Principles of Methodology

Research Design in Social Science

Perri 6 and Christine Bellamy

Los Angeles | London | New Delhi
Singapore | Washington DC

interests lie in public administration and public policy, especially information policy. She recently completed a major ESRC-funded project (with Perri 6 and Charles Raab) on how public agencies deal with tensions between client confidentiality and data sharing, and is currently engaged in an administrative history of identity management and privacy in British government.

Introduction

We have written this book because, in our own teaching, we found that most textbooks on methods or philosophy of social science say remarkably little about one of the most fundamental questions in research design: what will we be able to claim to know as a result of doing this research?

The book deals with the key issues in methodology – as opposed to methods – from basic through to advanced level. We have written it to address the needs of early career researchers, including graduate students taking masters' courses for which they are expected to carry out empirical research and doctoral researchers undertaking methods training in support of their PhDs. The early sections may also be useful for undergraduates preparing to undertake research-based dissertations. But we believe, too, that the important and original, and perhaps controversial, argument we develop will also engage experienced researchers, whether they work in universities or carry out or commission the growing volumes of social research now undertaken for businesses, government and social agencies.

What is 'methodology'?

This book is about methodology and research design. By 'methodology', we mean the understanding of how to proceed from the findings of empirical research to make inferences about the truth – or at least the adequacy – of theories. Its importance stems from the fundamental insight that findings about empirical facts are often most interesting when they enable us to make deeper judgements about what might be going on beneath those facts. The point of understanding 'methodology' is that it should enable us to design our research so that we can draw defensible conclusions about what might be *causing* the things we observe, including those causes stemming from ways in which people think about the world.

How to draw defensible conclusions is not the same problem as how to 'generalise' from our results. In its commonest form, generalisation is not a very deep problem. If we have looked at a set of cases – perhaps young people becoming homeless, or companies going bankrupt, or political parties recovering from

electoral defeat – and we have good reason to think that those cases are typical of other cases in the same 'population', then most of us would risk generalising to cases that we have not actually observed.

A much deeper problem is that of drawing conclusions, even about the cases we *have* observed, from patterns of facts about such things to 'explanations' or 'interpretations'. 'Methodology' is the study of this problem. It differs from the study of 'research methods', which deals with techniques for collecting data and for analysing them. Good methods are, of course, essential to doing good research. A defensible methodology is no use at all, if the data are collected poorly or coded sloppily, or the wrong statistical tests or qualitative analyses are carried out. But the converse is also true. A sound understanding of methodology will help us avoid the risk of collecting and analysing the data competently, but finding ourselves unable to draw any sound conclusions because the research design is flawed.

The study of methods is commonly divided by the type of data they use. Broadly, different methods are used for collecting, coding and analysing quantitative and qualitative data. By contrast, many problems of research design and methodology are much the same, irrespective of whether the research uses numbers or strips of text. We still need to ensure that we can draw interesting conclusions from whatever we discover. For this reason, this book does not have separate chapters on qualitative and quantitative approaches to research.

Methodological disagreement: the impossibility of neutrality

Methodology is a subject about which social scientists disagree, probably more than they disagree about method. This means that there are few non-contentious issues with which to introduce social research methodology. We must therefore issue a health warning. No book on methodology can be entirely neutral between methodological arguments. As practising researchers we inevitably hold our own positions, which have influenced how we have conducted our own research. We have also found from teaching this subject over many years that students benefit from taking part in vigorous academic debates.

For these reasons, this book takes sides in some important, current disputes. We have endeavoured, nevertheless, to cover the main lines of methodological debate from a range of perspectives that accurately reveal the rationale for competing positions. We intend that, in this way, the book will equip you to develop your own views. A reader who does not share our approach, or who holds a rival one, can, we hope, use this book with confidence that we shall not smuggle in unacknowledged assumptions or disguise the space for rival positions. We have tried throughout to set out clearly the available contrary arguments, even as we indicate what we consider to be their weaknesses.

A multi-disciplinary approach to the study of methodology

We set out to write a book that will be useful across the social sciences, including sociology, political science, anthropology, development studies, business and management studies, criminology, public policy, social policy, and science and technology studies. Indeed, we draw examples from each of these disciplines and fields. We have both principled and practical reasons for believing that this cross-disciplinary approach is appropriate.

The principled reason is that all these disciplines share common foundations in methodology and research design. This is not surprising, because they have all developed from a common source. As late as the mid-nineteenth century, these disciplines were hardly distinguished from one another, save as empirical fields within the broad remit of social science. The practical reason is that it is increasingly common for students to learn about method and methodology in multi-disciplinary groups. Indeed, in many countries, including the UK, this practice is actively encouraged by the public bodies which fund research training.

We welcome multi-disciplinary approaches to research training, because we know from our own experience how much students and early career researchers benefit from seeing how colleagues in other disciplines work through methodological issues. Just as important, it can be easier to understand a complex issue from an example outside our own field, because it is easier to recognise the problem's logical structure, without being distracted by the empirical facts of the particular illustration. Contours and contrasts became clearer and easier to read, when we look at something from a distance. So, for example, political scientists should find considerable value in looking at methodological problems through the lens of examples taken, say, from anthropology, public policy or business studies.

The scope of the book

This is a book on methodology, so it will not cover much that is usually found in texts on research methods, either qualitative or quantitative. We say nothing here, for example, about how to carry out standard statistical tests or how to conduct interviews or analyse transcripts. Rather, this book addresses the problem of what inferences you might expect to be able to draw from findings produced by such methods. So it deals with the problem of how to design research that is methodologically defensible and provides confidence that good use can be made of its findings.

Nor is this book principally about the philosophy of the social sciences. We devote Chapters 2, 3 and 4 to philosophical issues that are of central importance to methodology, but this book does not provide a comprehensive guide to the

thought of the key figures in the philosophy of the social sciences or to philosophical issues and problems.

We insist, too, that understanding research methods and the philosophy of the social sciences – important as they both are – is not enough to make a competent social researcher. Too many new researchers embark on their work with some background in method and philosophy, but too little in methodology. Worse still, they may not even realise that the methodology of research design deals with distinct issues from the study of method and philosophy.

The book is principally about what is often known as 'observational research'. We do not cover in any depth the methodological issues at stake in experimental work, although Chapter 4 does say some important things about experimental design in order to show what is distinctive about observational research.

The structure of the book

The book is divided into four parts covering, respectively:

- *Foundations* – what methodology is and does.
- *Designs* – the principal types of research design and the methodological challenges each type presents.
- *Achievements* – the kinds of products or arguments to which researchers can make inferences from the findings produced by analysis.
- *Combinations and trade-offs* – ways of bringing two or more research designs together to support each other, and how to think about the compromises that all types of research designs must make.

Part I raises two issues that recur throughout the book. We explain why the study of methodology is controversial in a manner and to a depth that, for example, the study of method is not. Second, all the good things prescribed by methodology do not go together well: we cannot have everything we would ideally like in the same research design. A research design may exhibit several different virtues, but we usually have to strike trade-offs between them. So a key part of any methodological defence of a piece of research is making the case that the trade-off struck between these virtues is a reasonable one.

The second part of the book deals with the concepts structuring the whole book. It describes the main types of research design by distinguishing between deductive and inductive design, within- and between-case analysis, and between variable-oriented, case-based and case-comparative research. The third part of the book is concerned with the main types of achievements, products or outcomes of social science research, or what research seeks to draw inferences *for*. These products are descriptions, explanations and interpretations.

In practice, however, some social science research combines these elements. In the final part, we examine the services that description, explanation and interpretation can provide for each other in a combined study. In the concluding chapter, we discuss some of the ways in which trade-offs can be struck between particular virtues appropriate to different research designs

Our experience of teaching this material to multi-disciplinary classes of masters and doctoral students suggests that the first nine chapters provide a basic or 'core' course, while the last nine chapters support an advanced-level course.

Methodology is a practical business, because understanding methodology allows us to deliver the goods promised by social science research. So, while this book deals with principles derived from the theory of methodology, it does so in the spirit of the old adage that 'there is nothing so useful as a good theory'. And so it is important that you should be able to see clearly how these principles may be applied to your own research. To this end, we provide two main kinds of pedagogical aids: examples and exercises. They appear in numbered boxes in each chapter.

Examples

The examples are taken from many disciplines, and are of two kinds. Some use influential pieces of published research as vehicles for showing how experienced researchers – including the authors of some classic studies in social science – have dealt with particular kinds of methodological challenges. We hope that you will want to read the original texts, because it is important for new researchers to become familiar with at least some of the principal items in the methodological canon. But we have attempted to provide a sufficiently clear explanation of each example for it to be useful to readers who do not have access to the original publications or the time to read them all. Others are fictional examples drafted to help you think about methodological problems that might well arise in your own research. We have taken both types of examples from a variety of social science disciplines, and have ensured that some of them deal with practical issues in business and policy research.

Running through the book are several fictional examples. Some are about investigating rough sleeping and household behaviour in recycling domestic waste. The purpose of creating linked examples is partly to show how different issues in methodology might play out in relation to the same research topic or even the same question. It also shows how different methodological strategies can illuminate different aspects of a problem and cumulate in providing an understanding of it. These series of examples speak to our overall purpose in this book, which is to help readers understand how different aspects of methodology contribute to good research design, sometimes, but not always, in the same study.

Exercises and further reading

We also include practical exercises in each chapter, to enable readers to test their understanding of key points. Some of these also run throughout the book. Most are appropriate for use in class discussion.

Suggestions for further reading at both basic and advanced level are offered at the end of each chapter. These are selective rather than comprehensive reading lists. We have indicated what we believe to be the most important and accessible articles or book chapters, whilst also encouraging readers to read some of the classic contributions to the methodology of social science. Because the list of references at the end of the book also includes all the recommended reading, it serves as a fairly comprehensive bibliography on methodology.

We are aware that even readers with a background in social sciences may not be familiar with some of the vocabulary used in this book. We have therefore provided a glossary of important technical terms.

PART I

Foundations

What methodology is and does

ONE

Inference and warrant in designing research

This chapter will:

- explain what is meant by methodology and how it differs from method;
- introduce the three main types of research question in social science, and how each is answered by drawing inferences from patterns found in data; and
- explain that methodology is always controversial, because all good things do not go together, and that trade-offs must be struck between the virtues of good research design.

How does methodology differ from the study of methods?

Since this book is about methodology, we should start with that term. But, first, we should point out that many standard textbooks use it loosely to refer to anything to do with research methods. So do not be surprised if you read books or articles that claim to be about methodology, but which deal with issues that are excluded from this book. Our definition, by contrast, is narrow and specific, and distinguishes clearly between method and methodology in social science.

In this book, we define *method* as the set of techniques recognised by most social scientists as being appropriate for the creation, collection, coding, organisation and analysis of data.

- *Data creation methods* are used to produce the raw material of research, namely well-structured data – or sets of information – that can be used to perform further investigations, of the kind described below. Data creation methods include ethnographic or participant observation, focus groups, individual interviews, questionnaire surveys and so on.

- *Data collection methods* are procedures for capturing what is important for answering the research question from the data that have been created. They may involve scanning text for particular themes, codes or content or undertaking counts or more advanced quantitative procedures. However, we can only count or code once we have decided how to identify what is important, as we show in Example 1.1.

EXAMPLE 1.1. STREET LIFE

It is claimed that the number of people sleeping rough on the streets in British cities fell fairly sharply in the four or five years after the British Labour government's initiative on rough sleeping in 1998. But it then levelled out, and at the time of writing appears to be increasing again.

However, as a consultation paper issued by the subsequent coalition government (Department of Communities and Local Government, 2010) shows, there is a major problem with this claim, in that no one believes that the data on rough sleeping are accurate, because counting the number of rough sleepers is far from straightforward.

The government is worried that its official definition of rough sleepers as 'people sleeping or bedded down in the open air' means that local councils do not count people who spend the night awake or sitting up in sleeping bags. But does it follow that councils should count all people on the street with sleeping bags? What, for example, about people who may use them as an aid to begging, but do not actually sleep rough? And should councils count people who sleep in tents, stairwells of blocks of flats or who take refuge on cold nights in shelters run by charities?

- *Data coding methods* are procedures for determining whether the information indicated by a particular datum or set of data meet the standards or thresholds required for them to be classified under a category, where that category is related to the research question or hypothesis.
- *Data organisation methods* are procedures for laying out whole sets or series of data, that have either been created, collected and coded by the researcher for the purposes of the project, or been taken from another source – for example, a national survey data set such as the British Crime Survey (BCS) or, as in Example 1.1, the British government's annual estimate of rough sleepers (available at http://www.communities.gov.uk/publications/corporate/statistics/roughsleepingcount2010). Data organisation involves setting out the data on a suitably common basis – for example, by tabulating them – so that they can be analysed.
- *Data analysis methods* are procedures for manipulating data so that the research question can be answered, usually by identifying important patterns. Statistical procedures are obvious examples. There are many qualitative analysis techniques too, such as open-ended content analysis, and a variety of theory-based comparative techniques for handling historical qualitative data of the kind we shall discuss in Chapter 17.

EXERCISE 1.1. RIGOROUS BUT ROUGH

Taking account of the challenges identified in Example 1.1 above, think about how you could develop a consistent and accurate count of rough sleepers. What criteria would you use? How would you justify them?

Interpretation

You will notice that nothing has yet been said about the 'interpretation' of data. The process of data collection almost always requires the researcher to 'interpret' the data, and that this is particularly so when – as in Example 1.1 – the things being studied do not fall neatly into convenient, unambiguous units. We shall consider some of these issues in more detail when we discuss the use and application of concepts in Chapter 9. Coding likewise involves interpretation, because the decision whether the data indicate that a case meets standards for a particular code is an interpretive act of scientific judgement.

'Interpretation' is also required in the process of determining whether the data analysis supports the general conclusions drawn from the research, to answer the research question. We call this support, its *warrant*. Warrant is a central issue in methodology, and therefore one that will be addressed throughout this book.

A third meaning of 'interpretation' in methodology is discussed briefly below and will be discussed again in detail in Chapters 15 and 16. This third meaning of interpretation is restricted to particular kinds of data and particular sorts of conclusions – namely, those which attribute beliefs, ideas, emotions or ways of classifying to people being studied.

But the point to emphasise here is that all methodological approaches rely to a large extent on 'data interpretation' and therefore 'interpretation' is not a separate stage or activity from the ones we list above. Although research proposals are often written with timetables describing 'data interpretation' as if it were the final stage of a project when conclusions are to be drawn about the theoretical or practical significance of the research, in fact interpretation is at the heart of the whole research process.

So what is 'methodology'?

The key lesson from this discussion is that *methodology* is not just – and is often *not very much at all* – a matter of method, in the sense of using appropriate techniques in the correct way. It is much more to do with how well we *argue* from the analyses of our data to draw and defend our conclusions. The *methodological* question posed by

our rough sleepers example is just what would allow us to claim that an increase in rough sleeping has occurred; that is, to make inference to a *description*. If we went on to claim that a rise in rough sleeping is being caused by the economic downturn, then this would be an inference to an *explanation*. Or perhaps it would be illuminating to explore what rough sleepers themselves would count as rough sleeping and why. This would require an inference to an *interpretation* in the third sense discussed above.

Because methodology is about arguments that show warrant for inferences, it makes no sense to break down the study of methodology according to the different stages involved in the research process, in the way that we have just done above for methods. Rather, we shall distinguish in this book between different approaches to methodology, and discuss the strategies appropriate to these approaches. We shall begin this discussion later in this chapter, when we discuss the differences between research designed to lead, respectively, to *description*, *explanation* and *interpretation*. But we stress throughout the book that each of these approaches raises the same basic methodological question – how and how far can you argue from the particular data to the particular conclusions, or, to put it another way, what argument, if any, do these data actually support?

Being able to draw sound conclusions depends on designing all stages in the project on sound methodological principles. Conversely, it is entirely possible to follow prescribed methods carefully, but still produce methodologically suspect research, if the conclusions drawn from it are not soundly based. These problems are inescapably theoretical ones, because the study of methodology involves theories about how and how far the research design enables us to draw sound inferences to conclusions that provide answers to our research questions, or that determine how far our hypotheses are supported or undermined.

And that is what this book is all about.

Inference and warrant

The core concepts in methodology are those of inference and warrant, and we should explain here why they are so important.

We are used to opinion pollsters drawing conclusions about the voting preferences of over forty million electors by sampling the opinions of around a thousand people. They do this by using widely accepted principles of statistical inference. This example illustrates the problem that we often need to draw conclusions about a large population from what we can find out about a smaller sample. A second problem is that we cannot always observe the things we are interested in directly, but are forced to work with proxies or indicators. For example, psychologists make inferences about the working of human or animal brains from observing very fine movements of eyes. Industrial sociologists make inferences about organisational morale from the way workers behave or describe their feelings. And anthropologists interpret how human beings make sense of their worlds from their stories or other cultural artefacts. In none of these examples can synapses firing in brains, 'morale' or 'sense-making' be directly observed.

Furthermore, researchers could not confidently make inferences without theories – however implicit or provisional – about the relationships between the things in which they are interested and those things which they can directly observe. For example, using cultural artefacts to interpret sense-making depends on a theory of culture.

We can therefore define *inference* as (1) the process of making claims about one set of phenomena that cannot be directly observed (2) on the basis of what we know about a set of things that we have observed where (3) the choice of research instruments depends on a theory of how those instruments work.

We can define *warrant* as the degree of confidence that we have in an inference's capability to deliver truths about the things we cannot observe directly. Warrant involves particular standards, which we shall discuss in more detail in subsequent chapters. We shall see, too, that some of these standards are more straightforwardly related to methods than others.

Observation

In the course of the book, we shall have occasion to use this slippery but absolutely unavoidable word in several ways. There are four different ways in which this word is used in social science methodology:

1 The value taken by a unit of data that is collected for, defined by and organised in a scheme of measurement. For example, the value ascribed to a variable entered into a cell on a spreadsheet or table is an observation on that variable. 'Observation' is used in this sense in the question, 'what do the observations show?'
2 A unit of data, such as a case in a sample or data set, as in the question, 'how many observations do you have?'
3 The systematic collection of data about behaviour or action, where the researcher cannot exercise experimental control over the regime of stimulus and constraint under which the research participants act, as in the term, 'observational research', which is the alternative to experimental research.
4 The activity of a researcher undertaking visual and/or audio inspection of participants' behaviour, as in 'a period of fieldwork observations'.

When we discuss some philosophical questions in Chapters 2, 3 and 4, we shall use the word in sense 1 a good deal. Chapter 5 considers observational research, in sense 3. In Chapter 6, when we discuss variable-oriented research, 'observation' will be used in sense 2. Although we bear sense 4 in mind throughout, it will come to the fore particularly in Chapters 15 and 16. You are warned to pause whenever you see the word, to make sure that you know what is meant. It will always be clear from the context which meaning is intended, but you can check either this page or the entry in the glossary if you need a reminder.

Some controversial claims about methodology

With those definitions in mind, it is time for us to make some big claims. Some are going to be controversial. You will find, as you read this book, that almost anything that is said in the field of methodology will attract disagreement. This is another big difference from the study of methods, because most people who study methods agree on what counts as, for example, transcribing an interview, or calculating a chi-squared test.

Here is our first big claim. *Making warranted inferences is the whole point and the only point of doing social research*, irrespective of what type of data and what style of research we use. The contribution to knowledge of any research consists in the inferences that can be made from it. Inferences are the principal products; they provide support for findings; and they are what make findings into findings rather than speculations, on the one hand, or raw data, on the other.

There are two reasons for making this claim. The first is a semantic one and the second rests upon a normative claim about what our ambitions ought to be and why social scientists get out of bed in the morning.

The semantic reason is that careful attention to inference, and what warrants it, is what distinguishes *research* from other kinds of investigation. Good journalistic reportage does not generally try to make inferences, beyond telling us what the reporter found. Interesting speculative or theoretical writing does not have to be so concerned with warrant: pure theorists and social commentators leave that to empirical researchers. Detective work by police officers, however, *is* concerned with inference and warrant. But it differs from most social research – although it does resemble some kinds of historical work – in that it is concerned only with warranted inference about the particular case under examination, whereas a good deal of social science research is interested in drawing inferences beyond the particular case to a wider population.

The bigger, normative reason is that warranted inference is worth doing, because it represents a strategy for making a contribution to knowledge that none of these other investigative activities can achieve, given their entirely proper purposes and limitations. We need to understand how social processes generally work, and this cannot be done adequately by nailing 'whodunnit' in a particular criminal case, or by news-hounds ferreting out facts, or even by reading the insights of literary giants.

Critics of inferential ambition

We acknowledged that our first big claim would be controversial, and we should therefore tell you who would object to it, and why. Those who would resist this claim tend typically to argue one of the following positions:

- Social research can be justified if it 'gives voice' to people – such as rough sleepers – whose perspectives on homelessness would not otherwise be available. For this purpose, it is claimed, warranted inference to general theories is not necessary and, indeed, may actually be harmful. What, rather, is needed is researcherly observation and analysis that is faithful to the views of the individuals studied.
- Social research cannot, because of its inherently subjective nature, achieve warrant for general inferences, and should be considered just as lacking in fundamental – or 'foundational' – warrant as journalism, speculative writing, *belles lettres* and detective work. On this view, the accounts, say, of rough sleepers, national government ministers, local charity managers and local mayors of why rough sleeping is a problem, and how big and significant a problem it is, are bound to be different: we cannot achieve a perfectly accurate description, let alone a true explanation, of this state of affairs. This view is shared by several schools of social thought, ranging from scepticism through relativism to anti-foundationalism and postmodernism.

An answer to the critics

We are not persuaded by either of these claims, and, for the record, we shall offer a couple of remarks to indicate why we disagree with the first of these views. The second, we shall leave for Chapters 2–4.

'Giving voice' involves attributing thoughts, emotions, practices, aspirations, memories and so on to other human beings. Researchers often want to reveal the preferences, experiences or ways of understanding the lives of the people they study. But none of these things can be observed directly, nor can they be read off unproblematically from what people say in interviews or do when they are observed. There is no getting away from using information from outside the particular situation, because even the concepts we use are taken from a wider vocabulary. And when we try to work out just what people think, we draw on information about other people we believe to be similar to those we are studying.

The very concept of 'rough sleeping', for example, is one drawn from government policy documents, and we would find it difficult to escape from concepts such as the 'vulnerability' or 'social exclusion' of rough sleepers if we tried to describe the impact on their lives of, say, the closure of a winter shelter due to spending cuts. People in charge of public policy, and researchers who write and read scholarly articles or monographs, use language in a very different way from many of those studied by social science research. And so, 'giving voice' often involves risky acts of translation or making risky attributions.

The only way to do it well – and to do it in ways that make us accountable to other academics or to participants in our research – is to adopt procedures that force us to be conscious of the inferences we make and to reveal all our workings-out. That is to say, one characteristic of good research design is that it enables us to demonstrate how we got from interviews and other observations to our conclusions about research participants' lives. This process *is* warranted inference.

An alternative way of proceeding is available, of course. We could write down what we happen to think and perhaps publish it in journalistic outlets. But our research

would then constitute a different kind of enquiry, undertaken for different purposes and with different kinds of accountabilities to the data, to research participants and to the wider academic community. Social research based on warranted inference makes a quite distinct – and distinctly valuable – contribution to understanding people from that made by journalism or any other type of enquiry. Specifically, the unique contribution of social science consists in the methodological care that we pay to the inferences we make.

Inference to what?

All this raises an important but obvious question: *to what* exactly do we make inferences? We have already seen that social scientists distinguish between three types of purposes for which inferences are made. These purposes are *description*, *explanation* and *interpretation*.

Descriptive inference

Descriptive inference is undertaken to answer certain questions about Xs (where X stands for any empirical topic for social research) when we cannot observe them at all, or cannot observe them all, or can observe only aspects of them, or cannot be sure that what we are observing of the Xs is quite what it seems. These questions are, 'what kinds of things are the Xs?', 'what kinds of statements can we make about them?' and 'how can we characterise them?' The product of descriptive inference is a set of claims about Xs. These claims may be about what is typical of Xs, what is generally true about Xs, or what is true about a subtype or across some spectrum of Xs.

One product of research on rough sleepers, for example, might be a description of how many rough sleepers there are in a particular town; what kind of people they are by age, gender and so on; how long, on average, they have been homeless; whether this period is becoming longer or shorter; and whether the number of long-term homeless people is rising or falling. This description would depend on inference, because – even if we could count directly everyone who sleeps rough in the town on a particular night – we would need to make assumptions about what proportion of rough sleepers we have observed. And we shall also have to make inferences from earlier data or from interviews with rough sleepers, about *changes* in patterns of rough sleeping and in the characteristics of the population of rough sleepers.

Some textbooks are very snooty about descriptive inference. This snootiness is – to coin a phrase – unwarranted. Description may be a modest ambition, but it is a necessary one. It is very difficult to go on to do anything more ambitious in social research if you have not got the descriptive inferences right. It is true that the most prestigious journals do not publish articles that offer *only* descriptive inferences. But the articles that they do publish rely, in a vital part of their overall argument, on the soundness of descriptive inferences, even if those parts of their workings are not shown.

Explanatory and counterfactual inference

Explanatory inferences are undertaken to answer the questions, 'why have the Xs done Z or become Y?', 'what brought this about?' and 'what *caused* the Xs to become Y or do Z?' In Chapters 10–13, we shall look in much more detail at what we understand by causation. We shall see that explaining how something came about raises methodological challenges of a higher order than describing it, although description can often be quite tricky too.

Suppose that we want to find out whether cuts in public spending have contributed, causally, to an increase in rough sleeping. Once the cuts have taken place, we can no longer look at rough sleeping in a particular town in the absence of those cuts. So we could never measure the impact of cuts on rough sleeping by comparing the situation we currently observe with one (in one and the same place) in which the cuts had never taken place. This difficulty is known as the fundamental problem of counterfactual causal inference. It is one reason why explanatory inference is tough. But it is often very important to try for explanations. Indeed, explaining why things happen is the main reason that anyone pays for social science research to be done, in the hope that explanation will help with the design of interventions in social problems.

There are weaker senses of the term 'explanation' which do not require *causes* to be revealed. For example, researchers write of *statistical* 'explanation'. This phrase refers to the process of showing that two variables are strongly associated with each other, but does not require us to draw any inferences about which direction any influence might run or to rule out the possibility that both variables are being influenced by a third variable. Other explanations are *logical* in character. That is, we may 'explain' a condition or event, by showing how it is derived logically from another. For example, we may explain the government's plans for extended sharing of personal information about individual citizens between government agencies on the grounds that this is a direct – that is, a logical – implication of an emphasis on multi-agency interventions in social problems such as homelessness.

Interpretive inference

Finally, there are *interpretive inferences*. Interpretive inference is addressed to a variety of questions, some of which we have already discussed.

We have seen that the most elementary interpretive inference is made when we determine whether something is to count, for a given research purpose, as falling within some category, and therefore decide that it is to be given a particular code or measure. We call this an interpretation of its *categorical significance*. Deciding, for example, who counts as a rough sleeper – a question which precedes the descriptive inference question of whether we can draw conclusions about the numbers of rough sleepers – is clearly a matter of interpreting the concept of a rough sleeper. And that, in turn, depends on our view of whether that concept captures the particular aspects of the underlying condition of utter homelessness in which we are interested.

Second, giving voice is only one way of accounting for how people think, feel, understand, frame issues and so on. Interpretive inference is not simply the development of descriptions of people's subjective experiences, but may also produce an *integrated* account – or interpretation – of the *subjective significance* for people's mental lives, in which the patterns observed make some larger sense. For example, in interpreting how managers of local council housing departments perceive the implications of the government's edict to count rough sleepers in new ways, we would probably need to go beyond a simple repetition of our descriptive data (e.g. 37% of respondents agreed with Proposal 1 put to them in our survey) by drawing inferences about the significance they attach to the government's proposals for the lives of their rough sleepers and for their capacity to help them. If the data allow, we could also, perhaps, make further inferences about what these managers believe to be the significance of these proposals more broadly for social justice or social inclusion and about the standards they appear implicitly to adopt in measuring justice and inclusion. Finally, there are inferences to integrated accounts in which the subject of the interpretation is not the mental life of a group – or groups – of people, but a set of events. Historical interpretation – which is very important in historical sociology, comparative political science, business history and even in institutional economics – is a case in point. Its aim is to detect overarching patterns of historical events – for example, those involved in the emergence of multinational business corporations in the period after the First World War or in the growth of a welfare state after 1945 – to provide the basis for an integrated account or interpretation of their *objective significance*.

Relationships between descriptive, explanatory and interpretive inferences

Much of the book will be devoted to considering separately the standards of warrant required for inferences in *explanatory* and in *interpretive* research. But we shall see that almost all explanatory and interpretive research rests upon descriptive and categorical interpretive inferences.

We shall see, too, that even those researchers who insist most fiercely on an exclusive focus on interpretation of *subjective* significance cannot, in practice, carry out that task without implying some kind of explanation of why people think as they do. It is very difficult to develop an account of, say, the ways in which people on low incomes think about or 'frame' the risks of health problems arising from their diets, without making some reference to categories that imply causation. For example, in trying to decide between interpretations that emphasise the limited dietary choices available to people on low incomes and those interpretations which emphasise their limited willingness seriously to consider eating health foods, researchers necessarily find themselves implying something about the causal role that beliefs might play in explaining unhealthy dietary behaviour.

This example illustrates the point that separating descriptive from causally explanatory categories is not straightforward, because we often describe by using categories that imply an explanation. For example, we might count the number of 'drug-dependent' people who are registered for treatment in the UK, but the very use of this category recognises addictive dependency as a significant *cause* of the use of illegal drugs and carries the implied claim that it should be treated rather than punished. We shall see in Chapters 15 and 16 that there are other and deeper reasons why it is difficult to do interpretive research without carrying any explanatory baggage.

The questions addressed by descriptive, explanatory and interpretive research are, nevertheless, analytically quite distinct. These three types of research ask, respectively, 'what's going on with the Xs?', 'why have the Xs done Y?', 'what do the Xs understand by the way they do Y?' and 'what is the wider significance of the fact that the Xs have done Y?'. It is therefore most helpful to consider separately the methodological challenges raised by each of these three approaches, and this is what we shall do in this book.

Trade-offs between virtues in warranting inference

In examining these challenges, we shall explore the virtues that should be exhibited by methodologically sound research, if it is to warrant the inferences that it seeks to support. Indeed, we have already noticed some of these virtues.

First, in discussing description, we have implied that a key virtue of a description is that it should be as *accurate* as possible within the limitations imposed by the ways in which the data have been created and collected. For example, the kind of accuracy we expect from a statistical description of broad trends is very different from the kind that can be achieved by a meticulous anthropologist who carefully checks each significant observation recorded in his or her field notes.

Second, we have assumed that our inferences should capture the significance of as many of the data in the set as is practicable. In other words, the account should summarise and integrate our findings, but with the minimum loss of the facts, nuances, differences and contrasts that are relevant to the question. The better our account does this, the better its *goodness of fit*.

Third, in contrasting social research with detective work and investigative journalism, we pointed out that social science researchers want to draw inferences beyond the particular case to some wider population of people, events or cases. That is, we are interested in achieving *generality* across some category.

Fourth, we mentioned that researchers often look for a few overarching patterns that are of the greatest significance in shaping thought styles or emotions or in explaining outcomes or events. Whilst it might be tempting to trace in great detail the interaction of a large number of complex factors that might *explain*, say, the rise in custodial sentences handed down by the criminal courts, it is both impractical and

distracting to continue piling up lots of different factors over a large number of cases. It may be better to compare the influence of a few, important factors such as changes in national sentencing guidelines, judges' attitudes in interpreting them, and beliefs held by judges and juries about how community sentences work. That is, another virtue of both explanations and of interpretations is *parsimony*.

There are other virtues, which we shall consider in due course. However, we shall also see that it is often impossible in the same research design to maximise accuracy, goodness of fit, generality and parsimony, let alone other virtues (Przeworski and Teune, 1970). For example, the more accurate we try to be, the more detail we accumulate and the closer we stick to the granularity of particular cases, the more difficult it becomes to generalise across cases. It also becomes more difficult to identify the effects of a few really central factors, because they will not consistently perform their explanatory or their interpretive work at the level of close detail. Conversely, the more parsimonious we want to be, the more likely it is that we shall be forced to restrict the domain of cases over which we can generalise, because rather few things are common to every case, especially those falling in wide categories like 'homelessness' or 'judicial behaviour'. This problem means that we have to strike trade-offs between virtues in designing our research.

The need to strike trade-offs between virtues of good research design is one reason why there is no such thing as a piece of research that is completely beyond methodological impugning. It is possible to complain about something in every piece of social research, and social scientists, being a quarrelsome lot, are not slow to find it. But that does not mean that anything goes in striking trade-offs. There are always better and worse trade-offs to be found to address a particular research question, and there are some that lie so far behind the trade-off curve, or so far to one extreme of that curve, that they would clearly constitute poor research design.

What is 'research design'?

But what, actually, is 'research design'? By the *design* of a research project, social scientists usually mean (1) the specification of the way in which data will be created, collected, constructed, coded, analysed and interpreted (2) to enable the researcher to draw warranted descriptive, explanatory or interpretive inferences (3) where the warrant is calculated to strike a reasonable trade-off between competing virtues; and (4) where the standards of warrant may vary slightly, but are based on a core set of virtues for each type of inference.

A research design is usually set out in advance of undertaking a project, in a research plan or *proposal*. A more detailed statement of the methodological defence of a research proposal is often provided in a *protocol*, which lays out in detail the steps through which the inference will proceed and the degree to which the conclusion can be supported, given the nature of the data and the nature of the methods used to create, collect, code, construct and analyse them.

Standards of good research design

The simplest standards of soundness in methodology are those of *reliability* and *validity*.

Reliability

Reliability has to do with how we measure – or, if we are using qualitative data, code – the things in which we are interested. A reliable system of measurement or coding is *consistent* in that, each time it is used on the same data, it yields the same measure or code. If two researchers work together, and both follow the same procedure on the same data, they should produce the same measures or codes. Redoing the coding or measurement, to see how reliable the procedure is, is called the 'test/retest' method of assessing reliability.

A second way to assess reliability at the level of method is called the 'internal consistency' method. This does not rely on repeating the coding or measurement of the same data, but on gathering additional data using the same design. In a questionnaire survey, for example, we might insert several questions, each phrased slightly differently, to ask the same thing. If they elicit the same answers from the respondents as did the first, then they provide some evidence that the first question was reliable.

Validity

Validity is, loosely, the degree to which our statements approximate to truth. It is conventional to distinguish between construct and conclusion validity, and between internal and external validity.

Construct validity is the degree to which the measures or codes used to operationalise a concept really capture what we intend to capture. For example, suppose we want to know how much 'goodwill' people have toward their neighbours in their own street. Goodwill is not a straightforward concept. We might ask about people's attitudes to other people in general and to their neighbours in particular. Perhaps we should ask about hypothetical future neighbours who might differ in important ways – for example, in their ethnic origins – from the present ones. But we should surely want to know, too, about how people actually behave toward different neighbours. Perhaps we would want to know about how they think they would behave in certain hypothetical situations, such as a severe fall of snow in the neighbourhood. We might also want to know about how they expect their neighbours to behave toward them.

Having settled on a set of measures or codes, we could assess their construct validity in several ways. The simplest way might be to look at theories of goodwill, and compare our measures or codes with the features used in those theories. If we are collecting quantitative data, we might use statistical analysis to determine whether there are common factors that run through each of our chosen

measures – such as whether goodwill is based, say, on 'social affinity' (sharing ideals and beliefs) or 'social reciprocity' (helping each other out), or whether in our observed cases, the values point in different directions: for example, goodwill may be strong where it depends on affinity but less so where it depends on reciprocity. If so, we might wonder whether, in fact, 'goodwill' is a single phenomenon after all, and instead stipulate different 'types' of goodwill. This process would increase the construct validity of our concept of goodwill, by giving it more operational precision.

Measurement validity is a subtype of construct validity. It captures the extent to which any given measure or code allows us to attribute values, say to different factors in, or dimensions of, 'goodwill' without importing systematic bias. Measurement validity is important whether we use cardinal (1, 2, 3...) or ordinal (1st, 2nd, 3rd...) numbers, on–off codes (yes/no) or qualitative values (such as 'strong/moderate/weak').

Conclusion validity concerns the warrant we have for making inferences from our conclusions. It relates to the degree of support which the patterns observed in the data provide for the conclusions drawn from them. If we conclude, for example, that goodwill based in social affinity tends to be stronger than that based in social reciprocity, the question is whether this conclusion is a reasonable statement of what the data show.

Internal validity applies within a study, regardless of whether we want to generalise to others. It concerns the warrant we have for inferring that an outcome can be explained by a particular causal factor. If we claim, for example, that our study shows that 'social reciprocity' becomes stronger the longer people are neighbours, regardless of factors such as race, then the test of the study's internal validity is the extent to which we can show from our data that this really is the case.

External validity concerns the warrant we have for inferring that our findings would hold in other situations or studies that were similar in relevant ways. Clearly, there is a gradient of similarity and dissimilarity. As samples or cases become less similar, external validity is bound to fall, along with our ability to generalise from the study. So, for example, the findings of a study of neighbourly goodwill in an American small town might be expected to hold in towns of a similar size and with a similar socio-demographic structure, but not in a city or in a small town with very different population. This means that a key issue in securing external validity is knowing what features of our cases or population are 'relevant' for this purpose, and what makes them 'similar' or dissimilar.

Trade-offs between validity and reliability

Just as there are trade-offs between different virtues in research design, so there can also be trade-offs between validity and reliability. At first blush, this might seem an odd claim. After all, as a measure or code declines in reliability, so it must also become less valid.

But there are some things we may want to measure or code in social science that are not amenable to straightforward measurement or coding. Suppose, for example, we want to understand the differences between people in respect of their capacity to make discriminating and thoughtful judgements in the fields of the arts such as music, theatre, literature and dance. Measuring taste, or aesthetic judgement, requires a cluster of different dimensions, because it is not just one thing. We should need to measure or code the breadth of the arts over which someone was capable of exercising judgement; whether they did so in consistent ways; and also the different ways in which they might be more and less articulate in their judgements; and so on.

Bringing all these measures or codes together into one composite indicator of taste could be done in a variety of different ways. We could, for example, increase the validity of our composite measure or code of taste by adding more subsidiary measures, such as scope, consistency and articulacy. That process would pick up more dimensions of this complex concept, but it would increase the difficulty of choosing a way of combining them, and our composite measure would be sensitive to whatever method we chose to weight and relate measures of particular dimensions of taste.

In other words, we would risk reliability for gains in validity. Beyond a certain point, too great a sacrifice of reliability will also ruin validity, and the range of *acceptable* trade-offs between the two values – for example, between reliable precision and valid relevance – especially in measuring complex or rich qualitative concepts such as taste is probably quite narrow. But there is usually more than one defensible trade-off to be struck dealing with this problem.

Sometimes, though, the trade-off problem can become vicious. Problems arise most acutely where the very process of doing the measurement or coding changes the thing being measured. For example, doing research about behaviours which are unlawful or which are regarded as immoral can cause the people being studied to behave more cautiously because they are being watched, or, alternatively, to show bravado by exaggerating their sinful behaviour. This is a problem that is well recognised, for example, among researchers who want to conduct ethnographic studies of institutional racism or bullying, where attempts directly to observe behaviour by means of non-participant observation end up by seriously undermining both validity and reliability.

This problem is also familiar to policy-makers. Goodhart's law was originally developed in the 1970s and 1980s, when new, more complex measures were developed by central banks of what counted as 'money'. The reason for measuring money in different ways was that central banks began to be charged with gaining control over the money supply, and needed to know how well they were doing. Unfortunately, the introduction of measurement and the use of policy measures to influence the money supply interacted in unexpected ways. Quite simply, when measures focused on one definition, people created money on some other definition instead: the central banks' work began to seem like squeezing a balloon in one part, only to expand the bulge at another.

The former Bank of England economist Charles Goodhart concluded that the very effort to measure money was making those measures less valid. He generalised his

finding to any situation in which measurement was associated with policy action and so had behavioural consequences. Goodhart's original formulation of his law concerned the application of policy action – 'any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes'. Later formulations have extended it to make the point that even introducing or publishing a measure will have behavioural consequences that reduce its validity for capturing the phenomenon of interest. The problem of Goodhart's law matters most in research conducted over a period of time, when the people being studied have time to react to the research. So it particularly affects longitudinal research or where the activity being studied is one about which people have normative views.

We shall return to these concepts in Chapter 6 to explore how they are applied to research designs that use variables. In the more advanced chapters about explanation and interpretation, we shall look at various ways in which internal validity can be pursued in observational research, not least because many method textbooks give only experimental examples of internal validity. Construct validity is of central concern in Chapter 9 on concept formation and is at the heart of the methodological challenges for good interpretive work that we discuss in Chapters 15 and 16.

TWO

Methodology and social science knowledge

This chapter will:

- examine what it is that social science seeks to describe, interpret or explain, by distinguishing events, trends, states of affairs and conditions;
- discuss whether social science methodology for describing, explaining or interpreting is fundamentally different from that used in natural sciences;
- distinguish three levels of intellectual claim that social science examines, namely paradigms or frameworks, theories and models; and
- examine the relationships between them.

Why do practical social researchers need to know some philosophy?

In the next three chapters we enter the terrain within the philosophy of social science which is specifically concerned with the status of the knowledge we claim to be able to find by doing social research. We shall assume no prior knowledge of philosophy in general or of philosophy of science in particular. But to make sense of the rest of the book, it is important that you first acquire some understanding of the principal terms used in the philosophy of social science, along with some familiarity with the contours of philosophical debates that are relevant to methodology.

Everyday practice of research in the social sciences continues to have a close involvement with philosophical debate. Rival traditions of substantive theory often claim to rely on rival philosophical arguments, and advocates of particular methods – such as the statistical analysis of questionnaire-based attitudinal survey work or ethnographic participant observation – often justify their claims on the basis of philosophical arguments. Unsurprisingly, therefore, some of the mistakes and muddles we get ourselves into are as much philosophical as they are to do with the application of particular methods.

In this chapter, then, we shall first examine the kinds of phenomena about which we seek knowledge when we carry out social science research. We shall then consider the principal kinds of knowledge yielded by social science in the forms of *models*, *theories* and *paradigms* (or *frameworks*), before going on to explore in the next chapter what is meant when we claim that social science research can *empirically test*, *confirm* or *falsify* them on the basis of our data. Social scientists often aspire to decide between rival theoretical statements. We shall therefore examine what is involved in comparing, judging and, when necessary, replacing theories, and how far this process can be said to occur in a once-and-for-all manner in social science.

In Chapter 4, we shall revisit the same arguments, this time using the lenses provided by the main philosophical positions to which social scientists appeal when they seek to justify their methodologies, and when they criticise those of other social scientists. We shall see that each of these stances holds particular views about each of the questions examined in the next two chapters. This discussion will help us to draw out the methodological implications of these questions.

What do we try to describe, explain and interpret?

Consider the empirical claims in Example 2.1. Each has been taken from highly reputable research in the social sciences.

EXAMPLE 2.1. SOME KNOWLEDGE CLAIMS MADE AS A RESULT OF SOCIAL SCIENCE RESEARCH

- In April and early May 1962, as [Soviet premier] Khrushchev was finalising his decision to send missiles to Cuba, he was confronting an enormous problem with Berlin. (Allison and Zelikow, 1999: 106)
- Child poverty ... has fallen by 600,000 to the nearest hundred thousand (or just under one-fifth) in the ten years since 1998–99 and needs to fall by a further 1.1 million in the remaining two years until 2010–11 to meet this element of the [government's] target. (Brewer *et al.*, 2001: 8–9)
- A loss of $x is more aversive [to people in general] than a gain of $x is attractive. (Kahneman and Tversky, 2000: 3)
- Azande speak of dreams as oracles, for they reveal hidden things (soroka). (Evans-Pritchard, 1976 [1937]: 174) [The Azande are a people living in an area today occupied by the states of Sudan, the Democratic Republic of Congo and the Central African Republic.]

The first statement in Example 2.1 describes an *event* – the Soviet leader's fateful decision that led to the Cuban Missile Crisis of 1962.

The second statement quantifies child poverty in the last ten years or so, and describes a once-and-for-all change that occurred over a defined period, or a *trend*.

Events, temporary states of affairs and trends have defined dates and places to bound them.

The third statement describes something that, if it is true, needs no bounding by time and place. The psychological generalisation of loss aversion is a *regularity*, a generalisation that rests on a very large number of observations, each of which shows the same consistent behaviour.

The fourth statement describes a *condition*, or a state of affairs which presumably has some boundaries in space or time but where those boundaries are not clearly defined. Moreover, a *condition* may or may not be tied to any regularity of *particular* behaviours or actions.

Social science research makes claims about each type of phenomena. To be sure, there are people who argue that events and *temporary* states of affairs might be proper matters for historians or political biographers, but not for social science research: so they could be properly consigned to the humanities. They would argue that social science should, like every other science, be interested in seeking *general* explanations, and should only be interested in particular cases – such as Khrushchev's decisions about Cuba and Berlin in 1962 – in so far as they are examples of more general phenomena, such as political decision-making.

This argument may be overstated. Individual cases may be of considerable interest for a number of reasons. They may offer an extreme example of something of great interest – such as political decision-making in an extreme crisis, such as the Cuban Missile Crisis – or because they are unique and thus provide us with important understanding of the limits to generalisation.

EXERCISE 2.1. IDENTIFYING EVENTS, TEMPORARY STATES OF AFFAIRS, TRENDS, REGULARITIES AND CONDITIONS

Think about a research project that you intend, or hope, to do.

Is the focus of this research an event, a temporary state of affairs, a trend, a regularity or a condition? And do you intend/hope to describe, explain or interpret it?

What do we get from social science research?

Each of the four statements in Example 2.1 sets out claims of fact. Moreover, each has been made on the basis of inference from data. The first statement is based on documents, such as state papers, memoirs, diaries and newspaper reportage. The second statement rests on published quantitative data about family income. The third statement rests heavily on results from laboratory experiments, rather than on data captured from people who are not subject to control by researchers. And the

fourth statement rests on data captured by fieldwork conducted over several years in the 1920s and 1930s, during which the researcher lived among the Azande, taking detailed notes and asking questions.

The result, in each case, is a *descriptive inference*. The data on which they are based are taken from very different sources, but what they have in common is that statements of putative facts are made on the basis of more or less *structured* evidence.

Description is useful, but social scientists aspire to more than descriptive inference, in that we also want either to *explain* or to *interpret* our findings.

Social scientists present the products of explanation and interpretation as *theories*, *frameworks* and *models*. Theories, frameworks and models work by *reduction*. That is, they help us to see more clearly what lies behind the blooming, buzzing profusion and variety of empirical events, states of affairs, trends, regularities and conditions, by identifying a small number of general, abstract phenomena which characterise them or may, even, cause them. So they help us to see the shape of the theoretical wood in the empirical trees. Indeed, it is often held that the more convincingly we can show that a small number of more, rather than less, abstract phenomena account for a wide a range of empirical phenomena – that is, the more parsimonious our account – the better. Example 2.2 shows how a parsimonious theory helps us to understand and deal with crime.

EXAMPLE 2.2. HOW THEORIES ACHIEVE REDUCTION OF VARIETY IN EMPIRICAL DATA

In the field of crime prevention, a very simple but very influential theory – known as the theory of situational crime prevention (e.g. Clarke, 1980) – proposes that criminal acts which appear, empirically, to be very different can all be explained by the presence of an opportunity and an opportunistic individual. This is because it holds that crime is the outcome of how a *personal disposition to crime* interacts with *situational opportunities* (or temptation) through the mediation of a *criminal's decision-making* process.

The theory thus *reduces* the empirical diversity of crimes such as robberies, assaults, fiddling expenses and intellectual piracy to these three general, abstract factors. It proposes that crime can be prevented – or at least its volume reduced – by reducing opportunities for crime, and that this can be done by changing key aspects of the situations in which it is committed.

Such measures include redesigning cars to make them harder to steal and introducing visible CCTV cameras in car parks and elsewhere.

Social and natural sciences

Is there anything philosophically distinctive about the methodologies used for description, interpretation or explanation, or about the accounts offered for events, trends, states of affairs and conditions in social science, from those offered by natural and physical science?

This may seem to be a question of no great practical importance for most social scientists who expect to conduct most of their research work within the confines of social science. However, the trend is toward ever-deeper collaboration between social scientists and natural sciences, so this might well become a practical issue for many readers. In development studies, for example, it is now commonplace for anthropologists and sociologists to work alongside biotechnologists and agronomists to examine the ways that innovations in crops, weedkillers and pesticides are actually used in rural communities and how outcomes in such settings may be different from those observed in field trials in research centres. Vast sums are also spent on collaborations between social scientists and information technologists to understand the ways in which people's use of computing and telecommunications tools shapes, and is shaped by, new technologies, producing outcomes that could not have been predicted from either social science studies of behaviour in offices or computer science models of technological change.

In collaborations such as these, reconciling methodologies for drawing inferences from data becomes central to the success of research. So the claim that there is something fundamentally different in social and natural sciences is a practically important one. Four justifications for this claim are usually given special importance.

The first can be disposed of quickly. This is the claim that natural and technological science is fundamentally based on the experimental method, in which the researcher can control the independent variables, whereas social science is basically observational, and thus the researcher cannot achieve such control.

Quite simply, both parts of the proposition are false. Much of natural science is conducted through mathematical modelling before any experiments are considered, and sometimes several years – or even centuries – before the necessary technology is available to conduct them. In natural sciences such as astronomy, for example, direct experimentation is far beyond human powers: experiments can be done only with physical models in which the relationship between the phenomena the astronomers want to understand depends on theory. Geology works with time periods so vast that the possibility of experimentation is very limited. Oceanographers have many theories that work at the scale of whole oceans or whole global systems and that simply do not admit of direct experiment. In many specialist fields of zoology, such as the study of primates and cetaceans, observational studies probably comprise the majority of what is published for both ethical and practical reasons.

Conversely, there are experiments in the social sciences, and not just the laboratory experiments so beloved of psychologists. As we shall see in Chapter 5, a rich vein of work in political science is concerned with conducting field experiments to find out, for example, whether certain kinds of intervention have any effect on people's propensity to vote. Similarly, in social policy, many more randomised controlled trials are being funded today to examine the potential impact of particular approaches, for example to the treatment of addiction and to crime prevention.

The second claim is that causation in human systems is fundamentally different from that in any natural or animal system, requiring fundamentally different forms

of modelling and therefore of research design. For many years, this was taken to be obviously true, but in fact it is far from clear that the premise is true or that, even if it is, the conclusion follows from it.

First, many animals clearly exhibit complex social organisation, perform rituals, engage in communication, collaborate, engage in competition and organised conflict, form kinship structures, exchange goods and services, use tools and so on. Human organisation may be more extensive, but it does not follow that the basic causal processes are fundamentally different. We used to believe that human systems are distinctive in being non-linear, complex in the sense of having sustaining self-organisation, and capable of irreversible causal processes and of exhibiting both positive and negative feedback dynamics. In fact, however, the social sciences have had to borrow every one of their models for these causal processes from natural sciences such as meteorology, oceanography and biology, and our use of such tools remains primitive by comparison with theirs.

A third claim is that natural science theories are more parsimonious than social science ones. It is true that the famous basic laws of physics, chemistry, microbiology and astronomy are indeed rather parsimonious. However, even in the natural sciences, the aspiration for reduction is sometimes qualified. For example, botany is committed to retaining quite a lot of blooming, buzzing profusion in its theories; so too are meteorology, geology, oceanography and many technological sciences such as agronomy.

Social science as the study of 'meanings'

The fourth claim is the only genuinely interesting one. This is the thesis that social scientists study '*meanings*'. This capacious term encompasses the full range of mental life including ideas, beliefs, desires, systems of classification, emotions, judgement and styles of thought. Some methodologists argue that studying meanings, in this sense, rather than just studying behaviour, makes a fundamental difference not only to the kinds of data with which we work, but also to our methods of analysis. It also makes, they claim, a fundamental difference to the methodologies we use for drawing inferences from data, and to what counts as an explanation.

The premise of this argument is true and no longer very controversial. Although in the 1950s and 1960s, behaviouralism in political science aimed to explain political behaviour and power without reference to meanings, it is not now judged to have been a success. Today, everyone is interested in meanings, from rational choice theorists who try to specify people's goals to describe their utility functions, through to ethnographers.

The second stage in the argument, likewise, is true. Data about meanings are different in kind from data about physical behaviour, whether they take the form of quantitative survey data about social attitudes or the less structured form of interview transcripts or observational field notes.

What remains deeply suspect, however, is the conclusion that methodologies used for data about meanings are, or ought to be, fundamentally different in character from those used for data about anything else. If we are engaged in causal explanation, we need to show how, for example, reasons or ideas operate to bring about action. And there can be true or false statements about what relationship a reason or an idea has to an action.

Think, for example, about Khrushchev's decision to deploy nuclear missiles in Cuba in 1962, which was mentioned in Example 2.1. We shall see later in this chapter that there are several theories about what was in his mind. But if one set of reasons was more important than the others, then that would change very fundamentally the way we explain the Cuban Missile Crisis. Beliefs and intentions have to be inferred, and form parts of coherent causal accounts exhibiting certain virtues before they can explain. Inference and building causal accounts are the core explanatory activities of all the sciences.

Later in the book, we shall discuss in more detail the argument that a science of meanings is methodologically distinct and that, furthermore, it must somehow eschew explanation altogether and deal with interpretation instead. In Chapters 15 and 16, we shall argue that one reason this argument fails is that it is very difficult indeed to produce an interpretation that does not imply some claims, however well disguised, about causal explanation. For the moment, we need only say that the argument is greatly overstated.

This conclusion is good news for those who engage in collaboration with natural scientists. But it does not follow from it that the social sciences should emulate the natural sciences. Indeed, this inference would make little sense. As indicated by some of the examples we listed above, there is no great consistency in the methodologies used by different natural sciences: those used by oceanography and fundamental physics are, for example, quite different and it would be impossible to emulate both at once.

The fundamental canon of methodology is all sciences must develop methods and methodologies that suit their theories, their puzzles and their data.

Paradigms, theories and models

Each social science discipline has its own theoretical traditions, but among those that cross several disciplines are rational choice, Marxism, Durkheimian and Weberian traditions, neo-realism and neo-liberalism, and various institutionalisms. We need not explain these traditions' ideas here, but we need to note the big methodological issue that it is quite easy to develop rival accounts from within the same tradition, as well as from different ones – to explain, say, an event or a regularity.

Paradigms

The intellectual core of a tradition consists in its *paradigms* or *frameworks* (the two terms are used interchangeably here). The canonical characterisation of a paradigm

comes from the work of the historian and philosopher Thomas Kuhn, whose monograph, *The structure of scientific revolutions* (1970 [1962]), continues to be an important point of reference in both philosophical and methodological debates. For our present purposes, we can define a paradigm as a:

> shared commitment by an identifiable group of scientists about what is to be observed or examined and therefore what counts as relevant data for answering those research questions which are deemed important, how those data should be interpreted, and what structure we should expect in satisfying answers to those questions.

In particular, the paradigm specifies what is required by its peculiar tradition to be included in any satisfying explanation of an event, temporary state of affairs, trend, regularity or condition. Example 2.3 shows how a specific paradigm – neo-realism – works in the field of international relations.

EXAMPLE 2.3. THE PARADIGM OF NEO-REALISM IN INTERNATIONAL RELATIONS

This paradigm proposes that we should look for explanations of the foreign policy decisions of national governments by reference to the interests which those states have in securing a better relative position vis-à-vis other states. It further proposes that when, say, prime ministers claim to be acting for the good of the whole international community, we should not interpret this claim at face value, but look, rather, at the underlying national interests served by their rhetoric.

That is, the neo-realist paradigm tells us to weigh evidence about the expected benefits and costs to a state's position more heavily than claims to shared benefits, and proposes, furthermore, that research questions about *interests* are much more important than research questions about either *ideas* or *institutions*.

How paradigms change

It might be thought that – as the guardians of the cores of traditions – paradigms are unchanging and more or less impossible to challenge, and that people who wish to dispute anything they have to say about what constitutes good research will be forced to find homes in other traditions.

This assumption is too simple. The broad traditions in social sciences – such as rational choice, institutionalism or Durkheimian social theory – could not survive unless they are sufficiently flexible to accommodate change. For example, Durkheimian social theorists used to explain arrangements such as family forms or industrial relations practices by referring to the function they performed for a 'society' as a whole. Today, Durkheimians would not do this. The idea that 'societies' – however they might be defined – have 'needs' has proven to be unhelpful, and can now be sustained only by being so heavily modified by other theoretical insights that it loses all claims to parsimony. Yet other kinds of functional explanations are still important in this tradition.

Again, conventional rational choice theory explains action by showing that people pursue their goals under constraints. Originally, this tradition assumed that people were basically neutral toward risk, but the advent of 'prospect theory' (Kahneman and Tversky, 2000) has challenged that assumption. This theory explains why people are averse to losses in some situations but willing to bear risks in others. However, prospect theory has preserved the basic idea of optimising subjective utility under constraint, and a debate is now going on about whether the change falls within the basic paradigm or breaks with it: for example, Levy (1997) finds the two approaches often empirically difficult to distinguish.

It appears from both these examples that 'peripheral propositions' – those which are less central to the identity of the tradition and less critical for its approach to explanation – can be surrendered if they prove unworkable, provided that 'core' or fundamental claims are maintained. Thus traditions persist, even though some aspects of their paradigms may change. Indeed, this view of how scientific traditions progress was set out in the work of an important philosopher of science, Imre Lakatos (1970).

Theories

Paradigms or frameworks do not, of themselves, describe, explain or interpret anything. But they tell us how descriptions, explanations and interpretations should be developed within the tradition in which we are working. We can then derive candidate explanations of particular regularities, trends or states of affairs or conditions from paradigms, so long, of course, as we have sufficient empirical data about these things. Candidate explanations of *particular* events, temporary state of affairs, trends, regularities or conditions are *theories*.

The theories with which we work for a particular piece of research may be extensions or refinements of ones that other people have previously developed, or they may be ones that we develop ourselves. But whenever we form a view about the *character*, *causes* or *significance* of events or conditions and so on, we are working, respectively, with a *descriptive*, *explanatory* or *interpretive theory* about them. Social researchers are, of course, usually aware of the theory from which they *deduce* the formal hypotheses that they propose to test in a piece of research. But, of course, the hunches or insights that lead us to stipulate research questions for more open-ended pieces of exploratory research are also based on ideas about states of affairs, conditions and so on. And so, the discussion in Chapter 13 about the use of deduction and induction in social science research will argue that there is necessarily an element of deduction from some kind of theory in the structure of all social science research.

Likewise, practical decisions made in government, business or social agencies depend on ideas, however implicit, about the character of particular social or business problems and the reasons why an intervention is likely to 'work' in the expected way. These ideas are, of course, theories, too. Indeed, some social scientists who work with such organisations (Pawson and Tilley, 1997) argue that one task of social science is to help decision-makers make explicit the theories on which practical decisions are based, partly so that these theories can be clearly stated and tested, but also to capture more systematically the tacit knowledge on which these decisions are based.

Suppose, for example, a local council decides to introduce CCTV cameras in the market square of their town. Their staff are, presumably, working with some ideas about why CCTV will reduce crime in that situation. These ideas – whether formulated in these terms or not – may be similar to, and can perhaps help further develop, those formally proposed by the theory of situational crime prevention we looked at in Example 2.2.

Rival theories within the same paradigm

Most paradigms allow us to generate several rival theories for the same phenomenon. So showing that one theory is false will not, by itself, falsify the paradigm. We discuss this point in Example 2.4.

EXAMPLE 2.4. THREE RIVAL THEORIES TO EXPLAIN ONE EVENT, ALL FROM ONE PARADIGM

Suppose we are interested in explaining why Soviet leader Nikita Khrushchev took the highly risky decision in 1962 to deploy nuclear missiles in Cuba, and to do so in secret.

The rational choice paradigm requires explanations based on individual desires, beliefs, capabilities and constraints. It assumes that a decision-maker has a set of intelligible goals derived from his or her strategic interests, and ranks them in a consistent and coherent manner. It assumes, too, that he or she arrives at this ranking using consistent inferences from the available information, and is not so constrained by circumstances that none of those goals can be pursued. Thus, a rational choice explanation asks for:

- a set of ranked preferences or goals, so that when the first preferences become unavailable, there are second preferences, and so on;
- a set of *beliefs* about how preferences can be pursued;
- a set of *capabilities* to act; and
- some *constraints on action*, such as limited information or resources, but not absolute constraints.

It is possible to develop several plausible theories that meet these requirements, indeed scholars of the Cuban Missile Crisis have done so.

(Continued)

One theory holds that Khrushchev was trying to protect Cuba from what he believed, on the basis of intelligence reports and inference from US policy, was a likely American invasion. Cuba had recently become a socialist republic following Fidel Castro's overthrow of the previous regime and the USA had sponsored an unsuccessful invasion by anti-Castro exiles in the previous year.

Another theory holds that Khrushchev was concerned only with trying to regain closer parity with the USA in terms of their mutual capability for threatening each other with nuclear war. This downplays Soviet concern for Cuba, other than as a theatre of superpower conflict.

A third theory holds that Khrushchev wanted to bind the fledgling Cuban socialist state into the Soviet system of alliances to stop it drifting under Chinese influence.

For each of these three theories, it is possible to supply plausible goals, reasons why Khrushchev might have believed that deploying Soviet missiles to Cuba would further them and reasons why he might have thought that the USSR had the capacity to do so.

Example 2.4 shows that three rival theories can be developed to account for Khrushchev's behaviour during the Cuban Missile Crisis, each consonant with the canons of rational choice. If, however, we could show that Khrushchev was not driven by any of the goals stipulated by these theories, we would not have shown that no satisfactory rational choice theory for his decisions could be developed – unless we were able to show, too, that there were *no* other goals that could provide a more convincing rational choice explanation.

Models

Each of the theories offered in Example 2.4 is schematic. By this we mean that, for each theory, we could specify slightly but *measurably* different accounts of the sets of:

- the constraints under which Khrushchev was working (financial, time, military and so on);
- the particular beliefs that the Soviet leader might have held; and
- the appetite among Soviet political and military leaders for different risks.

Each of the different ways of specifying the factors prescribed by the rational choice paradigm produces, in turn, a slightly different explanation of Khrushchev's decisions during the crisis, including their timing and the changing nature of his willingness to bargain with the Americans. So we could say that the different specifications of these factors prescribed by this paradigm would produce, for each of our three goal-based theories, a distinct *model* of Soviet decision-making in the Cuban Missile Crisis.

In social science, a model is a formal representation of exactly how a theory might be realised, showing how the explanatory factors are (1) to be measured, (2) predicted to be linked with each other and (3) how they relate to what is being described or explained. *Specification* is the development of precise statements of each of these three things.

In model-building, parsimony is usually weighted more heavily than goodness of fit or generalisability. This is because models are *reductionist* by intention: the main reason for building models is to isolate as clearly as possible the key factors and the relationships between them. This is why models are used in descriptive and particularly in explanatory inference, but to a lesser degree in interpretation which often tends to privilege other values, such as goodness of fit, over parsimony.

To revert to the example we used in Chapter 1, for example, a comprehensive model explaining a rise in the numbers of people sleeping rough in a town would need to include, say: changes in the characteristics of homeless people; routes out of homelessness; local labour markets; aspects of criminal justice policy and its implementation; demographic factors; and changes in national government policy or interventions by local councils. In practice, no one would attempt to construct a complete model of so complex a phenomenon. Instead, in striking trade-offs between various virtues of explanation, we would surely try to achieve more parsimony by excluding the less important variables, even at the expense of reduced goodness of fit or limiting the range of variance that could be explained.

A first test of a good model is that it is sufficiently 'specified'. In the case of an explanatory model, it must distinguish the important causal factors and specify their relationships with the outcome to be explained. And it must do this with sufficient clarity and detail to allow the model to be tested, provided the right data are available. Highly quantitative models often use algebraic specifications for testing statistical relationships between variables. But for other kinds of research, models can be specified adequately for testing using geometric representations (i.e. diagrams), or very rich, rigorous and careful qualitative statements.

The relationship between theories and models

Most theories are capable of being realised by several models, depending on how each factor is specified. Because theories can be realised by several models, it is usually easier to test a model than a whole theory. If one model fails, the same theory may yield another model that would be more consistent with the data.

It is sometimes possible to build a well-specified model, but lack a good theory to explain it. For example, we may be able to show a statistical association between a fall in gross domestic product (as a measure of economic activity in a country) and a rise in rough sleeping, but not know whether one causes the other. Our model will, nevertheless, still be of genuine interest if we believe that, sooner or later, we shall be able to develop a causal theory to explain the association. So when social scientists condemn someone else's work with that most damning of dismissals, that it is 'atheoretical', they may really be saying that a researcher has developed a model but has yet to specify a theory that provides a reason for being interested in it.

An important conclusion to be drawn from this discussion is that researchers should clarify both for themselves and for their readers just what inferential status they are claiming for their findings. For it is dangerously easy to claim more than the findings will warrant, for example by over-interpreting a finding as either supporting or undermining a theory, when, in fact, the research does no more than support or undermine a model. Similarly, it is dangerous to claim that a piece of research supports or undermines a paradigm when the findings do no more than support or disturb a theory.

EXERCISE 2.2. BUILDING THEORIES AND MODELS

Think about the rough sleepers example we have used in this chapter and the previous one, and try to formulate at least two, and preferably three, different *explanatory* theories that would account for a statistical association between an economic recession (defined as a fall in gross domestic product for two consecutive quarters) and the rise in the number of rough sleepers in a particular town in the same period.

Next, specify a model to test each theory. You can set it out in prose, if you prefer, or in diagrammatic form, or indeed algebraically.

Finally, identify the tradition(s) within which your theories are situated, and itemise the core features of its/their paradigms.

Basic reading

A good place to start for readers with little background in philosophical debates is Ford J, 1975, I beg your pardon, in J Ford, *Paradigms and fairy tales*, London: Routledge, 1–16.

Advanced reading

Kuhn TS, 1970 [1962], *The structure of scientific revolutions*, Chicago: University of Chicago Press.

A challenge to Kuhn which has attracted wide interest is Lakatos I, 1970, Falsification and the methodology of scientific research programmes, in I Lakatos and A Musgrave (eds), *Criticism and the growth of knowledge*, Cambridge: Cambridge University Press, 91–196.

THREE

Testing, confirming and falsifying

This chapter will:

- examine what is involved in confirming or disconfirming a model or a theory;
- discuss what counts as a test of a theory or model;
- examine how observations relate to theoretical claims made in theories and models;
- explore the risks presented by trying 'save a theory' by using auxiliary hypotheses;
- examine problems of comparing theories to see which works best; and
- consider the implications of this chapter for replacing theories and paradigms by better ones.

We have already introduced terms such as 'supporting', 'undermining', 'testing', 'confirming' and 'falsifying' to describe the relationships that data can have with models or theories. It is time to examine these terms in more detail.

Most theories of any interest have *something* going for them. Let us start with a simple example of a theory which is not deeply explanatory, but mainly descriptive. This is the widely discussed theory that since the 1960s, the way in which manufacturing production is organised has changed fundamentally. It holds that the system that held sway from the 1920s to the 1960s involved the mass production of highly standardised products for mass markets by vertically integrated firms that made their own tools and parts. This system was known as 'Fordism' because the Ford Motor Corporation was one of the earliest and most important companies to use it. According to this theory, Fordism has now broken down, and manufacturing has moved toward more flexible production systems capable of turning out a much greater variety of products aimed at small niches of consumers. These systems require shorter production runs, more variation in production tools and techniques, and thus more outsourcing of the supply chain for materials, parts and tools. Terms like 'post-Fordism' and 'flexible specialisation' are used to describe this new approach.

The theory does indeed have something going for it, and it is not hard to find examples that appear to confirm it. On the other hand, most theories also face at least *some* contrary evidence, as you will probably find if you do Exercise 3.1.

EXERCISE 3.1. IDENTIFYING CONFIRMING AND DISCONFIRMING INSTANCES

Write down some examples of trends in manufacturing and/or retailing that seem to be *consistent* with the theory of 'post-Fordism'.

Now write down some examples of trends in manufacturing and/or retailing that seem to be *inconsistent* with the theory of 'post-Fordism'.

If you are struggling to think of anything, it might to help to think about, say, ringtones for mobile phones or stores such as IKEA, and to consider what aspects of such phenomena are 'Fordist' or 'post-Fordist'.

Saving a theory from defeat by counter-examples

But if we rejected every theory that faced some contrary evidence, we would have to reject all or most theories in social science. This means that many – perhaps all – theories and models need to be refined in the course of their development, usually by supplying auxiliary hypotheses.

For example, many social scientists now offer explanations of welcome states of affairs such as low crime rates, good public health, impressive economic growth or peace between rival religions, by reference to 'social capital' (e.g. Putnam, 2000; Halpern, 2005). Social capital is defined and measured in somewhat different ways by different scholars, but a conventional characterisation emphasises the importance of benign social connections of acquaintanceship, friendship, mutual respect and trust. But only a moment's thought is needed to realise that successful criminals, too, must trust each other, and must have strong social bonds. So the simple version of the theory – that all positive social ties are beneficial in reducing crime – cannot be right.

But some auxiliary hypotheses can save the theory. For example, we can suggest (with Sampson *et al.*, 1997) that what is most important are the ties that connect law-abiding people (1) to criminals so that they know who is doing wrong and (2) to each other and to people in authority so that they will be willing to report crime. Thus, even if criminals are strongly bonded – provided other ties of the right kind exist – we might be able to explain reduced crime rates.

What have we done by adding these auxiliary hypotheses? First, we have added new content. Second, we have used that new content to qualify and modify the original theory, without abandoning the causal force it proposed. Third, we have made it much more specific. That is, we have moved closer to turning a general theory into a testable model. In doing so, we have saved the theory from falsification by some obvious contrary evidence.

EXERCISE 3.2. SAVING A THEORY USING AUXILIARY HYPOTHESES

Think about how to *save* the theory of post-Fordism from simple defeat by the counter-examples you thought about for the last exercise.

To do this, write some auxiliary hypotheses that could be added to the theory and which would allow it to be consistent with your counter-examples.

Crucial tests of theories and models

Ideally, we want to design research that provides a *crucial test* of a theory, or at least of a model. To do this, we need a test that would show that there is no other *plausible* explanation than that offered by the theory or model. Thus, a *crucial disconfirmation* occurs when we can show that that no plausible auxiliary hypotheses can be developed that make the theory consistent with the data – the theory has to be wrong. *Crucial confirmation* occurs when we can show that no plausible auxiliary hypotheses can be developed that would make the theory *inconsistent* with the data – the theory has to be right. Crucial confirmation tests for anything interesting are rare in social sciences, and it is much debated whether crucial disconfirmation tests are available at all.

In practice, it is often possible to find alternative explanations for most findings produced by social science research, at least if we are prepared to make some changes in our models to accommodate them. This means that it is rare that we find a *crucial* test.

EXERCISE 3.3. IDENTIFYING CRUCIAL TESTS

Go back to the theories and models that you developed in Exercise 2.2 for an apparent rise in rough sleeping.

Choose one. Can you think what it would take to design a piece of research that would be a crucial test of your explanation?

But this discussion raises an important question: what changes should we be prepared to make to our theory and what limits are there on making changes to save it?

Limits to saving theories and models

There is a view that it does not matter whether explanations in theories or in models *are plausible* or not, provided that they accurately fit the data – that is, if the predictions of the theory or model are borne out by the evidence, then that theory or model should be accepted. So any changes to a theory or model are also acceptable, as long as they fit the data. This view was famously argued by the economist Milton Friedman (1979 [1953]). One reason that he advocated this extreme position was that it enabled him to offer very parsimonious theories and models in economics, which posited little by way of unobservable mental life – beliefs, desires, motives, for example – among the people being studied, especially at the level of the firm. Provided the predictions of economic theory about how firms would behave in the face of changes in prices or whatever were borne out by their behaviour, it did not matter what was going on in the decision-making processes in the firms' board rooms to produce those outcomes.

Several criticisms have been made of Friedman's argument, not least that, to save our theories or models, we may have to pay too high a price in terms of their implausibility. Another is that it makes explanation just too easy to be interesting, because we can always come up with assumptions that will yield an exact fit with the data. In statistics, this inferentially risky procedure is called 'curve fitting'. There are situations in which two curves might show an equivalent goodness of fit with the distribution of the data points: so, on that ground alone, we could not choose between them. Requiring theories and models to be parsimonious may not allow us to choose between them.

So, when contrary evidence turns up for one or more of the things predicted by a theory or model, we have to make *scientific judgements* about what we are prepared to ditch. The problem is that, although various philosophers of science and methodologists have worried away at this issue for decades, there are no precise and generally accepted rules about the plausibility of auxiliary hypotheses and assumptions that could guide researchers.

But there are rules of thumb. Most people would say that if we find ourselves introducing more and more hypotheses to save a theory or a model, especially if they appear to be *ad hoc* in that they are not driven by the core ideas of the underlying theory or even the framework, and if they appear to do nothing but save the theory or model from one or two sets of recalcitrant pieces of evidence, then the theory or model is in deep trouble. The stock example in the textbooks is that of the attempt by Apollonius of Perga to save Ptolemy's theory of planetary movement. This attempt is described in Example 3.1.

EXAMPLE 3.1. AD HOC ADJUSTMENTS TO SAVE THE THEORY FROM THE DATA, BUT LOSING COHERENCE AND ADDING NO NEW EMPIRICAL CONTENT

Ptolemy's theory required the planets to move tidily in circles around the earth. Five of them persistently turned up in places in the sky that did not fit these circles. Undeterred by this inconvenient behaviour, Apollonius proposed to add some extra little circles (epicycles) to their paths, on top of the main circle (the deferent).

(Continued)

The resulting additional complexity produced some very odd-looking lines that did not seem to be driven by the basic idea of one big circle – but it did save the data.

One problem with these epicycles was that it was hard to derive any new predictions of observations from them. Put another way, they did not add any new 'empirical content' to the theory that could be tested against new observations: they just helped the theory better to fit the data that were already available.

Had there been a better theory available – one which did not require epicycles – then perhaps astronomers might have ditched Ptolemy's theory earlier than they did. Unfortunately, the only rival for a very long time was a theory – the heliocentric theory – that created more anomalies that also called for more epicycles to save the day. Trying ellipses worked better, because they required fewer *ad hoc* adjustments, but even they were not quite right. As a result, theory replacement took centuries.

The risk also arises in social science, of saving the data by means of *ad hoc* adjustments which can leave the theory appearing to degenerate. Marx, for example, famously propounded a theory of value which held that the value of a good or service was a function of the labour that went into making and distributing that good or administering that service, and that it had little to do with the capital investment involved. This required him to accept that prices and values could diverge very radically, but that was not a problem for Marx, who wanted to show that the capitalist system produced distorted prices. But as goods became steadily more capital intensive, and the amount of labour directly involved in making them fell, Marxists had to save the theory by arguing that the capital involved should be counted as a product of labour from other parts of the economy, perhaps even the distant past. This meant, however, that goods became progressively more valuable, even though they were getting cheaper. This stratagem detached the idea of a value from that of a price almost completely, so the theory was being saved from badly behaved evidence, but at the price of so extending the concept of labour that it had drifted far from its central theoretical imperative.

Observations and theory

This leads us to another important issue about how we should use observations to test, confirm or disconfirm models and theories.

In social science, we collect data in a variety of ways. We use social surveys based on questionnaires; we use semi-structured interviews; we ask participants to keep diaries for us; we use observation protocols to organise the collection of field notes; and, sometimes, we undertake either laboratory or field experiments. Each type of data rests on what we might call a *technology* of data collection. But this technology is, of course, itself a theory, one that makes claims, first, that the responses of people

are capable of being captured by such instruments, and, second, that the data so produced will suffice to test a theory.

In order to analyse our data, we need to conduct some interpretation. Interpretation is done through the way we record and code data. This process too must rest on theory of some kind. Example 3.2 shows that this is because the *terms* in which data are recorded and coded depend, for their meaning and significance, on the theories in which they are based.

EXAMPLE 3.2. CODES TAKE MEANINGS FROM THEORY

Suppose we are coding responses from a transcript of an interview conducted in a doctor's surgery and that the purpose of the interview is to find out about the surgery's practice in handling patient records. We use a code that rates 'willingness to disclose details about a patient to the police on request if the police claim they have evidence of an urgent threat' – an example would be a threat to a child arising from the patient's mental health problems.

Next we develop other detailed codes to rate other aspects of the surgery's practice in relation to disclosure of information from patient records.

After examining the data entered in all the relevant codes, we infer that we can enter '*moderate* protection of patient confidentiality' in a higher-level code that summarises information in these codes.

You will notice that we have used the verb 'to disclose'. Its meaning in our coding scheme derives from the rules and conventions that we adopt to tell us what counts as a 'disclosure' of personal information for the purpose of *our* research. These rules and conventions are, in turn, derived from the substantive theory we wish to test.

Suppose, for example, the research is being conducted to test a theory about the organisational practices that shape how doctors make judgements about the circumstances that justify breaches of patient confidentiality. For such purposes, correspondence with a hospital specialist in the context of referring a patient for medical tests would not count as a relevant 'disclosure', but supplying details of a patient's illegal drug use to the police or social work agencies for the purpose of detecting crime or preventing child abuse would. We would also need conventions, justified by theory, enabling us to judge consistently what counted as 'weak', 'moderate' or 'strong' practices for regulating disclosures thus defined.

This example illustrates the point that, when we come to analyse coded data, we are analysing a set of data that are not in any sense 'raw'. If they were raw descriptions of sounds heard in the interview, they would be useless to us for the purposes of research.

The point illustrated in Example 3.2 has important consequences for using data to reject theories. For example, suppose that our analysis of our coded data suggests that our initial theory about the kinds of confidentiality protection to be expected from doctors appears to be wrong. One possibility is to write an article showing that another poor theory has bitten the dust. But before doing that, we should consider other possibilities.

Could the theory about confidentiality be right, but could there, instead, be something wrong with the theories that underpin the technology we used for data collection, instrument design and coding? It might be possible, for example, to defend our theory of organisational practice by arguing that our empirical test was not a fair one, because

the other theories embedded in our 'technology' of interview schedules, coding frames, coding rules – including the meanings given to such terms as 'disclosure' – were faulty.

Some philosophers of science have suggested that there is potentially no limit to the number of theories that could be involved in a piece of research, because, in principle, each term or code we use will rest on still other theories and conventions that could also be challenged. The logic of this argument is that almost the whole of human theorising could be up for grabs in any one piece of research, if we are prepared to be sufficiently imaginative about saving some theories at the expense of others.

But in practice, of course, the situation is rarely so serious. Typically, we bound the set of theories we need to consider by refusing to sacrifice well-established ones about what terms mean or what a particular method does. Conversely, we also avoid defending theories that require *ad hoc* or degenerating adjustments unless there really is no practicable alternative. This brings us to a rule of thumb for deciding which theories to accept and reject when we find anomalous evidence. This can be called a *provisional rule of caution*, and it states that we should normally reject the fewest theories, and particularly those theories the loss of which would be most damaging for the body of social science knowledge. This is to recognise that all social science research necessarily stands on the shoulders of wider bodies of knowledge, including those relating to theories about method and methodology.

But this rule does not relieve researchers of the responsibility to specify models as clearly as possible and to focus empirical tests upon the principal, and especially the more controversial, theories upon which their judgements rely. In particular, methods and methodology sections in articles and theories require us to defend these specifications and designs on which our data collection, coding, analysis and interpretation rest. Exercise 3.4 illustrates this important point.

EXERCISE 3.4. DISTINGUISHING CORE FROM AUXILIARY THEORIES; MAKING SCIENTIFIC JUDGEMENTS ABOUT WHICH THEORIES TO AMEND, WHEN ANOMALOUS DATA ARISE

Go back to your work on rough sleeping for Exercise 2.2.

Take both your most plausible and your least plausible theory of the trend in rough sleeping. Now think about what you would do if some observations you made yield data which, apparently, conflict with these theories. If you were to save these theories, you would have to ditch other theories.

Write down some theories that you would, and some that you would not, be prepared to abandon for the sake of saving your explanation of the rise in rough sleeping, and explain your rationale.

Comparing rival theories

Some of the best, and certainly the most ambitious, social science is done when researchers take two bodies of theory, derive rival predictions from them and do some empirical work to decide which prediction is better supported by their data. In social science, it is probably done too rarely, because it is a hazardous process.

One reason that is so difficult has already been explored. Because paradigms do not yield only one possible theory about most matters of interest, and theories can be modelled in different ways by making changes to how dimensions are specified, achieving a genuinely crucial test to choose between theories is hard. Suppose, for example, that we want to choose between theories of acquisitive crime.

One theory – which is known as the 'market reduction' theory (e.g. Sutton *et al.*, 2001) and which is, in turn, a version of rational choice theory – predicts that overall levels of acquisitive crime, such as burglary, will fall when the demand for stolen goods falls, because criminals' motivation for committing acquisitive crime is reduced. We might develop another theory, derived from the 'prospect theory' approach of decision-making under conditions of risk. This theory would predict that if burglars see the price of stolen goods falling they may be tempted to risk more high-value burglaries to sustain the level of income from their trade.

If on the data from the UK, the market reduction theory outperforms this version of prospect theory, this fact would not enable us to infer that the latter theory is comprehensively defeated, even in the British case. For there may be yet other theories relating to the behaviour of burglars in response to risk that could outperform both theories. The most that we can claim for our work is that we have *shifted the onus of argument* onto those who still wish to defend the use of prospect theory in the study of acquisitive crime.

Is there a problem of incommensurability?

Another issue that has greatly preoccupied the philosophers of science is that the meanings of terms – like 'demand' and 'response to incentive' – may be defined *internally* so that they mean slightly different things in rational choice and the prospect theory. If that were true, then a test that compared the two theories would be misguided, because it would not be comparing like with like. Kuhn called this the problem of 'incommensurability' of meanings of theoretical terms.

EXAMPLE 3.3. THE MEANING OF A TERM IN TWO RIVAL THEORIES

Consider two well-known theories of what drives the behaviour of top civil servants.

One theory developed in the early 1970s (Niskanen, 1971) says that what they really care about is increasing the size of their ministries, their budgets and their payrolls, because this process increases their importance and salaries as a result of giving them greater responsibilities.

(Continued)

The theory recognises that they will not be able to succeed in achieving those goals all the time, because they face constraints from politicians and ultimately from voters, who will not indefinitely elect parties which keep raising taxes. But subject to these constraints, the theory of 'bureau budget maximisation' predicts that this aim will shape the behaviour of civil servants. In its support, the advocates of the theory pointed to big spurts in the size of the civil service in several countries through the 1950s and 1960s.

A rival theory developed in the 1990s (Dunleavy, 1991) holds that their status and power – and the possibility of having a congenial working life – come from being able to influence policy. That requires having ready access to politicians, and having time and responsibility for leadership in developing policy. What they do not want, this theory says, is to spend huge amounts of time in managing large-scale service provision. Better instead to find ways to slough off that sort of dull, low-status work to specialist agencies.

Again, there may be constraints on the extent to which politicians will let civil servants shed these responsibilities, and there are limits, too, on the extent to which they will allow top civil servants to monopolise policy-making. But, nevertheless, top civil servants will engage in 'bureau shaping' by trying to shape their ministries into tight, influential, policy-focused machines with strategic rather than managerial functions, even if this means that they have smaller budgets.

In evidence, the advocates of the theory pointed to the support given in the 1980s and 1990s by top civil servants in the UK for the devolution of management responsibilities to specialist agencies and indeed for the privatisation of functions from government departments.

At first sight, these two theories appear comparable, but in fact they are difficult to compare directly. This is because some important terms that appear in both theories are defined rather differently in each of them. Furthermore, these differences become very important at the point where we try to find empirical evidence to compare their performance in explaining changes in the structure of government.

Let us take the concept of a bureau budget. The bureau budget maximisation theory does not restrict the size of the budget to the annual sums of expenditure spent directly by a government department. They may well include the budgets of agencies that have been 'hived off' to handle specialist functions on its behalf. There are, after all, many ways to measure the size of one's bureaucratic empire, if one needs to boast about it.

In contrast, the whole point of bureau-shaping theory is to draw a tight distinction between the 'core' budget directly spent by the department and the budgets of devolved or privatised agencies.

Furthermore, 'bureau budget maximisation' can be tested by financial data that admit of year-on-year comparison and between-department comparison. But 'bureau shaping' is also about nurturing certain kinds of capabilities that may change their meaning between periods and between departmental contexts and need to be qualitatively assessed.

The constraints on civil servants' behaviour are of different kinds in these theories, too. In 'bureau budget maximisation', the constraints come from political control on spending, whilst in 'bureau-shaping' theory the constraints come from changing opportunities for involvement in interesting policy work. But this latter theory does not predict what top civil servants will do once all the service provision that it is politically possible to slough off has been shed or outsourced. In contrast, we know what bureau budget maximisation theory predicts they will do in any given period, even if it is just 'more of the same'.

In each theory, then, the concept of 'optimisation under constraint' is central. In each theory, there is a goal to be optimised, and there are constraints imposed by political acceptability.

(Continued)

But the meaning of optimisation under constraint appears to be very different in the two theories. Should we infer that the meanings of the term are 'incommensurable', such that no direct test of their rival strengths can be carried out? Well, not necessarily.

To enable comparison, we shall need to move down from the level of theory to the level of the model. Here, we can itemise the differences of meaning in the term 'optimisation under constraint'. If one theory is underspecified in its meaning in one respect where the other is specified – as in the question of what the bureau-shapers do when they have sloughed off all the work they can – we shall need to supply new specifications, to force out a comparison.

The simplest thing one could supply is the null hypothesis, but even that might be unavailable: after all, what would count for a top civil servant as 'doing nothing very much' about the size, scope and functions of the organisation he or she runs? But even if we could conduct a comparison by this means, the advocates of the failed theory could claim that 'it's the model, not the theory, that has failed; another model might show it to be perfectly sound'.

But this leaves the advocates of these theories with a dilemma. Either they show us a way of comparing theories directly – without the need to specify particular models – or they accept that their theories are indeed incommensurable, so that they cannot criticise their opponents. Otherwise, they must let us specify some models, to see whether we can test them in ways that will at least shift the onus of argument one way or the other.

Change and replacement of theories and paradigms

Much philosophy of science is concerned with the ways in which paradigms are replaced by others. This is not just a philosophical problem: it is also a practical problem about how we should understand developments in a body of research to which we are seeking to contribute.

Some writers who are familiar mainly with natural science tend to write as if they believed that paradigms succeed each other, and, once succeeded, die and become part of the history of science. And indeed there are some fields in the natural and the physical sciences of which that statement is not wholly false. It is not a complete travesty of the history of physics: for example, Newtonian dynamics has been replaced by a body of theory developed originally by Einstein.

In the social sciences, though, linear succession of traditions happens more rarely, with the consequence that traditions are hardly ever entirely superseded. For example, an emphasis on social and political institutions was very important in political economy of the 1920s and 1930s, as well as in the classical sociology of Weber and, in a different manner, of Durkheim. The behavioural turn after 1945 in political economy and sociology eclipsed institutionalism, which had come under fierce criticism. Yet institutionalism did not die. Some leading journals continued to publish some institutionalist work. By the 1980s, institutionalism had been reworked and recast but continued to draw on some of the same sources of theory as those which drove the interwar tradition. Today, it is a mainstream tradition of explanatory research in more or less all social sciences.

The same pattern can be observed in the history of the rational choice tradition. Rational choice of a kind flourished in eighteenth-century France, and was revived in late nineteenth-century England with the work of Herbert Spencer. Although it was eclipsed in the interwar years, it was revived immediately after the Second World War by several groups of mathematicians, mainly in the USA. Today, it is perhaps the pre-eminent approach in American social science and hugely influential everywhere else, too.

These examples emphasise the provisional character of what social science, or indeed any science, can offer. Better paradigms may, perhaps, emerge, or better theories than existing ones. Or perhaps not. Perhaps in a century from now, our descendants will still be reshuffling the same pack as the one that we now have and that was available in its main contours by the 1920s. But there is a significant possibility that over the next decades, social science will reject some, perhaps much, of what we believe that we have achieved. This possibility does not imply that there is no reason for confidence in what we are now able to do. Still less does it imply that any particular judgement about which theories to reject or to accept is defensible.

Basic reading

Ladyman J, 2002, *Understanding philosophy of science*, London: Routledge.

Okasha S, 2002, *Philosophy of science: a very short introduction*, Oxford: Oxford University Press.

Both texts are accessible, but are not specific to social science.

Advanced reading

A more advanced text which is specific to social science is Kincaid H, 1996, *Philosophical foundations of the social sciences: analysing controversies in social research*, Cambridge: Cambridge University Press.

Popper K, 1959, *The logic of scientific discovery*, London: Routledge, is the canonical statement of Popper's early argument about falsification, but the book is not easy reading for social scientists.

FOUR

Perspectives on findings from social research

This chapter will:

- show that the arguments we encountered in Chapter 3 represent some of the main positions in the philosophy of the social sciences;
- examine what is meant by 'positivism', 'relativism' and 'realism';
- briefly explore the meanings of 'constructivism' and 'constructionism'; and
- explain the implications of each position for deciding between rival theories.

Why philosophical disputes matter for methodology

The argument in Chapter 3 occupies some very specific terrain in the philosophical battles which have been fought over the status of claims to knowledge that can be made for the products of social science research.

There are some, very practical, reasons why it is useful for the study of methodology to understand the basic contours of these philosophical battles. The first is that when social scientists defend the methodology they have followed in their own research, or when they criticise the work of other social scientists, they often use labels for various '-isms' as shorthand for key issues at stake in these battles. Furthermore, even famous social scientists sometimes use these labels inexactly or as all-purpose terms of abuse for styles of research they do not like. As a result, mistaken notions have arisen as to what is implied by particular methodological choices. A more precise understanding of these labels, and what is implied by them, can help us recognise the more obvious elephant traps in such debates.

The most important reason for understanding philosophical arguments is that they are important for understanding the kinds of warrant that we can claim for our conclusions, especially when we want to confirm or undermine a model or theory. This is because different philosophical positions give quite different statuses to our

claims about inference – that is, to our claims to know about things that we cannot directly observe, such as people's preferences or their social networks or their political institutions.

There are a great many positions and perspectives in the philosophy of social science. Philosophers have a penchant for classifying them, and setting up debates between them. In this chapter we shall concentrate on three of the most important labels that are used as campaign badges in the philosophical wars among social scientists. We shall offer some definitions and also some warnings about what is, and what is not, implied for methodology, in taking these positions, Finally, we shall show how the arguments we are making in the first part of this book offer support for one of these positions.

The three positions on which we shall concentrate are 'positivism', 'relativism' and 'realism'. They have been chosen because these are the most common labels attached to the insults thrown at each other by social scientists. We shall also, in passing, issue brief warnings about 'constructionism', 'constructivism' and 'pragmatism'.

Positivism

Positivism is one kind of empiricism. Its basic idea is that all knowledge comes from experience – that is, from what we observe, in sense 3 identified in Chapter 1. Here, observation refers not only to visual observations but, in the broader sense, to what we can detect with all our senses or with the instruments we possess to extend and enhance them.

Empiricism argues, further, that when we develop abstract concepts and theories, we are doing no more than summarising our experience. This process, empiricists would argue, involves identifying patterns that recur in the data. These recurrent patterns are the real content of explanations. *Explanation* consists in no more, and no less.

Empiricism has a long history in human thought. Defined thus, it refers to a philosophical outlook that can be found in ancient Greek, Indian and Chinese thought. It was brought to a high degree of development in European thought in the seventeenth and eighteenth centuries, not least by Scottish Enlightenment thinkers such as David Hume.

However, in the recent philosophy of science, positivism has a more specific meaning than this. It holds that all claims to theoretical knowledge should be capable of being *reduced to* statements about empirical observations and the patterns found in them. It follows that explanation consists in the ability to project these patterns from observed to unobserved cases, to *predict* successfully what will be found by further observations. The quotations in Box 4.1 offer as simple a statement of the positivist position as can be found, without resorting to caricature. They are all taken from the work of a philosopher of science, Larry Laudan.

BOX 4.1. A DEFINITION OF POSITIVISM FROM LAUDAN (1990)

> Science is the attempt to codify and anticipate experience. The raw materials of science are raw data or observational measurements. We develop theories and laws to correlate, explain and predict those data. A science progresses just to the extent that later theories can predict and explain more phenomena than their predecessors did. (Laudan, 1990: 2–3)
>
> [T]heories are sets of universal statements which are genuinely empirical by virtue of their prohibiting certain states of affairs. (Laudan, 1990: 5)
>
> [A] theory's confirming instances and the things it explains are the same set. (Laudan, 1990: 21)

You have already met, in Chapter 3, one example of a positivist position. That was Friedman's argument that we can explain economic behaviour *as if* people optimised under constraint. So long as this explanation fits the data and serves as a basis for prediction, we need not worry about the plausibility of the assumptions on which it rests.

The following distinctive methodological principles therefore follow from the positivist position:

1 Abstract, unobservable things that we introduce into our explanations in social science – such as rules, institutions, interests, networks, power, organisation or organisations, beliefs or ideas – should be treated not as special kinds of entities or facts that are different from the observable behaviour of human beings, but simply as convenient shorthand for structured patterns of relationships between the measures or codes captured from our observations of that behaviour (that is, they show strong goodness of fit). If those patterns are useful, they will be capable of being projected onto new cases (that is, they exhibit strong generality), thus demonstrating the power of our explanation to predict future observations.
2 If we design the right research instruments, a theory, in principle, could be reduced to a set of empirically testable statements of generalisation about expected patterns between things we can observe.
3 The patterns that hold the key to theories are *regularities*. Events, states of affairs, trends and conditions are of interest for *scientific* social research when they can be shown to form regularities.
4 If two theories predict the same behaviour – or events or trends or conditions and so on – then they are equivalent theories, whatever the abstract terms they use suggest about possible differences. It is still possible, however, that one theory might be *better supported* by the empirical evidence. For example, one theory might perform better in the number or severity of anomalous or apparently disconfirming cases (Laudan, 1990: 64).

Positivism, verification and falsification

One form of positivism which was widely influential in the first half of the twentieth century adopted a view of research which emphasised *verification*; that is, a commitment to testing theories empirically by searching for *confirming instances*. Thus, if we wanted to test the market reduction theory of crime reduction we looked at in the previous chapter, we might apply it to an increasingly wide range of markets for illicit goods, such as illegal drugs or counterfeit products, as well as stolen goods.

The arguments of Karl Popper were widely considered to have been devastating to the idea of verification, and thus to this form of positivism. He pointed out that confirming instances can be found for most theories, but simply adding more and more confirming instances is not a fair test. Piling up evermore cases where market reduction leads to crime reduction does not test the theory adequately, because we shall still not know whether there are cases in which it does not work in the predicted way. Instead, Popper (1959) argued, we should concentrate on what a theory predicts will *not* be observed and then design research to see whether it *can* be observed, thus providing a *disconfirming* instance.

The key methodological point that emerges from this controversy is that to be useful, and to meet scientific standards of rigour, theories must be stipulated in ways that make them empirically testable and, according to Popper, falsifiable, too. The proposition that 'reducing the market for illicit goods leads to a fall in the numbers of crimes' is, in principle, both a testable and a falsifiable one, because a single case where this relationship does not hold would undermine the theory. The problem that we have already seen is that failing to find disconfirming instances does not prove that the theory is true, though increasing the number, and particularly the variety, of cases we test *might* enable us to find out how widely it holds.

The really interesting scientific judgements are made when a process of disconfirmation allows us to see more clearly what has to be done to preserve our theory. If, for example, it were found that the market reduction theory of crime prevention is generally supported in certain markets for illicit goods – such as stolen goods fenced through local shops or bars – but disconfirming instances were found in other markets – such as those for stolen works of art – then the theory would be increased in precision, and market reduction interventions by police forces could be accordingly refined.

Contrary to what is often said, Popper's approach to methodology – concentrating on finding disconfirming instances and then making judgements about which patterns appear most consistently to predict new observations – is perfectly consistent with the four core positivist positions on the nature of data and explanation that we have identified above. Indeed, Popper's fundamental break with positivism was based on arguments other than the one about falsification, but we need not discuss those arguments here.

Likewise, the point we made in Chapter 3 – that all data are necessarily stated in forms that imply some theory – is not a fundamental problem for positivism, correctly understood. Unlike its predecessor, empiricism, positivism does not require that experience is always captured in terms that are free of any theoretical implications. Rather, positivism requires only that we should design research that relies on theories of observation, measurement and coding that are – to some degree at least – more dependable than the theories we are testing (Laudan, 1990: 47–8).

Misapprehensions about positivism

There are, too, some other misapprehensions about positivism that frequently appear in discussions of social science methodology, especially by those who think their positions require them to be hostile to it.

Many social scientists, for example, accuse their peers of 'positivism' when they use statistical analysis of quantitative data sets to model social phenomena. However, nothing in the characterisation of positivism requires the use of any particular *kind* of data. Conversely, the use of a particular method or a technique does not imply a particular philosophical position.

Someone who undertook statistical analysis of quantitative data would be a positivist, only if they went on to claim that achieving a successful statistical prediction *is all there is to explanation*. A genuinely positivist view of explanation would be that of Milton Friedman who, as we saw in the previous chapter, claimed that no further consideration need be given to a model's assumptions, or to their plausibility, or to the problem of *how* the model is supposed to work. All we need to do, the positivist argues, is carefully examine the consistency of the assumptions in the model with existing knowledge, which, for a positivist, means the patterns that have already proved to be successful in making predictions. However, most social scientists who use quantitative methods to test statistical models would *not* make that kind of philosophical claim about the status of what they achieve with their statistical tests.

Another mistake – one often made by some social scientists – is to suppose that positivism implies *behaviourism*. Behaviourism refers to a *practical* approach to research that aims to eliminate accounts of mental life or normative prescriptions from substantive theories, to focus exclusively on what can be observed – namely, behaviour.

The conflation of positivism with behaviourism is important because many social researchers who are interested in the significance of what they often call 'meanings' – a term that covers people's ideas, desires, beliefs, systems of classification, preferences and so on – often claim that their approach to explanation is inconsistent with positivism. Strictly, this is not the case. Positivists have no objection to setting theories out in ways that include terms for unobservable entities – including mental states – provided that we concentrate only on what those theories predict about observable things rather than what they purport to claim about unobservable ones. For positivism has no room for knowledge claims based on untestable assumptions about the existence and character of abstract, unobservable entities. That is, positivists accept that, in practice, it would be cumbersome and practically useless to 'reduce' all theoretical statements only to the observations they imply. The issue is what status we can claim for a theory's assumptions, not what terms are allowed in the statement of the theory.

Positivism and explanation

We can see, then, that many arguments that are frequently used to attack the methodological claims of positivism turn out to be flaky. But this does not mean that we cannot make more powerful criticisms of positivism than the ones discussed so far. It does mean, however, that if we wish to reject positivism, we should do so for the right, and for quite precise, reasons.

One of the most controversial claims of positivism, in social science at least, is that, so long as we are confident of the accuracy of our data, we should privilege goodness of fit – albeit one achieved as the result of a good deal of data reduction – in any trade-off that we might have to make with other methodological virtues, such as the plausibility of our assumptions or causality or even generality. This is because the positivists' understanding of 'explanation' prizes predictive power – the achievement of a good fit between prediction and observations – over the search for 'explanation' in its everyday sense of understanding of how observations relate to deeper causes.

But to be told that a pattern can be successfully projected from observed data to new sets of data does not amount to a *satisfying* explanation. It does not tell us *how* or *why* the pattern works. Positivists claim that the questions 'how?' and 'why?' are illegitimate ones. But the only support they can offer for this claim is to assert that an explanation that does not consist in an empirically observable pattern must be a piece of metaphysics about things that are unobservable. In our view, this argument limits positivism to explanations that are bound to be more or less unsatisfying.

For example, if we found that market reduction theory worked in some, but not other, markets for illicit goods, we could re-examine the data from our cases to see whether it is possible to establish *why* that might be the case. We might, as a result, be able to develop a more refined *explanatory* theory about how different kinds of illicit markets operate and how each affects criminal behaviour.

Causality matters for making explanations intellectually satisfying and meaningful in ways that positivism cannot capture, because positivism values explanation only in so far as it yields predictions. Furthermore, plausible accounts of causality depend on making inferences about entities or facts that cannot be directly observed – such as the very notion of 'markets' for illicit goods or the idea of burglars responding to particular incentives – which should also be taken more seriously than positivism allows.

This leads directly to the second criticism, which is that the positivist's conception of scientific progress – including progress in social science – is *empirically* inadequate. A great deal of scientific progress has consisted not in predicting yet more empirical observations, but rather in conceptual development (Laudan, 1977), and especially in the development of theories about *how* causal processes work. Often the research that supports this work is done with existing data sets, in ways that claim to produce explanations that are richer, more satisfying and conceptually better organised, even when they do not predict new observations. And this progress in better specifying theories is sustained, in practice, not by the pursuit of maximising variance predicted by correlations but by the pursuit of explanations which take causality seriously, by proposing causal explanations which use facts which cannot be directly observed but only inferred. This means that any philosophical position that aspires to do better for methodology than positivism must recognise the important role of conceptual and theoretical development in the development of social scientific knowledge.

Relativism

Too often, this term is also used rather loosely. A more precise statement follows shortly, but, broadly defined, relativism claims that we cannot account for what we observe, independently of the ways in which we recognise, classify, code and analyse our observations. So the truth of statements must always be 'relative *to*' the paradigms and conceptual frameworks within which we collect and analyse data.

To turn again to our crime prevention example, relativists would claim that is there no possibility of achieving an account of the causes of acquisitive crime, which captures a truth of the matter in ways that are independent of the paradigms we apply to its study. They would emphasise that the market reduction approach is situated in the rational choice paradigm which assumes that crime is committed by offenders who are autonomous and rational human beings. They might also point out that, for example, work situated in an institutionalist paradigm – such as the work of Sampson we cited in Chapter 3 – would look to very different sets of constraints and opportunities to explain increases or reductions in crime, such as the changing strength and nature of social ties in the relevant communities.

Relativism claims, then, that *if* we categorise the problem in the way that rational choice demands, then we establish measures and gather and code observations in one way, but if we follow institutionalist categories, then we would measure and then gather and code data in a different way. And there is no way of determining which approach is superior that does not appeal to one or other of these paradigms, or perhaps to a third paradigm which has no greater authority. It follows that, for relativists, there can be no such thing as a science of criminology, possessing knowledge of facts that have determinate existence independent of the means of classifying, categorising and measuring them.

For the purposes of this book, we shall use a definition of relativism which is consistent with mainstream usage in the philosophy of social science, and say that *relativism is the doctrine that (1) there is no compelling reason to accept that scientific explanations are true or false on the basis of determinate facts; and (2) scientific methods of research and inference do not alone supply such a compelling reason.*

Some relativists in both philosophy and social science have recently reclaimed the old term 'pragmatism' (e.g. Rorty, 1990). This is confusing, because the historical founders of pragmatist thought were by no means all relativists, and even the few who came close to relativism had rather nuanced and qualified views. Moreover, there are important contemporary philosophers of science who call themselves pragmatists but who argue fiercely against relativism (e.g. Laudan, 1977).

Responses to the main arguments for relativism

The main arguments offered in favour of relativism have in fact already been rehearsed and rejected in Chapter 3, although at that stage, we did not present them as such, but rather as criticisms of an overly mechanical understanding of how data can be used to establish whether a theory is 'true'. We recap them here, and the rejoinders we offered to them:

'Theory is underdetermined by data.' This is the argument that data alone do not suffice to choose between theories. This is certainly true, and no one denies it. Positivists recognise that simple counts of confirming and disconfirming instances alone do not determine which of two theories we should accept, and that, even when we conclude that the data leave a model in trouble, we are making inferences. But, so what? It does not follow that, just because an inference requires scientific judgement and is risky, the statement inferred cannot be true or false save within one paradigm.

'Terms are incommensurable between theories.' It is sometimes true that the same term has different meanings in rival theories. But it does not follow that the two theories cannot be compared, because we have ways of moving down from theories to models that force us to specify meanings in ways that will permit comparison. This procedure requires us to draw more limited conclusions about theories, but it does enable us to shift the onus of argument in very important ways.

'Theories do not confront data alone, but in larger bodies of theory, and we have choices to make about which elements to amend in the face of anomalous data.' Again, this is quite true, but it does not support the relativist conclusion. The choices that we make about which changes are *reasonable* are not arbitrary ones, even if there can be no algorithm for making scientific judgements about them. On the contrary, we can make judgements on the basis of defensible rules of thumb which allow us to preserve the more robust parts of our bodies of knowledge, to eschew *ad hoc* modifications, and to maximise the empirical content of our theories.

'Data are not pure, but theory-laden.' This argument, too, rests on a true premise, and, indeed, we cannot state or describe our observations without using some theoretical terms. But, again, the relativist conclusion does not follow. The key problem is to design research on the basis of theories which are more dependable than the ones we are trying to test.

'Scientific theories are constantly being replaced, and the ones we have now will look quaint in a century's time.' Again, this is true but does not provide much succour for relativism, because it does not suffice to show that, over time, there is no improvement in our capacity to make scientific judgements about which theories to accept as being the best available at the time.

The main point to take from these arguments is that we do not have to adopt a relativist position as the best, or only available, response to the deficiencies of positivism. And although it is still common in social science writing to discuss these arguments as if they are mainstream ones – as they were in the 1970s in the work of writers such as Kuhn (1970), Lakatos (1970) or Feyerabend (1975) – philosophers of science moved on long ago, and today interest in relativism among philosophers has waned considerably.

A note about related terms: 'constructivism' and 'constructionism'

Before leaving relativism, we should make some points about 'constructivism' and 'constructionism', to which it is – mistakenly – often assumed to be closely related.

Constructivism and constructionism come in many forms which it is unnecessary to discuss in detail here. But what they have in common is that they both emphasise – what is undoubtedly true – that people understand the issues they face in ways that are influenced by biases, frames, theories, accounts, narratives and conceptual frameworks of various kinds, and that these construals of their situation and their experiences have important consequences for how they act and organise.

It used to be the case that 'constructivism' referred to an *empirical* argument in psychology – going back to the work of the Soviet theorist Vygotsky – about how *individuals* develop understandings, misunderstandings and ways of framing issues. But, confusingly for those who would have reserved the term 'constructivism' for theories of individual learning, it has come to be widely used to refer to constructionist accounts of social institutions and practices.

'Constructionism' was and is understood to be the study of the *social* interactions that led to *shared* understandings – or indeed conflicting ones – and to the development of practices based on those understandings. Thus institutions – such as money, constitutions, rights and occupational statuses – all function in the way they do because enough people accept them, and they cease to function – or at least function in the same ways – when enough people cease to accept them. So institutions are human inventions and do not last for ever, as is shown by the substitution of electronic money for paper and metal tokens, or by the abolition of inherited titles in the French and Russian Revolutions (cf. Searle, 1995, 2009; Berger and Luckmann, 1991 [1966]).

But constructionists recognise, too, that so long as they last, these institutions are social facts. They constrain our actions. They subject us to disciplines when we violate them. This view is quite different from relativism, which holds that all categories in social theory are merely conventions and that none can be regarded

as grounded in determinate facts about the world. To add to the confusion, relativism is sometimes called 'strong social constructionism', but we hope that you can see that it takes a quite different position from constructionism on whether facts about social life exist independently of the way we claim to know about them.

It follows that researchers who are interested in the *causal processes* by which such construals or framings cause action and social organisation are not required to hold relativistic views about the status of their explanations, and would probably find it intellectually uncomfortable to do so. This is simply another way of putting the obvious fact that in the social science we study meanings, but, of course, some of us also want to explain why people develop the meanings they do and how these meanings, in turn, affect their actions and social organisation.

This is not to say that some constructionists are *also* – or at least write in ways that suggest that they could claim to be – relativists. But this is a distinct philosophical step over and above the constructionist emphasis on the importance of framing. We shall not here enter the debate about whether those constructivists are *consistently* relativistic, but in Chapters 15 and 16, you will find some reasons to doubt that they will turn out to be so.

Ideational constructivism

Some theories that fly under the constructionist flag put forward causal explanations that take a further step, one that goes beyond an emphasis on the human invention of shared social practices and the dependence of these practices on our common acceptance of them.

It is one thing, for example, for political scientists or theorists of international relations to claim that the conventions governing exchanges between diplomats, or the conduct of business in the civil service, persist only because governments, civil servants and diplomats accept them. It is quite another thing for theorists to claim – as students of the history of British public administration have done (e.g. O'Toole, 2006) – that the practices, ethics and organisation of the British civil service were shaped directly by a particular set of ethical ideas developed in the late nineteenth century about the role of the state and the moral responsibilities of its agents. It would be quite consistent with the first claim to say that ideas and beliefs *follow* rather than precede institutions, but it might be more accurate to describe the second claim – that practices are created by the prior existence of specific ideas and beliefs – as 'ideational constructionism' to distinguish more clearly what is specific to this form of constructionism.

Ideational constructionism need not be interpreted relativistically: many who use the approach do consider that there are determinate facts on the basis of which statements about people holding particular ideas, and the consequences of their doing so, are true or false.

Positivism and relativism

Arguments about methodology often proceed as if the 'positivist' and 'relativist' positions were diametrically opposed in every respect. This is a mistake. Both positivism and relativism are based on a principle of caution about making inferences from data about observable facts to make claims about the substantive existence of determinate but unobservable facts about causal processes. Both positions claim that we should place confidence only in observations, and both assert that we should hesitate to draw any inferences beyond those which are about things that can be observed directly.

Positivism and relativism differ fundamentally only on one question. That is, the issue of whether scientific judgements about which theories to accept can be the outcome of independent reasoning, or whether they are essentially matters of convention that are influenced by the prevailing paradigms of the day.

Relativism and the practice of social science research

Just as there are many social scientists who confuse positivism with other philosophical positions, so both critics and advocates of relativism also confuse it with particular practices in social science research.

First, doing interpretive social science does not imply relativism. Most interpretive research is done by social scientists who are far from being relativists. Anthropologists have practised interpretive inference for decades without being relativists. Many have long argued that their interpretations of the mental worlds of those they study rest upon determinate facts about the way those people think, believe, hope and classify their worlds.

Interpretation, as we shall see in Chapters 15 and 16, is one product of social science research – along with description and explanation – and sometimes it is a precursor to explanation. There are many kinds of research questions that require interpretive work. Relativism – like positivism – is simply an argument about the status of the claims made in interpretations, and not the most convincing one. Nor is the converse position true, either. Relativists in social science are not required to eschew explanation and to confine themselves to interpretive research designs.

Second, just as there is no necessary relationship between positivism and the use of quantitative data, so there is no necessary relationship between relativism and the use of qualitative data. Most qualitative researchers do not take a relativistic view of the status of their explanations and interpretations, and there are relativists who use statistics. Neither group is being inconsistent. There is also no particular relationship between constructivism and qualitative research. Some of the ways that people construe and frame their experience lend themselves to quantitative survey research; others do not. Finally, for the record, interpretive work does not have to be qualitative, either. Later in the book we shall look at a study by a famous sociologist which offers an interpretive inference from quantitative data.

Realism

The argument thus far has led us to suggest some good, as opposed to the many bad, reasons for rejecting positivism, and we have discovered some good reasons for rejecting relativism too. The 'last position standing' in this chapter is realism, and so you probably expect to find that the arguments presented here turn out to support it. Indeed, this is the case, but there remain some important nuances to take into account.

For the last thirty years or so, more attention in the philosophy of science has probably been given to varieties of what is called 'realism' than to positivism or relativism. Realism, at its simplest, involves the claim that there are facts which are not directly observable about social phenomena – such as interests, preferences, institutions, bonds, norms, opportunities and constraints – and that, to make claims about these facts, we must make inferences from things that we can observe. Furthermore, realists claim that these inferences are true or false, depending on how well they conform with those facts.

A realist account of market reduction, for example, might depend on making inferences about the nature, and perhaps the volume, of the trade in stolen goods, from, say, interviews with thieves and fences, or from data in the British Crime Survey about respondents' experiences of purchasing stolen goods. Indeed, this is what Sutton and his colleagues did in the example cited in Chapter 3.

So a realist account of social science argues that we develop our understanding of such unobservable things by careful, progressive construction and empirical examination of testable theories and models. The following, simple, definition of scientific realism comprises two parts. The first refers to claims about what is going on in the social world – its *ontology*, to use the technical term. The second part adds the 'science' in 'scientific realism', and makes claims about the status of knowledge, or its *epistemology*.

Realistic ontology

Ontology is that sub-discipline of philosophy which is concerned with the question of what exists, and what status we are ascribing, especially to unobservable, abstract things, when we speak of them as existing, being 'real', and causing states-of-affairs or functioning in some way:

> Realism in ontology is the doctrine that there are determinate facts about unobservable processes, on the basis of which well-specified social scientific models and, thus, social science theories, are either true or false.

This definition contrasts realism sharply with both positivism and relativism. Thus, in the market reduction example above, realism departs from both positions in claiming that it is possible to make knowledge claims about things like trade or incentives that cannot be directly observed. By contrast, relativism denies that we can make ontological claims at all: its epistemology effectively precludes any ontological claims. Positivism argues that, because we should confine ourselves to predicting variance, we do not need to make ontological claims about unobservable things.

Note, too, that the realist ontological claim does *not* say that the 'basis' for judging the truth of theories and models is their 'correspondence' with the underlying reality. Those people who say that realism implies a 'correspondence theory of truth' are, strictly, mistaken. All that realism demands is that claims about truth must be rooted in more than convention or in the practical utility of our theories; they must be true or false, depending on the facts, which are determinate. 'Correspondence' might turn out to be the wrong way to understand that dependence on determinate facts. Realists recognise that our knowledge, at any one time, can never be more than an approximation to truth, because so much of it escapes direct observation, and because our grasp of it depends on the inferences we can make from what are bound to be partial observations. So we have no full access to 'truth', although we hope steadily to make advances in our approximations to it.

Two key questions are prompted by this discussion. The first is whether we are making *warranted* inferences from the facts we have observed to the claims we make about entities we cannot observe; and the second is whether we are using the most appropriate *concepts* to capture what is important about these unobservable and often abstract entities. These are, of course, questions about knowledge, but we shall discuss practical aspects of selecting appropriate concepts in Chapter 9 and procedures for inference in Chapters 13, 14 and 16.

Realist epistemology

Epistemology is the sub-discipline of philosophy concerned with the truth status of knowledge that can be achieved either by observation or by inference. We describe the theory which is principally under examination or test in any single piece of research as the 'focal theory' for that study.

Realism in epistemology is the doctrine that:

1. Scientific judgements about which focal theories to accept, and what trade-offs between such virtues as parsimony, goodness of fit, generality, causality and so on to draw, can be made on the basis of good reasons, based largely on valid inference from findings produced by sound application of correct methods using well-formed categories and evidenced propositions.
2. These scientific judgements rest on accepting, for the particular study, a network of other theories, which can be tested in other studies for which each of them would become focal theory.

3 Scientific judgements make inferences from findings to claims about determinate facts, on the basis of which a focal theory is true or false.
4 Predictions of determinate facts do not exhaust the content of the theories or models: they may begin with logical deductions from axioms; they also contain conceptual definitions and measurement statements.
5 The good reasons which can be provided for scientific judgements are not confined to the manner in which they support or undermine theories – important as this consideration certainly is – but may also involve theoretical and conceptual development.

As we saw above, making satisfying inferences depends on being able to capture what is important about determinate facts, a process that depends on the organising work supplied by concepts and the theories from which concepts are derived. But this dependence does not commit realists to defending the *present* body of theories and concepts as if they 'correctly' and permanently reflect some kind of fixed truth. A key tenet of realism, rather, is its claim that all empirical knowledge is bound to be provisional, as are all the theories and conceptual frameworks with which we organise our understanding of what cannot be directly observed. It follows that developing knowledge depends not only on developing more complete and accurate empirical accounts, but also on developing and refining the concepts and theories in which we capture them. Indeed, for realists, the two endeavours are inseparable.

Realism – but about what?

This doctrine captures what is central to realist ambitions, but what exactly does the task of theoretical and conceptual development consist in? This question can be discussed at three levels.

The first level of argument has to do with *particular empirical claims*. When we are trying to determine whether, for example, reducing the market for stolen goods does reduce the level of acquisitive theft – or whether there is, really, an increase in the number of rough sleepers in a particular town – then the methodological issue is the extent to which the cumulation of empirical findings is converging on some theoretical consensus that appears to be well founded.

The second level of argument has to do with whether or not the theories we use to design our research instruments, measures and codes really do allow us to claim an appropriate status for our findings. We saw the importance of this issue in Chapter 1, and need not discuss it further here.

The third level of argument concerns the models, theories and paradigms we use to organise knowledge as it develops. A key issue here is whether we should advance claims of truth about statements found only in models, or also in theories, or indeed also in paradigms or frameworks.

Consider again the example about the motivations of top civil servants that we explored in Example 2.5. We saw that it was probably easier to compare the two

theories discussed in this example, once we moved from the level of the theory to the level of the model where we could work empirically with the specific concept of 'optimisation under constraint'. And many social scientists would be happy to be realists in working with a well-specified concept of that kind, but would baulk at accepting either theory in its entirety, let alone buying into all the concepts involved.

Another reason offered for restricting realism to models begins from the fact that we can only claim plausibly to have falsified theories when we have exhausted all the possible models that we can derive from them. We saw in the previous chapter that falsifying *theories* is a residual and indirect business. Some people argue that, therefore, we should restrict the scope of realism, at least as an *epistemological* position, to the level of models.

Why it is difficult to restrict realism to the level of models

The problem with this proposition will become clear if we think further about the budget maximisation and bureau-shaping theories, and also about the earlier example in Chapter 2, which used rational choice theory to analyse Khrushchev's decisions during the Cuban Missile Crisis of 1962. Statements in these theories use many of the same terms as those used in the models that can be derived from them. Conversely, models do not consist exclusively in purely operational, observational terms: they are not always free from statements about abstract entities such as preferences or intentions. So, restricting the scope of claims for truth to models seems rather arbitrary.

More fundamentally, it is typically the theory rather than the model which supplies the *semantic content* for a satisfying explanation. For example, it would be difficult to understand how market reduction or situational crime reduction theories are supposed to work, without a theory of causation. Being realist only at the level of the model would, therefore, leave us unable to make claims about the very things which make explanations interesting and satisfying. And that would not be a realism worth the making of serious ontological and epistemological claims.

EXERCISE 4.1. PARADIGMS WITHOUT HOMES TO GO TO

Go back to the work you did in Exercises 2.2 and 3.4 on verifying and explaining the possible increase in rough sleeping.

Say whether you think your conclusions fell within the positivist, relativist or realist understanding of social science, and why.

How would researchers from the other two positions approach the same problem? What would be the main differences in the knowledge claims made for the results of their research?

Conclusion

We have argued through the last four chapters that, to be a complete social scientist, the researcher needs to be able to combine design choices made at three levels, namely:

- *method* – the selection of appropriate types of data and of appropriate strategies for collecting and analysing those data;
- *methodology* – the theories that explain why the methods used provide warrant for inferences from the data to be collected and analysed, to the kind of explanation or interpretation required by the research questions; and
- *philosophy of social science* – the account of the status of the explanations and interpretations, and the justifications presented for granting them that status.

Each level has its place in a research design, and each supports the other two. Unfortunately, many textbooks either run them together or entirely miss out either methodology or philosophy. For example, it is still far too common to read research proposals that deal with methods and with philosophy, and assume that methodology is covered simply by combining them. However, we hope that you can now see that making a philosophical argument – and still less stating a philosophical position – does not justify, let alone require, *any* particular choice of research design.

On the other hand, philosophical issues are of real practical importance for methodology. They provide us with ways to *understand* what it is that we are claiming, when we use particular research designs to claim warrant for particular inferences. So they force social scientists to be aware of, and to think through, the full range of judgements they make when they argue for the acceptance and rejection of theories, and when they make trade-offs between rival virtues of a good explanation or a good interpretation. And understanding some philosophy of social science equips us to answer critics who challenge us to show why they should take seriously the description, explanation, interpretation we produce by doing our research.

In this chapter, we have argued that, in our view, the most convincing response to such issues takes the form of 'provisional realism' and that this position requires us to be clear about the models and theories that are at stake in our research.

Basic reading

For a series of accessible, fictional dialogues between stock positions, see Laudan L, 1990, *Science and relativism: some key controversies in the philosophy of social science*, Chicago: University of Chicago Press.

For an accessible guide to realism in social sciences, see Pawson R and Tilley N, 1997, *Realistic evaluation*, London: Sage.

Advanced reading

Kincaid H, 1996, *Philosophical foundations of the social sciences: analysing controversies in social research*, Cambridge: Cambridge University Press.

Those interested in realism might look at Manicas PT, 2006, *A realist philosophy of social science*, Cambridge: Cambridge University Press, and Sayer A, 1999, *Realism in social science*, London: Sage. (Be aware that Manicas combines his commitment to realism with a substantive commitment to methodological individualism and rational choice, and that Sayer holds a particular philosophical position known as 'critical realism' which goes far beyond the present argument.)

A well-known statement of 'pragmatism', regarded by many as effectively relativistic, is offered by Rorty R, 1979, *Philosophy and the mirror of nature*, Princeton, NJ: Princeton University Press.

A famous relativist statement is by Feyerabend P, 1993 [1975], *Against method*, London: Verso, although he qualified his views in later writings.