PART I
WHAT IS RESEARCH DESIGN?

1
THE CONTEXT OF DESIGN

Before examining types of research designs it is important to be clear about the role and purpose of research design. We need to understand what research design is and what it is not. We need to know where design fits into the whole research process from framing a question to finally analysing and reporting data. This is the purpose of this chapter.

Description and explanation

Social researchers ask two fundamental types of research questions:

1. What is going on (descriptive research)?
2. Why is it going on (explanatory research)?

Descriptive research

Although some people dismiss descriptive research as 'mere description', good description is fundamental to the research enterprise and it has added immeasurably to our knowledge of the shape and nature of our society. Descriptive research encompasses much government sponsored research including the population census, the collection of a wide range of social indicators and economic information such as household expenditure patterns, time use studies, employment and crime statistics and the like.

Descriptions can be concrete or abstract. A relatively concrete description might describe the ethnic mix of a community, the changing age profile of a population or the gender mix of a workplace. Alternatively
the description might ask more abstract questions such as 'Is the level of social inequality increasing or declining?', 'How secular is society?' or 'How much poverty is there in this community?'.

Accurate descriptions of the level of unemployment or poverty have historically played a key role in social policy reforms (Marsh, 1982). By demonstrating the existence of social problems, competent description can challenge accepted assumptions about the way things are and can provoke action.

Good description provokes the 'why' questions of explanatory research. If we detect greater social polarization over the last 20 years (i.e. the rich are getting richer and the poor are getting poorer) we are forced to ask 'Why is this happening?' But before asking 'why?' we must be sure about the fact and dimensions of the phenomenon of increasing polarization. It is all very well to develop elaborate theories as to why society might be more polarized now than in the recent past, but if the basic premise is wrong (i.e. society is not becoming more polarized) then attempts to explain a non-existent phenomenon are silly.

Of course description can degenerate to mindless fact gathering or what C.W. Mills (1959) called 'abstracted empiricism'. There are plenty of examples of unfocused surveys and case studies that report trivial information and fail to provoke any 'why' questions or provide any basis for generalization. However, this is a function of inconsequential descriptions rather than an indictment of descriptive research itself.

**Explanatory research**

Explanatory research focuses on why questions. For example, it is one thing to describe the crime rate in a country, to examine trends over time or to compare the rates in different countries. It is quite a different thing to develop explanations about why the crime rate is as high as it is, why some types of crime are increasing or why the rate is higher in some countries than in others.

The way in which researchers develop research designs is fundamentally affected by whether the research question is descriptive or explanatory. It affects what information is collected. For example, if we want to explain why some people are more likely to be apprehended and convicted of crimes we need to have hunches about why this is so. We may have many possibly incompatible hunches and will need to collect information that enables us to see which hunches work best empirically.

Answering the 'why' questions involves developing causal explanations. Causal explanations argue that phenomenon Y (e.g. income level) is affected by factor X (e.g. gender). Some causal explanations will be simple while others will be more complex. For example, we might argue that there is a direct effect of gender on income (i.e. simple gender discrimination) (Figure 1.1a). We might argue for a causal chain, such as that gender affects choice of field of training which in turn affects occupational options, which are linked to opportunities for promotion, which in turn affect income level (Figure 1.1b). Or we could posit a more complex model involving a number of interrelated causal chains (Figure 1.1c).

**Prediction, correlation and causation**

People often confuse correlation with causation. Simply because one event follows another, or two factors co-vary, does not mean that one causes the other. The link between two events may be coincidental rather than causal.

There is a correlation between the number of fire engines at a fire and the amount of damage caused by the fire (the more fire engines the more damage). Is it therefore reasonable to conclude that the number of fire engines causes the amount of damage? Clearly the number of fire engines and the amount of damage will both be due to some third factor — such as the seriousness of the fire.

Similarly, as the divorce rate changed over the twentieth century the crime rate increased a few years later. But this does not mean that divorce causes crime. Rather than divorce causing crime, divorce and crime rates might both be due to other social processes such as secularization, greater individualism or poverty.
WHAT IS RESEARCH DESIGN?

Students at fee paying private schools typically perform better in their final year of schooling than those at government funded schools. But this need not be because private schools produce better performance. It may be that attending a private school and better final-year performance are both the outcome of some other cause (see later discussion).

Confusing causation with correlation also confuses prediction with causation and prediction with explanation. Where two events or characteristics are correlated we can predict one from the other. Knowing the type of school attended improves our capacity to predict academic achievement. But this does not mean that the school type affects academic achievement. Predicting performance on the basis of school type does not tell us why private school students do better. Good prediction does not depend on causal relationships. Nor does the ability to predict accurately demonstrate anything about causality.

Recognizing that causation is more than correlation highlights a problem. While we can observe correlation we cannot observe cause. We have to infer cause. These inferences however are 'necessarily fallible . . . [they are only indirectly linked to observables' (Cook and Campbell, 1979: 10). Because our inferences are fallible we must minimize the chances of incorrectly saying that a relationship is causal when in fact it is not. One of the fundamental purposes of research design in explanatory research is to avoid invalid inferences.

DETERMINISTIC AND PROBABILISTIC CONCEPTS OF CAUSATION

There are two ways of thinking about causes: deterministically and probabilistically. The smoker who denies that tobbaco causes cancer because he smokes heavily but has not contracted cancer illustrates deterministic causation. Probabilistic causation is illustrated by health authorities who point to the increased chances of cancer among smokers.

Deterministic causation is where variable X is said to cause Y if, and only if, X necessarily produces Y. That is, when X is present then Y will 'necessarily, inevitably and infallibly' occur (Cook and Campbell, 1979: 14). This approach seeks to establish causal laws such as: whenever water is heated to 100 °C it always boils.

In reality laws are never this simple. They will always specify particular conditions under which that law operates. Indeed a great deal of scientific investigation involves specifying the conditions under which particular laws operate. Thus, we might say that at sea level heating pure water to 100 °C will always cause water to boil.

Alternatively, the law might be stated in the form of 'other things being equal' then X will always produce Y. A deterministic version of the relationship between race and income level would say that other things being equal (age, education, personality, experience etc.) then a white person will always earn a higher income than a black person. That is, race (X) causes income level (Y).

Stated like this the notion of deterministic causation in the social sciences sounds odd. It is hard to conceive of a characteristic or event that will invariably result in a given outcome even if a fairly tight set of conditions is specified. The complexity of human social behaviour and the subjective, meaningful and voluntaristic components of human behaviour mean that it will never be possible to arrive at causal statements of the type 'if X, and A and B, then Y will always follow.'

Most causal thinking in the social sciences is probabilistic rather than deterministic (Suppes, 1970). That is, we work at the level that a given factor increases (or decreases) the probability of a particular outcome, for example: being female increases the probability of working part time; race affects the probability of having a high status job.

We can improve probabilistic explanations by specifying conditions under which X is less likely and more likely to affect Y. But we will never achieve complete or deterministic explanations. Human behaviour is both willed and caused: there is a double-sided character to human social behaviour. People construct their social world and there are creative aspects to human action but this freedom and agency will always be constrained by the structures within which people live. Because behaviour is not simply determined we cannot achieve deterministic explanations. However, because behaviour is constrained we can achieve probabilistic explanations. We can say that a given factor will increase the likelihood of a given outcome but there will never be certainty about outcomes.

Despite the probabilistic nature of causal statements in the social sciences, much popular, ideological and political discourse translates these into deterministic statements. Findings about the causal effects of class, gender or ethnicity, for example, are often read as if these factors invariably and completely produce particular outcomes. One could be forgiven for thinking that social science has demonstrated that gender completely and invariably determines position in society, roles in families, values and ways of relating to other people.

Theory testing and theory construction

Attempts to answer the 'why' questions in social science are theories. These theories vary in their complexity (how many variables and links), abstraction and scope. To understand the role of theory in empirical research it is useful to distinguish between two different styles of research: theory testing and theory building (Figure 1.2).

Theory building

Theory building is a process in which research begins with observations and uses inductive reasoning to derive a theory from these observations.
These theories attempt to make sense of observations. Because the theory is produced after observations are made it is often called post factum theory (Merton, 1968) or ex post facto theorizing.

This form of theory building entails asking whether the observation is a particular case of a more general factor, or how the observation fits into a pattern or a story. For example, Durkheim observed that the suicide rate was higher among Protestants than Catholics. But is religious affiliation a particular case of something more general? Of what more general phenomenon might it be an indicator? Are there other observations that shed light on this? He also observed that men were more suicidal than women, urban dwellers more than rural dwellers and the socially mobile more than the socially stable. He argued that the common factor behind all these observations was that those groups who were most suicidal were also less well socially integrated and experienced greater ambiguity about how to behave and what is right and wrong. He theorized that one of the explanations for suicidal behaviour was a sense of normlessness – a disconnectedness of individuals from their social world. Of course, there may have been other ways of accounting for these observations but at least Durkheim’s explanation was consistent with the facts.

Theory testing

In contrast, a theory testing approach begins with a theory and uses theory to guide which observations to make; it moves from the general to the particular. The observations should provide a test of the worth of the theory. Using deductive reasoning to derive a set of propositions from the theory does this. We need to develop these propositions so that if the theory is true then certain things should follow in the real world. We then assess whether these predictions are correct. If they are correct the theory is supported. If they do not hold up then the theory needs to be either rejected or modified.

For example, we may wish to test the theory that it is not divorce itself that affects the wellbeing of children but the level of conflict between parents. To test this idea we can make predictions about the wellbeing of children under different family conditions. For the simple theory that it is parental conflict rather than divorce that affects a child’s wellbeing there are four basic ‘conditions’ (see Figure 1.3). For each ‘condition’ the theory would make different predictions about the level of children’s wellbeing that we can examine.

If the theory that it is parental conflict rather than parental divorce is correct the following propositions should be supported:

- Proposition 1:children in situations (a) and (b) would be equally well off That is, where parental conflict is low, children with divorced parents will do just as well as those whose parents are married.
- Proposition 2:children in situations (c) and (d) should be equally poorly off That is, children in conflictual couple families will do just as badly as children in post-divorce families where parents sustain high conflict.
- Proposition 3:children in situation (c) will do worse than those in situation (a) That is, those with married parents in high conflict will do worse than those who have married parents who are not in conflict.
- Proposition 4:children in situation (d) will do worse than those in situation (b) That is, those with divorced parents in high conflict will do worse than those who have divorced parents who are not in conflict.
- Proposition 5:children in situation (b) will do better than those in situation (c) That is, children with divorced parents who are not in conflict will do better than those with married parents who are in conflict.
- Proposition 6:children in situation (a) will do better than those in situation (d) That is, children with married parents who are not in conflict will do better than those with divorced parents who are in conflict.
No single proposition would provide a compelling test of the original theory. Indeed, taken on its own proposition 3, for example, would reveal nothing about the impact of divorce. However, taken as a package, the set of propositions provides a stronger test of the theory than any single proposition.

Although theory testing and theory building are often presented as alternative modes of research they should be part of one ongoing process (Figure 1.4). Typically, theory building will produce a plausible account or explanation of a set of observations. However, such explanations are frequently just one of a number of possible explanations that fit the data. While plausible they are not necessarily compelling. They require systematic testing where data are collected to specifically evaluate how well the explanation holds when subjected to a range of crucial tests.

What is research design?

How is the term ‘research design’ to be used in this book? An analogy might help. When constructing a building there is no point ordering materials or setting critical dates for completion of project stages until we know what sort of building is being constructed. The first decision is whether we need a high rise office building, a factory for manufacturing machinery, a school, a residential home or an apartment block. Until this is done we cannot sketch a plan, obtain permits, work out a work schedule or order materials.

Similarly, social research needs a design or a structure before data collection or analysis can commence. A research design is not just a work plan. A work plan details what has to be done to complete the project but the work plan will flow from the project’s research design. The function of a research design is to ensure that the evidence obtained enables us to answer the initial question as unambiguously as possible. Obtaining relevant evidence entails specifying the type of evidence needed to answer the research question, to test a theory, to evaluate a programme or to accurately describe some phenomenon. In other words, when designing research we need to ask: given this research question (or theory), what type of evidence is needed to answer the question (or test the theory) in a convincing way?

Research design ‘deals with a logical problem and not a logistical problem’ (Yin, 1989: 29). Before a builder or architect can develop a work plan or order materials they must first establish the type of building required, its uses and the needs of the occupants. The work plan flows from this. Similarly, in social research the issues of sampling, method of data collection (e.g. questionnaire, observation, document analysis), design of questions are all subsidiary to the matter of ‘What evidence do I need to collect?’

Too often researchers design questionnaires or begin interviewing far too early – before thinking through what information they require to answer their research questions. Without attending to these research design matters at the beginning, the conclusions drawn will normally be weak and unconvincing and fail to answer the research question.

Design versus method

Research design is different from the method by which data are collected. Many research methods texts confuse research designs with methods. It is not uncommon to see research design treated as a mode of data collection rather than as a logical structure of the inquiry. But there is nothing intrinsic about any research design that requires a particular method of data collection. Although cross-sectional surveys are frequently equated with questionnaires and case studies are often equated with participant observation (e.g. Whyte’s Street Corner Society, 1943), data for any design can be collected with any data collection method (Figure 1.5). How the data are collected is irrelevant to the logic of the design.

Failing to distinguish between design and method leads to poor evaluation of designs. Equating cross-sectional designs with questionnaires, or case studies with participant observation, means that the designs are often evaluated against the strengths and weaknesses of the method rather than their ability to draw relatively unambiguous conclusions or to select between rival plausible hypotheses.
a point of confusion ... has been the unfortunate linking between the case study method and certain types of data collection – for example those focusing on qualitative methods, ethnography, or participant observation. People have thought that the case study method required them to embrace these data collection methods ... On the contrary, the method does not imply any particular form of data collection – which can be qualitative or quantitative. (1993: 32)

Similarly, Marsh (1982) argues that quantitative surveys can provide information and explanations that are ‘adequate at the level of meaning’. While recognizing that survey research has not always been good at tapping the subjective dimension of behaviour, she argues that:

Making sense of social action ... is ... hard and surveys have not traditionally been very good at it. The earliest survey researchers started a tradition ... of bringing the meaning from outside, either by making use of the researcher’s stock of plausible explanations ... or by bringing it from subsidiary in-depth interviews sprinkling quotes ... liberally on the raw correlations derived from the survey. Survey research became much more exciting ... when it began including meaningful dimensions in the study design. [This has been done in] two ways, firstly [by] asking the actor either for her reasons directly, or to supply information about the central values in her life around which we may assume she is orienting her life. [This] involves collecting a sufficiently complete picture of the context in which an actor finds herself that a team of outsiders may read off the meaningful dimensions. (1982: 123–4)

Adopting a sceptical approach to explanations

The need for research design stems from a sceptical approach to research and a view that scientific knowledge must always be provisional. The purpose of research design is to reduce the ambiguity of much research evidence.

We can always find some evidence consistent with almost any theory. However, we should be sceptical of the evidence, and rather than seeking evidence that is consistent with our theory we should seek evidence that provides a compelling test of the theory.

There are two related strategies for doing this: eliminating rival explanations of the evidence and deliberately seeking evidence that could disprove the theory.

Plausible rival hypotheses

A fundamental strategy of social research involves evaluating ‘plausible rival hypotheses’. We need to examine and evaluate alternative ways of explaining a particular phenomenon. This applies regardless of whether the data are quantitative or qualitative; regardless of the particular research design (experimental, cross-sectional, longitudinal or case
But these data are not compelling. There are at least three other ways of accounting for this correlation without accepting the causal link between school type and achievement (Figure 1.6). There is the selectivity explanation: the more able students may be sent to fee paying private schools in the first place. There is the family resources explanation: parents who can afford to send their children to fee paying private schools can also afford other help (e.g. books, private tutoring, quiet study space, computers). It is this help rather than the type of school that produces the better performance of private school students. Finally, there is the family values explanation: parents who value education most are prepared to send their children to fee paying private schools and it is this family emphasis on education, not the schools themselves, that produces the better academic performance. All these explanations are equally consistent with the observation that private school students do better than government school students. Without collecting further evidence we cannot choose between these explanations and therefore must remain open minded about which one makes most empirical sense.

There might also be methodological explanations for the finding that private school students perform better academically. These methodological issues might undermine any argument that a causal connection exists. Are the results due to questionable ways of measuring achievement? From what range and number of schools were the data obtained? On how many cases are the conclusions based? Could the pattern simply be a function of chance? These are all possible alternative explanations for the finding that private school students perform better.

Good research design will anticipate competing explanations before collecting data so that relevant information for evaluating the relative merits of these competing explanations is obtained. In this example of schools and academic achievement, thinking about alternative plausible hypotheses beforehand would lead us to find out about the parents' financial resources, the study resources available in the home, the parents' and child's attitudes about education and the child's academic abilities before entering the school.

The fallacy of affirming the consequent Although evidence may be consistent with an initial proposition it might be equally consistent with a range of alternative propositions. Too often people do not even think of the alternative hypotheses and simply conclude that since the evidence is consistent with their theory then the theory is true. This form of reasoning commits the logical fallacy of affirming the consequent. This form of reasoning has the following logical structure:

- If A is true then B should follow.
- We observe B.
- Therefore A is true.
If we apply this logic to the type of school and achievement proposition, the logical structure of the school type and achievement argument becomes clearer.

**Initial proposition:**
- Private schools produce better students than do government schools.

**The test:**
- **If** $A$ **then** $B$. If private schools produce better students ($A$) then their students should get better final marks than those from government funded schools ($B$).
- **$B$ is true**. Private school students do achieve better final marks than government school students (observe B).
- **Therefore $A$ is true**. Therefore private schools do produce better students ($A$ is true).

But as I have already argued, the better performance of private school students might also reflect the effect of other factors. The problem here is that any number of explanations may be correct and the evidence does not help rule out many of these. For the social scientist this level of indeterminacy is quite unsatisfactory. In effect we are only in a position to say:

- If $A$ or $C$, or $D$, or $E$, or $F$, or ... then $B$.
- We observe $B$.
- Therefore $A$ or $C$, or $D$, or $E$, or $F$, or ... is true.

Although explanation ($A$) is still in the running because it is consistent with the observations, we cannot say that it is the most plausible explanation. We need to test our proposition more thoroughly by evaluating the worth of the alternative propositions.

**Falsification: looking for evidence to disprove the theory**

As well as evaluating and eliminating alternative explanations we should rigorously evaluate our own theories. Rather than asking 'What evidence would constitute support for the theory?', ask 'What evidence would convince me that the theory is wrong?' It is not difficult to find evidence consistent with a theory. It is much tougher for a theory to survive the test of people trying to disprove it.

Unfortunately some theories are closed systems in which any evidence can be interpreted as support for the theory. Such theories are said to be non-falsifiable. Many religions or belief systems can become closed systems whereby all evidence can be accommodated by the theory and nothing will change the mind of the true believer. Exchange theory (Homans, 1961; Blau, 1964) is largely non-falsifiable. It assumes that we always maximize our gains and avoid costs. But we can see almost anything as a gain. Great sacrifices to care for a disabled relative can be interpreted as a gain (satisfaction of helping) rather than a loss (income, time for self etc.). We need to frame our propositions and define our terms in such a way that they are capable of being disproven.

**The provisional nature of support for theories**

Even where the theory is corroborated and has survived attempts to disprove it, the theory remains provisional:

> falsificationism stresses the ambiguity of confirmation ... corroboration gives only the comfort that the theory has been tested and survived the test, that even after the most impressive corroborations of predictions it has only achieved the status of 'not yet disconfirmed'. This ... is far from the status of 'being true'. (Cook and Campbell, 1979: 20)

There always may be an unthought-of explanation. We cannot anticipate or evaluate every possible explanation. The more alternative explanations that have been eliminated and the more we have tried to disprove our theory, the more confidence we will have in it, but we should avoid thinking that it is proven.

However we can disprove a theory. The logic of this is:

- If theory A is true then B should follow.
- B does not follow.
- Therefore A is not true.

So long as B is a valid test of A the absence of B should make us reject or revise the theory. In reality, we would not reject a theory simply because a single fact or observation does not fit. Before rejecting a plausible theory we would require multiple disconfirmations using different measures, different samples and different methods of data collection and analysis.

In summary, we should adopt a sceptical approach to explanations. We should anticipate rival interpretations and collect data to enable the winnowing out of the weaker explanations and the identification of which alternative theories make most empirical sense. We also need to ask what data would challenge the explanation and collect data to evaluate the theory from this more demanding perspective.
Summary

This chapter has outlined the purpose of research design in both descriptive and explanatory research. In explanatory research the purpose is to develop and evaluate causal theories. The probabilistic nature of causation in social sciences, as opposed to deterministic causation, was discussed.

Research design is not related to any particular method of collecting data or any particular type of data. Any research design can, in principle, use any type of data collection method and can use either quantitative or qualitative data. Research design refers to the structure of an enquiry: it is a logical matter rather than a logistical one.

It has been argued that the central role of research design is to minimize the chance of drawing incorrect causal inferences from data. Design is a logical task undertaken to ensure that the evidence collected enables us to answer questions or to test theories as unambiguously as possible. When designing research it is essential that we identify the type of evidence required to answer the research question in a convincing way. This means that we must not simply collect evidence that is consistent with a particular theory or explanation. Research needs to be structured in such a way that the evidence also bears on alternative rival explanations and enables us to identify which of the competing explanations is most compelling empirically. It also means that we must not simply look for evidence that supports our favourite theory: we should also look for evidence that has the potential to disprove our preferred explanations.

TOOLS FOR RESEARCH DESIGN

To achieve a reasonable research design we need to attend to a number of matters before we arrive at the final design. The first section of this chapter outlines these preliminary steps that precede design. It then expands on the idea of alternative rival hypotheses that was introduced in Chapter 1. The second section introduces a number of concepts that are fundamental to designing good research — internal validity, external validity and measurement error.

Before design

In the same way that an architect needs to know the purpose of the building before designing it (is it an office building, a factory or a home?) social researchers must be clear about their research question before developing a research design.

Focusing and clarifying the research question

The first question to ask is, 'What question am I trying to answer?' Specifying a question is more than identifying a topic. It's not enough to say, 'I'm interested in getting some answers about family breakdown.' What answers to what questions? Do you want to know the extent of family breakdown? Who is most vulnerable to family breakdown? Changing rates of breakdown? Over what period? Where? Or are you really looking at the causes of breakdown? The effects of family breakdown? All the effects or just particular ones?

FOCUSING DESCRIPTIVE RESEARCH QUESTIONS

To narrow the focus of descriptive research we need to specify the scope of what is to be described. The following guidelines, using family breakdown as an example, help narrow down a descriptive research topic into a researchable question.

1. What is the scope of the core concepts? What is to be included in the concept family breakdown? Do we mean divorce? What about
separation? Are we referring only to the breakdown in marital relationships? Do we mean the total breakdown or simply poor relationships? What about relationships between parents and children or between children? Until we specify what we mean by our core concepts it is going to be impossible to begin the description.

2. What is the time frame for the description? Is our interest in change over time or just about contemporary levels of family breakdown? If it is about change, over what period?

3. What is the geographical location for the description? Is the interest in family breakdown in a particular community, in different regions or the whole nation? Is it comparative, looking at breakdown in different types of countries (e.g. highly industrialized versus rapidly industrializing versus impoverished countries)?

4. How general is the description to be? Do you want to be able to describe patterns for specific subgroups (e.g. among those who married as teenagers, among those who are in de facto relationships, second marriages etc.)?

5. What aspect of the topic are you interested in? Is the interest in rates of breakdown? The experience? Laws? Attitudes and beliefs?

6. How abstract is your interest? Is your interest in family breakdown or in family breakdown as a reflection of something more abstract (e.g. social fragmentation, social conflict, individualism, the role of the state in the private lives of citizens)?

7. What is the unit of analysis? The unit of analysis is the ‘thing’ about which we collect information and from which we draw conclusions. Often this is a person (e.g. divorced person) but it may be ‘things’ such as organizations (divorce courts), a family as a whole, events (e.g. divorces), periods (divorce in different years), places (communities, countries).

The questions we can answer will depend on the unit of analysis. We could compare divorced individuals with non-divorced individuals (individual as the unit of analysis). We could study a series of divorces and examine what the process of becoming divorced was (event as unit of analysis). We might use year as the unit of analysis and track changes in divorce rates since 1945. Using countries we might examine the different divorce rates in different types of countries with a view to comparing the patterns in different types of countries. Alternatively families might be the units of analysis and we may want to look at the characteristics of divorcing families (e.g. size, family income, family type, nature of relationships in family) compared with those of non-divorcing families.

Thinking beyond individuals as units of analysis broadens the range of research questions we ask and broadens the range and sources of data available. For example, if years were the unit of analysis we would obtain statistics from the relevant national collection agencies regarding divorce for each year. We would also collect other information about each year (e.g. unemployment level, inflation rate, changes to laws) that was relevant to the hypotheses.

Focusing explanatory research questions

In framing explanatory questions we need to further specify our focus. Explanatory research explores causes and/or consequences of a phenomenon, so the research question must be clear about the style of explanatory research and identify which causes or consequences it will investigate.

Before outlining some different types of explanatory research it is useful to introduce some terms.

- **Dependent variable** This is the variable that is treated as the effect in the causal model: it is dependent on the influence of some other factor. The dependent variable is also referred to as the outcome variable and in causal diagrams it is conventionally designated as the Y variable.

- **Independent variable** This is the variable that is the presumed cause. It is also called the predictor variable, the experimental variable or the explanatory variable and is designated in causal diagrams as the X variable (as in education (X) – income level (Y)).

- **Intervening variables** These variables come between the independent variable and the dependent variable in a causal chain. They are the means by which cause X produces effect Y. Intervening variables are represented in causal diagrams by the symbol Z, as illustrated in Figure 2.1.

- **Extraneous variables** Two variables can be correlated without being causally related. The correlation may be due to the two factors being outcomes of a third variable (see Chapter 1). This third variable is called an extraneous variable and is also symbolized as Z in causal diagrams, and the form of this relationship is illustrated in Figure 2.2.

Searching for causes or effects This is the least focused type of explanatory research. It involves identifying the core phenomenon (e.g. changes in divorce rate since World War II) and then searching for causes or consequences of this. Searching for causes would involve identifying possible causal factors (e.g. changing values, decline in religion, changing population mix, economic changes, social reforms, changes in welfare policy for lone parents). We would then design research to evaluate which of these causes helps explain changes in divorce rates. This form of research question is illustrated in Figure 2.3.

Alternatively we might focus on the consequences rather than causes of changes in divorce rate (Figure 2.4).
Exploring a simple causal proposition. A more focused research question will specify a particular causal proposition to evaluate (Figure 2.5). It might propose an impact of a particular factor or examine a specific consequence. For example, we might propose that changes in government benefits to lone parents have led to an increase in divorce rates since World War II.

More complex causal models. Such propositions are simplistic in that they do not spell out the mechanisms by which the two factors might be related. We might develop more complex models that spell out some of the mechanisms in the causal chain. This fuller model then becomes the focus of the research and provides the framework within which the research design will be framed (Figure 2.6).

When clarifying a research question it is helpful to draw diagrams like those in Figures 2.1–2.5. It is also helpful to ask four key questions:

1. What am I trying to explain (i.e., what is the dependent variable)?
2. What are the possible causes (what are the independent variables)?
3. Which causes will I explore?

4. What possible mechanisms connect the presumed causes to the presumed effects (what are the intervening variables)?

Another way of framing research questions is to formulate different ways of understanding a phenomenon and then compare which of these competing approaches best fits the facts. For example, we might compare three different ways of understanding changes in divorce rates: an economic approach, a social values approach and a legislative approach. The economic approach might argue that changes in divorce rates stem from economic factors such as the level of affluence, access of women to work and levels of welfare support for lone parents. The social values explanation might seek to explain the increased divorce rate in terms of increasing levels of individualism and greater social acceptance of divorce. A legislative approach might focus on the impact of legislative changes such as no-fault divorce or new rules regarding property division.
Idiographic and nomothetic explanations

The above examples focus on particular causes and consequences and attempt to provide only a partial explanation of changes in divorce rates. These explanations are called nomothetic explanations. Partial explanations of a class of cases rather than a fully explanation of a particular case. They involve an examination of relatively few causal factors and a larger number of cases. In contrast, idiographic explanations focus on particular cases and develop as complete an explanation of each case as possible. They involve examining as many factors as possible that contribute to the case including unique factors.

A nomothetic explanation of divorce might focus on the importance of a single contributing factor such as the intergenerational transmission of divorce. Are couples more likely to divorce if one of their parents have divorced? The study might involve comparing people who had divorced with those who had not to see if the divorcees were more likely to have a family history of divorce. In contrast, an idiographic approach might focus on a particular divorced couple and build a full explanation of why this couple divorced. The explanation would consider the family history of the couple and examine this along with many other contributing factors. The idiographic explanation would provide us with a good understanding of the case while the nomothetic explanation would provide an understanding of the influence of a factor.

Identifying plausible rival hypotheses

In Chapter 1 I stressed the importance of identifying plausible rival ways of accounting for the phenomenon being studied. Obviously it makes sense to anticipate these alternatives before carrying out the research. But how do we identify these alternative explanations? There are two main types of rival hypotheses and these suggest ways of anticipating alternative explanations.

Theoretical and substantive rivals

There is no magical way of producing a set of alternative substantive or theoretical explanations. In the end the researcher must formulate them. The more familiar the researcher is with the particular substantive topic and with social science models and theories, the more likely they are to anticipate different ways of interpreting a given set of results. The following provide sources of alternative explanations.

The theoretical literature. Broad approaches in a discipline can present different ways of viewing any question. Suppose, for example, that we wanted to understand why some people seem to have happier marriages than other people do. An explanation could concentrate on the personal attributes of the couple their personality, their values and beliefs and their interpersonal skills. It could focus on economic factors and interpret marital happiness in terms of the costs and benefits to each partner. Alternatively, the explanation could adopt a life course perspective that interprets differences in marital happiness according to where the couple is in the life course (e.g. newly married, before children, young children, adolescent children, empty nest, later life). A feminist approach might try to explain marital happiness in terms of gender roles, division of labour and power within the relationship. A demographer might focus on the birth cohorts (e.g. depression marriages, postwar marriages, 1990s marriages etc.). Other researchers might seek to explain marital happiness in structural terms such as the couple’s level of social disadvantage. Social network theorists might concentrate on the extent to which a couple is integrated into the wider family and community and is able to receive support from these networks.

This list is not exhaustive. The point is that different approaches (e.g. psychological, life course, exchange, structural, feminist) will have a particular ‘angle’ which alerts us to different ways of looking at an issue. When thinking about a problem, ask yourself questions such as ‘How might a feminist account for this?’ ‘How might a Freudian psychologist explain this?’ ‘How would an exchange theorist account for this?’ ‘What might a conflict theorist say?’

Other researchers Previous research on the topic can be a rich source of competing explanations. Read the literature in journals and search electronic databases. Review articles and introductory overviews can be extremely helpful.

Practitioners, key informants, policy makers, advocates ‘Insiders’ with practical knowledge of a field can be invaluable. In a study on divorce, marriage counsellors, married couples, divorced people and advocacy groups can provide valuable insights. Literature such as novels and plays can also provide keen ideas that can be tested systematically (e.g. Tolstoy’s novel Anna Karenina provides one way of interpreting marital unhappiness).

Own experience, hunches, and intuitions Do not ignore your own experience, your own intuitions and hunches. In the end all explanations start with hunches that spring from individuals who have ideas and observe things around them. Use these insights, experiences and observations and test them systematically.

There is no right way of developing ideas. Do not limit yourself to formal social science research or to research on the very specific topic you are working on. Use diverse sources of ideas. Brainstorm with a group and
debate the topic. Think laterally: if your topic is marriage breakdown, look beyond the marriage and divorce literature.

**Technical/methodological rivals**

If findings are likely to be due to poor measurement (see below) then any theoretical interpretation of these results will be unconvincing. Throughout this book I will identify many technical and methodological factors that can undermine the conclusions we draw from our research. Good research design will minimize the threat from these sources.

I will not at this point go into these methodological issues in detail. However, it will be helpful to highlight the types of methodological rivals that will be examined. These are outlined in Goldenberg (1992).

1. Demand characteristics of the situation.
2. Transient personal characteristics such as the respondent’s mood, attention span, health etc.
3. Situational factors such as anonymity, people present when data were being collected, gender, age and class of investigator and respondent.
4. Sampling of items: Are the concepts well measured?
5. Nature of the sample: Can we generalize from the sample?
6. Lack of clarity of the instrument: Are the questions clear and unambiguous?
7. Format of data collection: Are the results an artifact of the data collection method? Would different patterns be found if a different method was used (e.g., observation rather than questionnaire)?
8. Variation in the administration of the instrument in studies tracking change over time.

**Operationalization**

Most social science research involves making observations that we presume tap concepts. If we were conducting a study on the effect of marital breakdown on the wellbeing of children we would need first to work out what is meant by marriage breakdown, wellbeing and children. This involves defining these concepts, which in turn requires developing a nominal definition and an operational definition of each concept.

Concepts are, by their nature, not directly observable. We cannot see social class, marital happiness, intelligence etc. To use concepts in research we need to translate concepts into something observable—something we can measure. This involves defining and clarifying abstract concepts and developing indicators of them. This process of clarifying abstract concepts and translating them into specific, observable measures is called operationalization and involves descending the ladder of abstraction.

**Clarifying concepts**

Before developing indicators of concepts we must first clarify the concepts. This involves developing both nominal and operational definitions of the concept.

**Nominal definitions**

Concepts do not have a fixed or correct meaning. Marriage breakdown could be defined in terms of the law (such as whether the decree nisi has been granted), the quality of the relationship or practical arrangements (living apart). Similarly, we need to define children. Is a child defined by a blood relationship, a legal relationship (includes adopted), a social relationship (de facto parents), chronological age, level of dependency or some other criterion? By deciding on the type of definition we provide a nominal definition: it specifies the meaning of the concept but remains abstract.

Different definitions produce different findings. Consequently, defining concepts is a crucial stage of research. It needs to be done deliberately and to be systematically and carefully justified. There are three steps in developing and narrowing down a nominal definition.

First, obtain a range of definitions. Look at review articles, discipline dictionaries (e.g., a dictionary of sociology), encyclopaedias (e.g., an encyclopaedia of social sciences) and journal articles. Look for both explicit and implicit definitions.

Second, decide on a definition. From your list either select one definition or create a better definition from the common elements of several definitions. Explain and justify your approach.

Third, delineate the dimensions of the concept. Many concepts have a number of dimensions and it is helpful to spell these out as they can help to further refine your definition. This can be illustrated using the concept of the child’s wellbeing. We can think of a number of dimensions of wellbeing: emotional, psychological, physical, educational, financial, social, environmental, legal etc. If we are arguing that marriage breakdown affects a child’s wellbeing, what sort of wellbeing are we talking about? All of these? Just one or two? Having delineated various dimensions you will need to decide which are of interest in the present study. You may examine all aspects or limit yourself to one or two.

A concept may have subdimensions. Suppose we focused on social wellbeing. This broad concept could incorporate subdimensions such as the level of safety in the neighbourhood, the nature of the child’s relationships and her experiences of social discrimination. The subdimension of ‘relationships with others’ could be further divided into relationships with particular people such as mother, father, peers, siblings and grandparents. Having settled on the particular relationships with which we are going to deal, we would need to identify what aspects of the relationships to measure and decide how to measure them (see below). In Figure 2.7 relationships are measured according to the level of
contact, conflict, closeness, helping and type of shared activities. These measures or indicators would then provide the core around which to frame specific questions or to focus an observational study.

In this example I have focused on one dimension at each level but I could have been exhaustive and developed measures for each conceivable dimension and subdimension. The decision of whether to adopt a focused or an exhaustive approach will depend on what you are interested in. The advantage of adopting a systematic approach as outlined in Figure 2.7 is that it helps focus and refine the research question and forces you to make deliberate decisions about how to measure a concept.

Operational definitions Having defined the concepts you will develop an operational definition – the observations to measure the concept. What indicators of marriage breakdown will you use? We might operationally define marriage breakdown according to the quality of the relationship. Using this definition we might measure breakdown according to level of conflict, type of communication, signs of lack of affection and level of cooperation or lack of cooperation.

How well such indicators tap the concept as defined will have a critical bearing on the value of the conclusions drawn from the study. If the indicators tap something other than what we claim they do then our conclusions can be easily challenged. For example, if our measures of marital breakdown simply tap social class differences in marital style rather than breakdown then our conclusions about marital breakdown will be suspect.

Once the operational definition of the concept has been developed we come to the final stage of operationalization. This entails the precise way in which the indicators will be measured. This might involve developing questions for a questionnaire or identifying what and how observations will be made. Articulating the mechanics of these strategies is beyond the scope of this book.

Concepts for research design

Two concepts, internal and external validity, are fundamental to developing research designs. Ideally research designs should be both internally and externally valid.

Internal validity

We need to be confident that the research design can sustain the causal conclusions that we claim for it. The capacity of a research design to do this reflects its internal validity.
Imagine a research project that compares the emotional adjustment of children from divorced families and intact families. It finds that children from divorced families are less well adjusted than children from intact families. Can we conclude that divorce caused emotional maladjustment? Not on the basis of these results. The design does not enable us to eliminate alternative explanations. The poorer adjustment of children with divorced parents might be due to adjustment differences that predated parental divorce.

A different research study may deal with this problem by tracking children before parents divorce and for some years afterwards. It may find that these children do show a significant decline in emotional adjustment after their parents’ divorce. Does the research design show that divorce is producing this decline in adjustment? No. The decline in adjustment may simply reflect a general decline in emotional adjustment as children get older. The same decline may be evident among children from intact families.

Yet another study might try to overcome this problem by tracking changes in adjustment of children before and after their parents divorce and changes among children from intact families as well. If children from intact families show less deterioration in adjustment than children from divorcing families, this must surely demonstrate the effect of divorce. No. We would need to be sure that the two groups of children were comparable in other relevant respects (e.g. age). The different rates of change in adjustment could be because those from divorcing families were younger on average. Maybe younger children show greater changes in adjustment over a particular period than older children. It may be age differences rather than having divorced parents that account for the adjustment changes of the two groups of children.

Internal validity is the extent to which the structure of a research design enables us to draw unambiguous conclusions from our results. The way in which the study is set up (e.g. tracking changes over time, making comparisons between comparable groups) can eliminate alternative explanations for our findings. The more the structure of a study eliminates these alternative interpretations, the stronger the internal validity of the study. A central task of research design is to structure the study so that the ambiguities in the above examples are minimized. It is impossible to eliminate all ambiguities in social research but we can certainly reduce them.

External validity

External validity refers to the extent to which results from a study can be generalized beyond the particular study. A study may have good internal validity but its value is limited if the findings only apply to the people in that particular investigation. The critical question is whether the results are likely to apply more widely. The most common threat to

our capacity to generalize more widely from a research study is the use of unrepresentative samples. This, and other threats to external validity, will be discussed more fully in Chapters 5, 8, 11 and 14.

Measurement error

A further threat to the conclusions that can be drawn from any study is measurement error. This occurs when we use flawed indicators to tap concepts (see Chapter 1).

Types of measurement error

Indicators must meet two fundamental criteria. They must be both valid and reliable. A valid indicator in this context means that the indicator measures the concept we say it does. For example an IQ test is used to measure intelligence. If it really measures intelligence the test would be valid. If the IQ test measured something else instead, such as education level or cultural background, then it would be an invalid measure of intelligence.

Reliability means that the indicator consistently comes up with the same measurement. For example, if people consistently obtain the same IQ score on repeated IQ tests, then the test would be reliable. If their results fluctuate wildly depending on when they take the test, then it would be unreliable.

Validity

The earlier discussion of internal validity related to the validity of the research design. It addressed the question: is the research design delivering the conclusions that we claim it delivers? In addition we need to examine the validity of the measures used in any piece of research. The validity of a measure depends both on the use to which it is put and on the sample for which it is used. For example, the validity of using frequency of arguments between partners to measure marital happiness turns on what we mean by marital happiness. The validity of this measure may vary for different cultural groups and for the same cultural group in different historical periods. Measures of children’s emotional adjustment will vary according to their age and their cultural group.

There are three basic ways of assessing validity. Criterion validity is best suited to situations where there are well-established measures of a concept that need adapting, shortening or updating. It involves comparing the way people rate on the new measure with how they rate on well-established measures of the concept. If ratings on the new measure match those of an established measure we can be confident of its validity.

Criterion validity has two limitations. First, it requires that the established benchmark is valid. If the benchmark is invalid then it is of little value in assessing the new measure. Second, there are no established measures for many social science concepts.
Sometimes criterion groups can be used to assess criterion validity. Instead of comparing a measure against an existing benchmark measure, the new measure can be trialled. For example, a new measure of marital happiness could be trialled on couples who seek marital counselling. We would expect that this group of couples would normally obtain low scores on a valid measure of marital happiness. If these couples actually obtain high scores on the marital happiness measure we would probably want to question whether the measure was really tapping marital happiness.

Content validity evaluates how well the measures tap the different aspects of the concept as we have defined it. A test of arithmetic skills that only tested subtraction skills would clearly not be a valid measure of arithmetic skills. Similarly a measure of marital happiness that only asked about the frequency of arguments between partners would probably lack content validity unless we had defined marital happiness simply as the absence of arguments. Measures of marital happiness could also include the nature of the arguments, leisure activities shared by partners, communication, methods of resolving conflict, the quality of the sexual relationship etc.

Given disagreement about the ‘content’ of many social science concepts it can be difficult to develop measures that have agreed validity. Even if we can agree on the concept and measure it using a whole battery of questions, we then face the problem of the relative importance of the various components of the measure. For example, should measures of the frequency of arguments be as important as the nature of the arguments, the method of conflict resolution, the style of communication or statements about level of subjective marital satisfaction?

Construct validity relies on seeing how well the results we obtain when using the measure fit with theoretical expectations. To validate a measure of marital happiness we might anticipate, on the basis of theory, that happiness will vary in predictable ways according to stage in the life cycle. If the results of a study provide confirmation, this could reflect the validity of our measure of marital happiness. However, this approach to assessing validity relies on the correctness of our expectations. If our theory is not supported this could be for one of two reasons: the measure of marital happiness could be wrong or the theory against which the measure is being benchmarked may be wrong.

There is no ideal way of assessing validity. If a measure passes all three tests it is more likely to be valid but we cannot be certain. In the final analysis we will need to argue for the validity of our measures.

Reliability. A reliable measure is one that gives the same ‘reading’ when used on repeated occasions. For example, assuming there was no actual change, a reliable measure of marital happiness should yield the same ‘reading’ if it is used on different occasions. A thermometer that measured body temperature as 97.4° F one minute and 105° F the next would be useless. Which measurement is right? Does the change reflect real change or just measurement ‘noise’?

Unreliability can stem from many sources. Poor question wording may cause a respondent to understand the question differently on different occasions. Different interviewers can elicit different answers from a person: the match of age, gender, class and ethnicity of an interviewer and interviewee can influence responses. Asking questions about which people have no opinion, have insufficient information or require too precise an answer can lead to unreliable data. The answers to some questions can be affected by mood and by the particular context in which they are asked.

Measures need to be both valid and reliable. Although these two concepts are related they are not the same. A measure can be reliable without being valid. That is, a measure can be consistently wrong. For example, people consistently underestimate their level of alcohol consumption in questionnaire surveys. Alcohol consumption measures are reliable but do not accurately tell us about the true level of alcohol consumed.

Measures will never be perfectly reliable and perfectly valid. These are not all or nothing concepts and the goal is to maximize the reliability and validity. If these aspects of measurement are weak then the results of the study that uses them might plausibly be attributed to poor measurement rather than telling us anything about social reality.

FORMS OF MEASUREMENT ERROR

Error can take different forms and the consequences of error will vary depending on its form. These forms of error are random, constant and correlated.

Random error is that which has no systematic form. It means that in some cases a measurement for a variable might be too low while in others it is too high. The measurement of someone’s weight might display random error. Sometimes people underestimate their weight while others may overestimate it, but if these errors are random there will be the same number of over- and underestimates and the size of the overestimates will be the same on average as the underestimates. When the average (mean) is calculated for the whole group it will be accurate because the overestimates and the underestimates cancel each other out. Furthermore, these mis-estimates are not correlated with any other characteristic (e.g. gender, age) but are truly random. Because random error does not distort means and is uncorrelated with other factors, it is less serious than other forms of error.

Constant error occurs when there is the same error for every case. For example, if everyone underreported their weight by 5 kilograms we would have constant error. Such error is uncorrelated with other characteristics. Although purely constant error will be rare there will be
variables for which there will typically be a component of constant error (e.g. overstatements of frequency of sexual intercourse and understatement of the amount of alcohol consumed). Because such error is constant it does not cancel out but has an effect on sample estimates. Thus the average weight of the sample would be an underestimate to the extent of the constant error.

Correlated error takes place when the amount and direction of error vary systematically according to other characteristics of respondents. For example, if women tend to overestimate their weight while men underestimate theirs then this error would be correlated with gender. If the format or language in a questionnaire is difficult then mistakes in answering questions may be correlated with education. This would produce results that make it appear that people with different levels of education behave or think differently while in fact it is only their capacity to understand the question that differs.

A crucial goal of the design and administration of survey instruments is the minimization of the various forms of measurement error. Achieving this entails paying careful attention to question wording, indicator quality, interviewer and observer training, and to ways of identifying social desirability responses and other forms of deliberate misrepresentation by respondents. In many cases it is difficult to identify the extent to which such errors actually occur. However, this does not reduce the need to do all that one can to minimize their likelihood and to have built-in checks to identify some sources of error. Such checks include looking for inconsistencies in answers, using multiple questions rather than single questions to tap concepts, identifying social desirability response sets, making inter-interviewer checks and careful fieldwork supervision.

Summary

This chapter has emphasized the importance of clarifying research questions and concepts before developing a research design. Lack of clarity regarding the research question and the central research concepts will severely compromise any research design. Guidelines were provided to help focus both descriptive and explanatory research questions and to clarify the concepts they employ.

It is also unwise to develop a research design unless alternative ways of understanding the matter at the heart of the research question have been identified. Since one of the purposes of research design is to help identify which of a range of alternative explanations work best it is desirable that these alternative explanations be identified before the research design is developed. The design can then be structured in such a way that relevant data are collected to enable us to choose between these alternatives. Guidelines were provided to assist in identifying these alternative explanations.

Finally, three core concepts that are at the heart of good design were discussed. These were the concepts of internal validity, external validity and measurement error. Later chapters will evaluate the various designs using these concepts.
Establishing causal relationships is at the heart of explanatory research design. However, it is not a simple matter to establish that one event causes another (Blalock, 1964; Hage and Meeker, 1988). The main reason why it is difficult to establish causal relationships is because we cannot actually observe one phenomenon producing change in another. Even though one event might always follow another we do not know that this is because one event causes the other. Causal relationships must therefore be inferred rather than observed. The purpose of research design in explanatory research is to improve the quality of our causal inferences.

**Inferring causal relationships**

**Criteria for inferring cause**

In Chapter 1 I distinguished between probabilistic and deterministic concepts of causation. Probabilistic approaches to causation are those that argue that a given factor increases (or decreases) the probability of a particular outcome. For example, we may argue that there is a causal relationship between gender and working part time – that gender affects the probability of working part time.

In order to infer that a probabilistic causal relationship exists between two variables, two basic criteria must be met. First, there must be co-variation of causal and outcome variables (e.g. between gender and being a full time or part time worker); and second, the assertion that one variable affects the other must make sense.

**Co-variation**

If two factors are causally related they must at least be correlated: they must co-vary. If X causes Y then people who differ from one another on X should tend to differ from one another on Y. For example, if we were to argue that working in the private sector rather than the public sector makes people more achievement oriented at work we would, at the very least, need to demonstrate that private sector workers had a higher achievement orientation than public sector workers. If two factors did not co-vary – that is public and private sector workers had identical levels of work achievement – then we would be hard pressed to argue that they are causally related.

However, while co-variation is a precondition it is not enough for us to assert that the variables are causally related. Correlations can also reflect non-causal relationships. When two variables or events are correlated but not causally related the relationship between the two variables is said to be spurious (see Chapter 1). For example, the correlation between sector of employment and employment orientation might be due entirely to a third variable such as age (Figure 3.1). Younger people might be more likely than older people to work in the private sector and younger people might also have a higher achievement orientation than older people. These two patterns will mean that sector and achievement orientation are correlated (more young people in the private sector, so therefore the private sector is associated with achievement orientation). However the link between sector and achievement orientation is coincidental rather than causal.

**It must make sense**

Any assertion that co-variation reflects a causal relationship must be plausible. It must make sense at three levels.

**Time order** If two variables are correlated the cause must come before the effect. Causal reasoning has no time for the assertion that a future event can have a present effect (teleological explanation). Our causal proposition must be such that the causal variable occurs before the presumed effect. The time gap between cause and effect can be minutes or may be years (e.g. the effect of education on income can take many years to show itself).

Even though two variables might be causally related it can sometimes be difficult to work out which variable comes first and therefore to
establish which variable is the cause and which is the effect. For example, does sector of employment affect achievement orientation or is it the other way around? Even where we assert that one variable comes first the causal relationship may be two-way. That is, sector of employment may affect achievement orientation which in turn influences future decisions about the sector of employment in which one works. Causal relationships can be reciprocal (two-way) rather than one-way.

**Dependent variable must be capable of change.** If we say that a correlation between two variables is because one is causing the other, we must make sure that the dependent variable (the effect) is capable of being changed. If it cannot be changed then a causal account of the relationship makes no sense. For example, any causal relationship between sex and income could only be in the direction of sex affecting income. The opposite proposition (income—sex) makes no sense.

Theoretical plausibility. The causal assertion must make sense. We should be able to tell a story of how X affects Y if we wish to infer a causal relationship between X and Y. Even if we cannot empirically show how X affects Y we need to provide a plausible account of the connection (plausible in terms of other research, current theory etc.). For example, to support the assertion that sector of employment affects achievement orientation we might argue that the private sector fosters the development of an achievement orientation by strategies such as paying performance bonuses, developing a culture of higher expectations, providing better resources and creating less job security. When backed up by this type of reasoning, any correlation between employment sector and achievement orientation can be plausibly interpreted in causal terms.

**Types of causal patterns**

**Direct and indirect causal relationships**

Causal relationships can be either direct or indirect. A direct relationship is one where we assert that the cause affects the outcome directly rather than via other variables. An indirect causal relationship is one where the cause has its effect by operating via its influence on another variable that, in turn, produces the effect. The variable through which the two variables are related is called the intervening variable: it comes in time and in a causal sequence between the initial cause and the effect. For example, we might argue that the way the private sector produces higher achievement orientation is by making employees fear for their jobs (the intervening variable) (Figure 3.2).

Indirect causal relationships may be simple (as in Figure 3.2) or consist of an extended causal chain or a number of different causal paths (Figure 3.3).

---

**Figure 3.2 An indirect causal relationship**

a) Long causal chain

\[ X \rightarrow Z \rightarrow W \rightarrow V \rightarrow P \rightarrow Y \]

b) Multiple indirect paths

\[ X \rightarrow W \rightarrow V \rightarrow Y \]

\[ P \rightarrow Q \]

---

**Figure 3.3 More complex indirect causal relationships**

**Types of relationships in a three-variable model.**

Any relationship between two variables will consist of two components - a causal component and a non-causal (spurious) component. The causal component can consist of a direct component, an indirect component or both.

It follows then that any relationship between two variables can be interpreted as:

- a direct causal relationship
- an indirect causal relationship
- a spurious relationship
- any combination of these.

Figure 3.4 illustrates the possibilities where we have three variables which, for the purpose of the example, I will call X, Y and Z. The relationship between X and Y could be any of the following:
(a) Direct causal: Y follows X in time, Y is capable of being changed, and it is plausible that X could produce changes in Y. In the absence of finding any other variable that is responsible for this relationship we may continue to argue that the observed correlation is direct and is causal (Figure 3.4a).

(b) Indirect causal: Y follows X in time, Y is capable of being changed, and it is plausible that X could produce changes in Y. However, in this case we are spelling out the mechanism by which X affects Y. We may think of Z as a single mechanism or a whole lot of intervening variables (Figure 3.4b).

(c) Spurious: X and Y are not causally related to one another. Even though Y might follow X in time and be capable of being changed, both X and Y are joint effects of some third variable Z. X and Y covary purely because Z has a simultaneous effect on both X and Y (Figure 3.4c).

(d) Both direct and indirect: The effect of X may be partly via its effect on an intervening variable and partly direct (Figure 3.4d).

(e) Direct and indirect causal and spurious: The relationship between X and Y could consist of three components: a direct causal part (X → Y), an indirect causal part (X → Z → Y), and a spurious part (X → Z → Y) (Figure 3.4e).

(f) Direct causal relationship combined with a spurious component (Figure 3.4f).

(g) Indirect causal relationship combined with a spurious component (Figure 3.4g).

This set of possibilities results from situations in which we have only three variables. The more variables we take into account, the more complex matters become.

When we collect and analyse data it can be helpful to draw diagrams to spell out the ways in which we propose variables are interrelated. We need to specify:

1. whether relationships are presumed to be causal or spurious
2. whether causal relationships are expected to be direct or indirect
3. the mechanisms (intervening variables) underlying any indirect causal relationships.

Resolving these matters allows us to articulate our research question and the most plausible line of explanation.

Providing a frame of reference

The logic of making comparisons is fundamental to testing causal models. Consequently a central element in designing an explanatory
**WHAT IS RESEARCH DESIGN?**

**Proposition** Divorce leads to emotional problems in young children.

\[ \text{divorce} \rightarrow \text{emotional problems of children} \]

**Observation** We find that many children with divorced parents have emotional problems.

**Proposition** Private schools produce high academic performance among their students.

\[ \text{private schools} \rightarrow \text{high performance} \]

**Observation** There are many instances of students from private schools achieving highly.

**Proposition** Youth unemployment is responsible for youth suicide.

\[ \text{unemployment} \rightarrow \text{high suicide rate} \]

**Observation** 50% of young people who commit suicide are unemployed.

**Figure 3.5** Propositions and observations without a comparative frame of reference.

Research study is providing a comparative frame of reference. Different research designs go about the task of providing comparisons in different ways.

**Comparing groups**

By making comparisons we provide a frame of reference within which to try to make sense of particular findings. The importance of this can be seen in the illustrations in Figure 3.5. In these examples the observation in each case hardly provides convincing support of the proposition. One reason for this is that there is no frame of reference within which to make sense of the observations. There are no groups with which to compare and contrast the observations. It is only by making comparisons that our observations take on much meaning and we are able to eliminate alternative explanations.

The proposition that *divorce leads to emotional problems in young children* would encourage us to expect to find emotional problems among children whose parents are divorced. But finding such evidence will not get us very far. To go further down the path of explanation we must make comparisons. Are emotional problems greater than, the same as or less than those among children from intact families? Do the observations among children of divorced parents simply reflect that which we would find among any group of children?

**Proposition** Divorce leads to emotional problems in young children.

\[ \text{Parental marital status} \rightarrow \text{level of emotional problems} \]

**Restated proposition** Children whose parents have divorced are more likely to develop emotional problems than children whose parents have not divorced.

\[ \text{parental marital status} \rightarrow \text{level of emotional problems} \]

**Proposition** Private schools produce high academic performance among their students.

**Restated proposition** Private schools have higher levels of academic performance than government schools.

\[ \text{school type} \rightarrow \text{level of academic performance} \]

**Proposition** Youth unemployment is responsible for youth suicide.

\[ \text{unemployment} \rightarrow \text{high suicide rate} \]

**Restated proposition** Unemployed youth are more likely to commit suicide than employed youth.

\[ \text{employment status} \rightarrow \text{suicide rate} \]

**Figure 3.6** Using propositions with an explicit comparative frame of reference.

It will be recalled that when we say that two variables are related it means that *variation* or difference on one variable is linked to differences on the other variable. In this instance we have two variables: (1) parents’ marital status (married versus divorced) and (2) child’s emotional adjustment (low versus high). If our research design included only divorced parents we could say nothing about the impact of divorce on children. Even if 100% of the children from divorced parents had a low level of emotional adjustment we could draw no conclusions about the impact of parental divorce on children’s emotional adjustment. We would need to compare the adjustment of the children with divorced parents with the adjustment of children from intact families. The difference in adjustment levels in the two groups will provide strong evidence regarding the alleged effect of divorce on children’s emotional adjustment.

The propositions in Figure 3.5 only mention one particular group (children of divorced parents; private school students; suicidal youth). It can be helpful to restate the propositions to highlight the implicit comparisons. Figure 3.6 restates these propositions using explicit comparisons.

**Multiple comparison groups**

The independent variables above (parents’ marital status, school type, youth employment status) have all been treated as two-category variables (dichotomies). However, we are not restricted to comparisons of...
just two groups. Multiple comparison groups are possible and will arise from at least two circumstances.

**Independent variables with more than two categories** Where independent variables have more than two categories we can compare more groups and provide stronger and subtler tests of hypotheses. For example, if investigating the effect of divorce on children we could simply classify marriages into divorced or intact, or we could classify them as single (never married), intact, separated, widowed, divorced (re-partnered) and divorced (not re-partnered). If we took the latter approach we could make multiple comparisons. In doing so we can get a clearer picture of what is going on. If we find that it is the children of divorced parents who have re-partnered who have the most problems we achieve a more focused understanding of the impact of parental marital status on children. We have learned that it is not divorce per se but the arrangements that follow divorce that are critical. Had we simply compared divorced and intact marriages we might never have identified this.

**Combined effects of different independent variables** It is possible that it is only when people have a particular combination of characteristics that an effect is produced. For example, it may be the joint effects of gender and age rather than each independently that is linked to suicide. We may hypothesize that when a person is both male and young, the likelihood of suicide is at its highest. If, for the purpose of this example, we think of age as a dichotomous variable (15–35 = young; 35+ = old), then we can think in terms of four groups based on the combination of these two independent variables. These are illustrated in Figure 3.7.

**Comparing time points**

A finding that children from divorced families are less well adjusted emotionally than those from intact families presumes that divorce is responsible for this difference. The problem with a simple comparison between groups (divorced and intact) is that it does not tell us whether divorce has actually produced any change in the emotional adjustment of children. Any conclusions would be much more convincing if we could track the emotional adjustment of children both before and after divorce to see if there was any change in the emotional adjustment of children following divorce.

By tracking children over time we could see if there was a change in emotional adjustment of the children and evaluate whether this was attributable to their parents' divorce.

**Multiple pre-tests and post-tests**

In the above example the measurement of emotional adjustment at the beginning of the study is often referred to as the 'pre-test' and the re-measurement at the end is called the 'post-test'. Between the pre-test and the post-test certain critical events occur – in this case parental divorce (or non-divorce as the case may be). These events serve as the independent variables in the study – i.e. the presumed causal variable behind the observed change. In some designs these intervening events are called the 'intervention' or the treatment (see Chapter 4).

We can collect data about more than two time points rather than being restricted to simple 'before' and 'after' data collections. We might have repeated measurements over an extended period to track 'ups' and 'downs' and to track trends before and after any critical event. Multiple 'pre-tests' and multiple post-tests can help distinguish between short term and long term trends. They also help identify the effect of the 'intervention' or independent variable. For example, a simple measure of emotional adjustment before and after divorce might show a decline in adjustment. But what if there was strong evidence of the decline having commenced well before the divorce? Only multiple measures before divorce would show that a trend had already begun. Similarly, a simple post-test might show that adjustment is lower after divorce but multiple post-tests might show poor adjustment immediately after divorce but a steady improvement over time (see Chapter 6).

**Making meaningful comparisons**

Ideally the groups we are comparing should be the same in all relevant respects except in regard to the independent variable. For example, if we want to test the idea that non-government schools produce students that achieve better academically than government schools we would need to be confident that both types of schools contained comparable students. We would need this so that we could be confident that the only relevant difference between the two sets of students is in the type of school attended. If the students differed in additional ways, how would we
know which of the differences was responsible for any differences in academic achievement?

Comparisons of children from divorced and intact families are complicated by the fact that typically children of divorced parents are older than children from intact families. Any differences between the level of emotional adjustment of the two sets of children could be due to age differences rather than the marital status of their parents.

The same problem can arise when looking at the same group or category of people over time. Ideally with comparisons over time the only differences between the pre-test and the post-test should be the event we are proposing as the cause of any change. However, many events can occur which can account for the change and these may confuse our comparisons and our attribution of what lies behind the change in the outcome variable (see Chapter 4 for further discussion).

Since comparisons are central to good research design we must ensure that they are meaningful. The more we can remove unintended and unknown differences between groups, the more we reduce the risk of mistaking spurious relationships for causal relationships - that is, the more we eliminate alternative plausible explanations.

There are four main strategies for maximizing the comparability of groups.

**Matching**

When recruiting the groups we can deliberately match them on relevant characteristics. Thus when comparing students attending government schools on the one hand and fee paying private schools on the other hand we should ensure that the two sets of students are similar in terms of intelligence, aspirations, parental resources, education values, family history, gender and age. By comparing like with like we should be able to isolate the effect of the type of school on academic achievement.

The problem in comparing like with like is to establish the identity of all the variables on which we need to match the groups. We can match for the characteristics that we know might contaminate our results but there may be other factors that we have not thought of.

**Ex post facto matching** Ideally groups should be matched at the beginning of the study before critical events (e.g. divorce, attending the school). Another approach referred to by Spector (1981: 48) as a 'patchwork procedure' involves creating matched groups from a whole pool of study participants after all the data have been collected. For example, we might have a large number of students from government and fee paying private schools. From this pool we could extract a group of government school students and a group of fee paying private school students who are comparable in terms of intelligence, aspirations, parental resources, values about the importance of education, family history, gender, and age. We could then compare these groups and see if there were differences in their level of academic achievement.

There are many problems with this approach. The obvious one is that we must have the relevant information on which to match. Since this will always be limited, the groups are likely to remain unmatched on important but unknown factors. Another problem is that many cases from both groups simply will not have matches from the other group and will have to be discarded. If we match on more than a small number of factors we will often end up with only a very small number of people in the study since the number of cases from both groups that match can be quite small.

**Randomization**

A simpler and more effective way of making groups comparable is to randomly allocate people to different groups at the beginning of a study. By randomly assigning individuals to each group any differences between groups should be random rather than systematic. So long as groups are large enough, random assignment should automatically produce groups with comparable profiles on both known and unknown factors. From a statistical perspective, random assignment of people to groups will make the groups identical for all intents and purposes and provides what Davis calls 'the all purpose spuriousness insurance of randomization' (1985: 35). We control for an infinite number of plausible rival hypotheses without specifying what any of them are (Campbell, 1989).

This is the approach taken in drug trials. Individuals are randomly assigned to one of several groups and each group is then given a different drug. Since the groups should have virtually identical profiles to begin with any differences in outcomes between the groups should be due to the different treatments administered to each group. However, this approach is often not applicable in social research because practical and ethical considerations preclude us assigning people to groups and then doing something to one group to see what effect it has (see Chapter 5). Advantages and drawbacks of this approach to social research are discussed in Chapters 4 and 5.

**Matched block designs**

While randomization minimizes the chances of initial differences between groups there is still the chance that there will be differences between the groups - especially in smaller groups. Where we want to be certain that the groups are comparable on a particular characteristic we can take an additional step to guarantee that the groups are comparable at least on specific variables. This is achieved with what is called the randomized block design - a combination of both randomization and matching.

Suppose we wanted to evaluate the effect on academic performance of three methods of delivering university courses: (1) face-to-face delivery
of lectures and tutorials, (2) correspondence using written materials and
(3) electronically using the internet and e-mail. Further, suppose that we
could assign individuals to one of these three modes. Random allocation
would be the obvious way. But suppose we want to be absolutely certain
that the students in each of the three modes are of comparable academic
ability.

To achieve this we would do the following:

1. Obtain a measure of academic ability at the pre-test.
2. Rank all the students from highest to lowest ability.
3. Select the three students with the highest ability (because we have
   three learning modes). This group of three students is block 1.
4. Randomly allocate each of the three students to a group (one to
   face-to-face; one to written correspondence; and one to electronic
   learning).
5. Select the next three most able students (block 2) and repeat the
   random allocation to groups.
6. Repeat this until all students have been allocated to one of the three
   groups.

STATISTICAL CONTROLS

An alternative way of making groups comparable is to do so at the data
analysis stage after data have been collected. It involves multivariate
analysis that, in one respect, matches groups on specified variables (see
Chapter 12 for a discussion of the logic of some of these techniques).

Although the procedures of multivariate analysis can be very complex,
the essential logic is simple. Suppose we want to compare the emotional
adjustment of children from divorced and intact families but we believe
that any comparison would be confounded by the fact that, on average,
children whose parents divorce are older than those from intact families.
Any greater maladjustment among children whose parents have
divorced could occur because they are older, as older children display
greater maladjustment.

Multivariate analysis removes any effects that might be due to age
by comparing like with like – by selecting, say, preschoolers and comparing
those from divorced and intact families to see if, despite similarity of age,
the children show different levels of adjustment. The same comparisons
could be repeated among say 5–8 year olds, 9–12 year olds, 13–15 year
olds and so on.

The obvious shortcomings of this approach are similar to that of
matching. We can remove the influence of variables that we have
thought of and on which we have data but we cannot remove the effects
of unknown variables, or those for which we have no data. Since we can
never know what factors we have missed there is always the danger that
factors we have not thought of may be contributing to group differences.

Interventions and independent variables

Throughout this chapter I have used the terms ‘intervention’, ‘treatment’,
‘independent variable’ and ‘groups’ more or less interchangeably.
Research designs vary in terms of the type of independent variable
employed and in the number of independent variables built into the
study.

TYPES OF INDEPENDENT VARIABLES

When thinking in causal terms (X—Y) X is the independent variable.
When we conduct the research and do the analysis we compare the
outcomes (Y) for different groups. The groups are defined according to
which category of the independent variable they belong to. In some
studies the investigator introducing an active intervention defines the
independent variable (e.g. allocate different people to one of three modes
of course delivery – face-to-face, correspondence or electronic). In other
situations the independent variable is defined by a naturally occurring
‘intervention’ (e.g. a person retires or remains employed). In other
situations the independent variable is determined by the relatively fixed
attributes of participants (e.g. sex, education, race etc.) rather than by
interventions of any sort.

NUMBER OF INTERVENTIONS

Research designs can also differ in terms of the number of interventions
that are made. Where the independent variable involves either an active
or a passive intervention we can examine the effect of single versus
repeated interventions. We can see whether a single intervention has a
different impact than the cumulative impact of repeated interventions.
For example, we might be evaluating the impact of teachers giving
students negative feedback on their work. Initial criticism of the work
may boost a student’s effort and performance but repeated negative
feedback may lower performance as the student loses confidence. A
multiple intervention design helps refine our understanding of the way
in which interventions affect participants and enables us to rule out
alternative ways of interpreting results.

DIMENSIONS OF A RESEARCH DESIGN

The above discussion has identified six core elements of a research
design. The particular mix of the elements in any study will yield
different designs.

These six main elements in producing a research design are:
The number of groups in the design. Designs will vary from those with no comparisons (e.g., single case design or case study) to those with many different comparison groups.

The number of 'pre-test' measurement phases. Designs vary from those with no 'pre-test' (e.g., cross-sectional designs and some experimental designs) to those with a series of 'pre-tests' which establish pre-existing trends before an event.

The number of 'post-test' measurement phases. All designs require at least one 'post-test' - the measurement of an outcome variable. In some designs (e.g., cross-sectional) there will be one 'post-test', while other designs can have many post-tests to help distinguish between short and long-term outcomes.

The method of allocation of cases to groups. In multiple group designs, groups can be made comparable by allocating people to different groups by random allocation, 'pre-test' matching, post hoc matching or block matching or by using statistical controls in the analysis phase.

The nature of the intervention. Studies that rely on existing variation (cross-sectional designs and those with 'fixed' independent variables) have no interventions. Other designs rely on interventions between a pre-test and a post-test. These interventions may be either active or natural.

The number of interventions. Designs with an intervention can have either a single intervention or multiple interventions. Multiple interventions can be used to identify the effect of cumulative 'treatments'.

A range of research designs.

The combination of the possibilities created by these six elements of a research design yields a large number of possible designs. A useful summary discussion of many of these designs is provided by Spector (1981). To impose some order on this range of possibilities it is helpful to think in terms of four broad types of design. For each design type, decisions taken by the investigator will produce variations within the type. These four types provide the structure for the rest of the book. The four broad types of design are experimental, longitudinal, cross-sectional, and case study. I will examine each of these in turn.

Experimental design.

The classic version of the experimental design has the following elements:

1. One pre-intervention (pre-test) measure on the outcome variable.
2. Two groups: one group that is exposed to the intervention (the experimental group) and one group that is not exposed to the intervention (the control group).

Figure 3.8: Classic experimental design

<table>
<thead>
<tr>
<th>Method of allocation to groups</th>
<th>Pre-test</th>
<th>Intervention (X)</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random (experimental group)</td>
<td>Measure on outcome variable (Y)</td>
<td>'Treatment'</td>
<td>Measure on outcome variable (Y)</td>
</tr>
<tr>
<td>Random (control group)</td>
<td>Measure on outcome variable (Y)</td>
<td>No 'treatment'</td>
<td>Measure on outcome variable (Y)</td>
</tr>
</tbody>
</table>

3. Random allocation to the groups before the pre-test.
4. One intervention (test/treatment).
5. One post-intervention (post-test) measure on the outcome variable.

This design is illustrated in Figure 3.8.

The analysis of any effect of the intervention focuses on changes in the experimental group before and after the intervention and a comparison with the rate of change in the control group. If the change is greater in the experimental group than in the control group the researcher will attribute this to the impact of the independent variable (the treatment).

Longitudinal design.

The basic form of this design involves:

1. One group.
2. One pre-intervention measurement on the outcome variable.
3. One 'intervention' where everyone receives the 'treatment'.
4. One post-intervention measurement on the outcome variable.

In effect, this design is similar to the experimental design except that there is no control group and typically only one 'experimental' group. The design is illustrated in Figure 3.9.

The analysis in this design compares the pre-intervention measures with the post-intervention measures. Change in these scores can reflect the influence of X on Y. However, the absence of a randomized control...
group makes it difficult to know whether the intervention or some other factor produces any change.

Cross-sectional design
The basic elements of the cross-sectional design are as follows:

1. Instead of interventions the cross-sectional design relies on existing variations in the independent variable(s) in the sample.
2. At least one independent variable with at least two categories is present.
3. Data are collected at one point of time.
4. There is no random allocation to 'groups'.

This design mirrors the post-intervention phase of the classic experimental design but without any random allocation to 'groups' being made. The data for this design are collected at one point of time and are analysed by examining the extent to which variation in the outcome variable is linked with group differences. That is, to what extent do those in different categories of the independent variable differ in relation to the outcome variable? Causal relationships are established by utilizing statistical controls rather than by random allocation of people to groups.

This design is illustrated in Figure 3.10. In this case the 'intervention' is simply being in a different category of the independent variable.

Case studies
Case study designs rely less on comparing cases than on an exhaustive analysis of individual cases and then on comparing cases. A distinguishing characteristic of case studies is that contextual information is collected about a case so that we have a context within which to understand causal processes.

Case study designs might consist of a single case study (e.g. a community study, a study of an organization) or a series of case studies with perhaps each case testing a theory from a different angle. It is useful to think of a case study in a similar way to an experiment. We do not finally reject or accept a theory on the basis of a single experiment; we try to replicate an experiment and conduct it under a variety of conditions. Similarly, a case study project that entails a single case study is analogous to a single experiment. If similar results are found in repeated case studies, or predictable differences in results for particular cases in the study, then we develop greater confidence in the findings of the cases in the same way that we gain confidence in experimental results that are found in repeated experiments.

Summary
This chapter has examined ways of structuring research designs to help draw convincing causal inferences from the research. Since causes cannot be observed they must be inferred from observations. However, incorrect inferences can easily be made. The chapter has considered ways of structuring research to improve the quality of these inferences. A range of ways of interpreting correlations between variables was outlined and criteria required for inferring that a correlation reflects a causal connection were provided.

The chapter emphasized the importance of making meaningful comparisons between groups as a core element of drawing causal inferences.
One of the tasks of research design is to structure the research so that meaningful comparisons of outcomes between groups can be made.

Finally, the chapter outlined six core elements of research designs: number of groups, number of pre-tests, number of post-tests, nature of allocation to groups, type of interventions and number of interventions. Research designs vary in the way these elements are dealt with. Four main types of research design – experimental, longitudinal, cross-sectional and case study – were then briefly described. These broad categories of design provide the framework for the remainder of this book.

Notes

1 There can be exceptions to this where a suppressor variable may be operating to mask a causal relationship. See Rosenberg (1968: Chapter 4).
2 We shall see later that this problem of loss of cases can be overcome with certain forms of multivariate analysis.
3 A naturally occurring intervention is an event that takes place between pre-test and post-test without the investigator initiating any intervention. An example might be a study of families over time in which some families experience parental divorce. Divorce is the ‘natural intervention’.