is usually taken to mean in the same calendar year. The cross-sectional design permits age-related changes to be gauged.

cross-sectional design

A **sequential design** will choose samples from a particular condition (e.g. a specific age cohort) but will study them at different times. The periodicity in sequential data gathering varies across studies. A simple sequential design might involve sampling the 21-year-old cohort of 1989, the 21-year-old cohort of 1979, and the 21-year-old cohort of 1969. This type of design would be targeted at revealing whether changes in a particular age group are affected by factors which are associated with their specific sociohistorical era.

sequential design

When studying patterns of change that are age-related there are always three factors which could possibly explain observed relationships: development tied to the ageing of the individual; characteristics associated with the particular age cohorts studied; and impact of the specific time of measurement. Time of measurement is the term suggested by Schaie (1965) to refer to the set of pressures upon the individual generated by the socio-environmental context at the point data are collected. The difficulty facing researchers interested in explained age-related changes lies in establishing which of these three factors is the source of the change. The strategy adopted by most researchers is to keep one of the factors constant. For instance, the longitudinal design keeps the cohort constant. The cross-sectional design keeps the time of measurement constant. The sequential design keeps the chronological age constant. Of course, this means that explanation of any observed age-related trend remains problematic since these designs always leave two of the three explanatory factors free to vary simultaneously. Irrespective of which of these three designs is adopted, two explanatory factors will be confounded. This represents the major methodological drawback in using such relatively simple designs.

There is a secondary problem. By holding one factor constant, the design obviously rules out the possibility of exploring the effects of that factor in

interaction with the others. Yet, in virtually all complex systems of change, one would expect interaction effects between developmental, cohort and time of measurement factors. The solution to this fundamental methodological problem has been to integrate the three design types in what is known as a longitudinal cohort sequential design. This combines the longitudinal follow-up of a series of cohorts first sampled simultaneously as in a cross-sectional study with the sequential addition of new cohorts of the same ages to the study at each subsequent data collection point.

Even if a researcher believes the psychological construct under investigation is not influenced by the chronological maturation of the individual and not affected by the sociohistorical context of the data collection, the burden of proof rests upon that researcher to show that they are not important. It used to be thought that only a developmental psychologist really needed to consider whether to use a longitudinal cohort sequential design. Now, particularly as lifespan development becomes an accepted stipulative adjunct to most theories of psychological functioning, all researchers need to understand the implications of these different types of design.

1.2.4 Amount of manipulation

Research designs differ in the extent to which they rely upon the researcher manipulating the experience of participants in order to induce reactions. The fundamental differences between experimental, quasi-experimental, and other non-intrusive (or, more accurately, less intrusive) approaches are described in subsequent chapters in this book. For current purposes, it may be enough to highlight that researchers must make decisions about the nature of the interventions and control they will deploy in order to create the context in which they can study their target variables. Some research traditions eschew all manipulation and seek only to record naturally occurring phenomena. Others engage in highly elaborate environmental and social manipulation to create artificial but closely controlled conditions under which data are collected. Understanding where you are in any particular research design on this continuum of manipulation is important. It is important for two reasons. First, the greater the level of manipulation, the greater the degree of artificiality in the data and the greater the need to check whether the results can be generalized beyond the research context. Second, one of the most frequent reasons why research fails is that the manipulations used are inadequate. They can be inadequate in a number of ways:

- they can fail to reflect the construct or variable the impact of which the researcher wishes to study (for instance, the researcher wants to threaten the participant's sense of self-esteem and seeks to do this by providing false feedback of failure on an IQ test but the feedback represents what the participant would have expected to achieve);

- they can introduce unanticipated changes in ancillary variables that the researcher does not wish to study (for instance, the researcher wants the participant to focus on their family history and presents a family photograph but the photo includes in the background a fair and the participant focuses upon the fairground not the family element in the stimulus);
- they can fail to mean the same thing to the participant as they mean to the researcher (for instance, the researcher wishes to frighten someone and uses a manipulation involving the sudden presentation of a large spider, but the participant finds spiders not frightening but comic).

Ironically, the problem with manipulation is that it is difficult to control. Care in choosing manipulations pays dividends. **Manipulation checks** are now the norm in good research. These are designed to test whether the manipulation you think you introduced worked in the way you thought it would. In evaluating the research of others it is always a good idea to assess the effectiveness of the manipulations used.

Manipulation checks

1.2.5 Treatment of data as qualitative or quantitative

Research methods can be differentiated according to whether data are submitted to a qualitative or quantitative treatment. A **qualitative treatment** describes what processes are occurring and details differences in the character of these processes over time. A **quantitative treatment** states what the processes are, how often they occur, and what differences in their magnitude can be measured over time. Subsequent chapters in this book deal in detail with data treatment techniques and this is not the place to go into them in detail.

qualitative treatment

quantitative treatment

It is important to reiterate that these five dimensions on which a piece of research can be described are independent of each other. Data type, elicitation technique, the design for monitoring change, amount of manipulation and the qualitative or quantitative treatment of the data can be put together in many varieties. For instance, it is possible to use a qualitative treatment of data acquired as part of an experiment conducted in a longitudinal study.

A researcher, in structuring a study along these five dimensions, will have to make hard decisions. The decisions will in part be determined by whether theory building is at an inductive or deductive phase. A broader range of data types, elicitation techniques with lower control, cross-sectional designs, and qualitative treatment of data may be most appropriate in the early inductive phase. The deductive phase leading to testable predictions is likely to be linked to the narrowing of data types, direct and controlled data elicitation, a mixture of change monitoring designs, and the quantitative treatment of data. Sadly the decision is also too often influenced by preconceptions, prejudices and fears. Researchers get trapped into one methodological approach (i.e. a package of one type of data, one

elicitation technique, one design, and one data treatment). Once a routine sets in this can be easier than getting to know (or even remembering) how to do the other things. Also, of course, often researchers acquire their reputation on the basis of using a specific sort of methodology. To relinquish it is tantamount to abandoning their claim to fame. The solution may lie in practising eclecticism of methodology from an early stage in a research career.

Such eclecticism is fostered by forcing yourself, when faced with the task of constructing any study to test a hypothesis deduced from your theory, to provide at least two realistic alternative methodologies. Then weigh the pros and cons of each. Work out the differences between what they will tell you. In most cases, even minor variations of methodology will substantially affect what you can conclude. Ultimately, researchers have to choose between alternative feasible methodologies in the full knowledge of what they might lose by passing over those which they reject. The chapters in this book make an attempt to help you to see what are the strengths and weaknesses of various techniques, designs and data treatments.

1.3 INTEGRATING FINDINGS FROM DIFFERENT METHODOLOGIES

If you understand different methodologies and use them in concert, there comes a point when you must ask yourself: how do I put the findings from one methodology together with those from another? The easy answer focuses upon the theory. Assuming that each methodology is used to test one or more hypotheses derived from the theory, as long as the various methodologies yield conclusions which are compatible with the theory there is no problem. They are merely vehicles for theory testing; they may travel by different routes but they get to the same destination ultimately.

The problems arise when different methodologies produce contradictory or inconsistent conclusions about the hypothesis tested. In the baldest terms, one may support the hypothesis, whereas another may generate evidence which indicates that it is incorrect. The first step in this situation is to check that the methodologies were both executed properly. If they were, you should, if possible, collect further data using the same methodologies. If the inconsistent result is repeated, it is necessary to examine whether there is some identifiable attribute differentiating between the methodologies which could explain their inconsistent results. If such an attribute can be identified, it should be incorporated into another study in a controlled way so that its effect can be studied systematically. This may support the introduction of some caveat into the original hypothesis. If no such attribute can be identified, the hypothesis should be retested using a series of completely different methodologies. If these yield contradictory evidence, it is reasonably certain that the hypothesis will need to

be reformulated. The combination of evidence from the various methodologies should show where its limitations lie and point to an appropriate revision.

Obviously, all this procedure of iterative data collection takes time and resources. The researcher will have to decide whether this aspect of the theory is sufficiently important to merit such effort. If the procedure is not followed, it is essential that the original finding which refuted the hypothesis is treated seriously. The temptation to dismiss the finding in such a situation must be resisted. There are many siren voices which will offer ways of discounting the finding in terms of the relative merits of the methodologies. Unless you clearly stated on an a priori basis that one methodology would be given priority in the event of inconsistencies in the findings, the methodologies must be treated retrospectively as having equivalent standing.

When there are inconsistent results, an integrated approach to the use of several methodologies may be inconvenient but it also has great advantages. Every methodology has its limitations. The nature of these limitations differs. Using a series of methodologies allows you to compensate for the weaknesses of one methodology in a domain by supplementing or complementing it with another methodology which is stronger in that domain. The development of a coherent strategy for integrating methodologies, designed to test clearly defined hypotheses comprehensively, is the basic foundation for researching psychological processes.

1.4 FURTHER READING

There are some excellent handbooks that offer comprehensive coverage of the central issues. Denzin and Lincoln (2005) provides a clear and concise introduction to the major methods of qualitative research, with details of how data can be interpreted. Scott and Xie (2006) provides a fundamental introduction to the key quantitative methods in the social sciences beyond the discipline of psychology, assuming no prior knowledge of the statistical methods necessary to analyse quantitative data. Scott (2006) presents the variety of ways in which documentary evidence is interpreted, and is valuable particularly because it shows how texts are used by scholars outside social science, for instance in literature or history. Finally, Smith (2005) is useful for those who would unearth some of the greater complexities of the philosophical arguments that underlie the choice of a method of data collection or analysis.