Comparative design thus presents the researcher with considerable challenges, especially when different countries are being compared. The researcher must select a theoretical problem that is best illuminated by comparative research: for example, why women are included in the political elite in Denmark but excluded in Britain and France (Siim, 2000). Relevant and equivalent data should then be collected and hypotheses tested, such as the impact of the electoral system or diverse traditions of citizenship in the various countries, and appropriate conclusions drawn. Comparative analysis sharpens our understanding of the context in which theoretical problems occur and enables causal inferences to be drawn. However, as comparative analysis usually involves only a relatively limited number of cases, caution has to be maintained about the levels of generalization that can be made.

Conclusion

The planning and execution of a research project are critical to its success. Planning involves determining the objectives of the research, developing research questions, transforming these questions into hypotheses and deciding on the appropriate research design to test the hypotheses and to convince a sceptical audience that the evidence is valid and reliable and that the conclusions drawn from the analysis are accurate. The choice of research design is fundamental to this process as it will determine how the evidence will be generated and analysed. The better the design, the clearer will be the evidence of cause and effect and the more likely that the findings and explanations will be accepted.

This chapter has also discussed the process of research and suggested that there are two ways of describing this, namely the linear model and the research labyrinth. The linear model suggests that the research process is smooth and ordered, moving from theoretical speculations to the collection of data, the analysis of the findings and finally the publication of the results. The research labyrinth emphasizes the complexity and pitfalls involved in research, especially the false starts, the need for inspiration, the difficult negotiations with funders and subjects, ethical dilemmas and competition and conflict with colleagues.

The issues raised in this chapter are discussed in greater detail in the following chapters, beginning with the next chapter on the comparative method.

Chapter 3

Comparative Methods

To make comparisons is a natural way of putting information in a context where it can be assessed and interpreted. This is especially true when we encounter new information about some issue and begin to integrate it with previous knowledge. For example, we might know that in Germany, Denmark and the Netherlands the political executive is based in the legislature and can only survive with support from the legislature; this is what makes these countries so-called parliamentary political systems. Then, turning to France, it is striking that its political executive is linked to the legislature in quite a different way: the French president is a directly elected head of government (and head of state) whose survival does not depend on the support of the legislature. To understand fully the significant political consequences of the French semi-presidential political system, it is quite instinctive to compare it with the already familiar parliamentary systems, and in this way to integrate old and new knowledge. By comparing parliamentary and presidential systems, it is possible to acquire a greater understanding of each; as Rudyard Kipling said in his poem, 'The English Flag': 'what should they know of England who only England know?' Research from virtually all political science research traditions and sub-fields of study can (and does) fit under the label 'comparative'. There are examples of quantitative and qualitative comparative research, large-n, small-n and even single-n comparative research, and inductive and deductive comparative research, spanning every conceivable substantive topic (see Rogowski, 1993, for an extensive survey of comparative research). The goal of comparative research, as set out in an influential book from 1970, is to be able to remove proper names, and to reason instead in terms of variables (Przeworski and Teune, 1970), meaning that ultimately the uniqueness of each case itself (such as US states Florida, Texas, Louisiana, Kentucky, the Carolinas ...) is less important than the case understood as a combination of values on a number of specific variables when it comes to generating general theories of politics.

In order to gauge what the comparative method can and cannot achieve, it will be placed alongside the experimental and statistical methods in the first section of this chapter. The subsequent section details some basic comparative research designs, while the advantages and disadvantages of this type of research are spelt out in the next two sections. Case selection is the topic of the final section. Case selection merits some special attention because the quality of a piece of comparative research depends very largely on what cases are included. Applying the comparative method in a rigorous attempt to test some proposition is only possible if the cases are comparable. This is the sort of comparison that this chapter addresses. In contrast, comparisons of cases as a means of persuasion without a clearly thought out case selection and analysis are not, strictly speaking, using the comparative method.

Comparative political science: substance and method

Calling oneself a comparativist, or saying that one is studying comparative political science, can mean at least three different things (Mair, 1996, pp. 309–10). First, comparative political science can refer to the study of foreign countries. This type of comparative political science often consists of single-country studies (such as Italian politics, New Zealand politics, Canadian politics, and so on) which can be considered as implicitly comparative if they draw on more widely applied theories or models of politics. If so, then case studies can be seen as part of a larger, comparative body of research. Nevertheless, this type of comparative research tends to be more focused on collecting and presenting facts about a single case than on making a sustained contribution to the development of theories and hypotheses. Second, there is a significant body of explicitly comparative research: that is, research covering more than one case. Systematic comparisons of some aspect of the political systems of two or more countries often provide the empirical basis for building and refining general political science theories. The third meaning of comparative political science refers to the methods used to carry out comparative research, and this is the focus of the remainder of this chapter.

In a seminal article from 1971, Arend Lijphart placed the comparative method, which he defined as 'a broad-gauge, general method, not a narrow, specialized technique' (1971, p. 683), alongside experimental and statistical methods in political science

(see also Ragin, 1987, pp. 61–4, on the relationship between comparative and statistical control). The article posited that these three methods aim to establish scientific explanations consisting of two elements: first, a specified empirical relationship between two or more variables, and second, the potential effects of all other variables on that relationship are held constant (that is, they are somehow eliminated or controlled for in the research design). It is, strictly speaking, not necessary to include the second element since it follows automatically from the first element: unless 'all other variables' are controlled for, it would not be possible to specify an empirical relationship between two or more variables. The experimental method normally has a better ability to generate this type of explanation than the statistic and comparative methods. However, in political science such explanations are rare because the research environment is impossible to control fully. Take the example of some political scientists who are studying an election because they want to learn about the impact of the particular electoral system used in that election. In order to be certain that it was the electoral system and not something else that had a particular consequence, it would be necessary to turn back time, change the electoral system in some major respect, and then rerun the election. The observed differences between the first and second time the election took place could then safely be put down to the electoral system. The impossibility of controlling the research environment leads to what King, Keohane and Verba (1994) call the fundamental problem of causal inference, and it is a very fundamental problem because without experimental control it is impossible to say with complete certainty that one's conclusions are correct (causal inference is discussed in more detail in Chapter 6).

The experimental method is the best known way of establishing explanations that fulfil the criteria of, first, having a specified empirical relationship between two or more variables, and, second, controlling for possible effects of other variables. When experiments are not possible, the statistical method is often available to political scientists. This method 'entails the conceptual (mathematical) manipulation of empirically observed data – which cannot be manipulated situationally as in experimental design – in order to discover controlled relationships among variables' (Lijphart, 1971, p. 684). That is, when it is not possible to manipulate a research setting as in an experiment, then the statistical method can be a useful way to assess the relationship between two or more variables. The statistical method does not have as strong a control

function as the experimental method, and thus statistically established empirical relationships cannot be viewed with the same level of confidence. However, there are techniques for assessing how much confidence one can have in a given statistical relationship (see Chapter 6).

The comparative method is about observing and comparing carefully selected cases on the basis of some stimulus being absent or present. The comparative method operates on the same logic as the experimental method, which has been described as 'nothing but the comparative method where the cases to be compared are produced to order and under controlled conditions' (Parsons, 1949, p. 743). For example, we might like to know whether and how decreases in trade union membership (a phenomenon that occurred in many western European countries in the last decades of the twentieth century) affect class-based voting. In order to assess the impact of trade union membership on class-based voting, then, it is useful to compare levels of trade union membership and class-based voting in different countries (or over time). However, since the ability to control the political environment is so limited, these comparisons do not reach experimental standards. The conclusions are drawn from comparisons, not experiments. As a consequence, the comparative method (and the statistical method) makes claims about empirically observed relationships without rigorous controls for other variables, and here the comparative method is even weaker than the statistical method. Quantitative comparative research bridges the two research methods. If a comparison involved enough cases for statistical control to be possible, then, as Lijphart points out (1971, p. 684), there would be no real difference between the statistical and comparative methods. In fact, the only thing that prevents (much) comparative analysis from being statistical is the number of cases included. The number of comparable cases available seldom satisfies the assumptions of statistical control techniques.

Political science proceeds by imposing some sort of order on processes, events and phenomena that do not easily conform to any sort of order. The real difficulty of political science is to make 'convincing statements about the causation of political phenomena, given the complexity of interactions among the whole range of social phenomena and the number of external sources of variance' (Peters, 1998, p. 28). Researching politics can be seen as a process of shifting focus from the level of *particular* pieces of information to the *general* level of theory and hypotheses. If, for instance, we are interested in the fairly recent electoral successes of

far-right parties across western Europe, we might decide to study the rise of the Austrian Freedom Party and the Vlaams Blok in Belgium. The reason to study these *particular* cases is to formulate *general* conclusions about far-right parties. A general theory of far-right parties is the 'order' we want to impose on flourishing far-right parties.

The comparative method can help alleviate but rarely, if ever, resolve the fundamental problem of causal inference inasmuch as it can emulate the experimental method. A chemist might, for example, pour a measure of a chemical into two bowls, and then add a second chemical to one of the bowls and observe what happens. The chemist can safely attribute the reaction between the two chemicals to the addition of the second chemical, because in the bowl containing only one chemical no reaction at all took place. In a similar vein, in political science the comparative method can facilitate the isolation of certain factors of interest, if the cases that are selected for study have been carefully chosen with this in mind. This is the logic of making comparisons. Exploiting the logic of comparison means choosing cases that isolate one or a small number of factors that appear relevant in producing a particular political outcome.

Designing comparative research

There are two basic comparative research designs: most similar research designs, and most different research designs. The most important aspect of formulating either a 'most similar' or a 'most different' research design is to select cases that make it possible to conclude something interesting about one's research question. Comparative case selection should take place on the basis of three selection principles: cases should 'maximise experimental variance, minimise error variance, and control extraneous variance' (Peters, 1998, p. 31). In less technical language, by choosing cases that isolate the effect of the factor or factors that are being investigated, we exploit the logic of comparison as much as possible (we 'maximise experimental variance'). To do so, it is necessary to choose cases that are representative, not one-off or unusual ('minimise error variance') and to minimize the effect of all other factors ('control extraneous variance').

Variables (or factors) can be divided into three categories: dependent variables (usually denoted y), independent variables (x),

and other variables, which we can call spurious or intervening variables (these are in fact also x variables, or independent variables, but ones that compete with the theory we want to test rather than being 'part of' that theory). Dependent variables are the phenomena that we want to explain in the research. Independent variables are the things we suspect influence the dependent variable. Everything else (that is, everything that makes up the social, economic and political context and backdrop of the dependent and independent variables) fits into the third category. Such variables might be spurious (that is, falsely appear to have some bearing on the relationship between the dependent and independent variables) or intervening (that is, actually having some bearing on the relationship between the dependent and independent variables).

The work of J. S. Mill, especially his so-called method of difference, lays the foundations for these two types of comparative research (although many other writers since have of course continued to develop the thoughts and arguments Mill formulated in 1843 in his *System of Logic*). In brief, what we can take from the method of difference is that if two units or cases are exactly alike in every respect (a ceteris paribus, 'all other things being equal', assumption), except that in case A y occurs and in case B y does not occur, and except that in case A x was present and in case B x was absent, then we may conclude that x caused y, if we also have some theoretical basis for believing that there is a causal process, not merely a correlation (that is, we must be able to present some plausible account as to how x 'produces' y, and does not merely seem to occur at the same time and in the same place as y (Shively, 2005, p. 75)). This logic is nevertheless easier said than done to implement in a research project, mainly because the ceteris paribus assumption almost never holds: two cases are never exactly alike except for the condition we want to investigate. One way of getting around this that might seem reasonable would be to observe the same case twice (say A), once with x and once without. However, this is almost as unrealistic as the ceteris paribus assumption, and has been labelled the fundamental problem of inference (Holland, 1986, p. 947).

To return to the idea of 'most similar' and 'most different' research designs introduced at the beginning of this section, one simple way of understanding what they are, and how they differ from each other, is to keep the three categories of variables outlined above in mind (dependent, independent, spurious/intervening). The first point to note is that the dependent variable is irrelevant

at the research design stage, in the sense that cases should not be included or excluded on the basis of their values on the dependent variable. This can be called 'the principle of not selecting on the dependent variable', and there is an example of this in the final section of this chapter (see also Geddes, 1990, pp. 134–41). The second point to note applies to most similar research designs (see Box 3.1). They compare two or more cases that are as different as possible in terms of the independent variable(s) and as similar as possible on all the spurious and intervening variables ('backdrop variables'). In contrast, a most different research design compares two or more cases that are as similar as possible in terms of the two or more cases that are as similar as possible in terms of the independent variable(s) and as different as possible on all the spurious and intervening variables.

The logic behind most similar research designs is that by a basic process of elimination all the variables in the 'other' category can be ruled out of the research. If they do have an effect on the dependent variable, they have the same effect on the dependent variable across all cases. This leaves the independent variable, which has a differential effect on the dependent variable in the two or more cases (which is the reason why the cases were selected in the first place). This means that any observed differences between cases with respect to the dependent variable can be associated with the only variable that makes the cases different: the independent variable. In brief, 'the more circumstances the selected cases have in common, the easier it is to locate the variables that do differ and which may thus be considered as the first candidates for investigation as causal or explanatory variables' (Castles, cited in Pennings, Keman and Kleinnijenhuis, 1999, p. 12).

For example, the well-known hypothesis about economic voting holds that voters judge government performance on the basis of the state of the economy and vote to return or replace the government on the basis of this performance indicator (Kinder and Kiewiet, 1979; Lewis-Beck, 1988; Lewis-Beck and Lockerbie, 1989; Markus, 1993; Gavin and Sanders, 1997). Here, 'the state of the economy' is the independent variable, which is hypothesized to generate a specified political outcome on the dependent variable, 'government popularity'.

The logic of comparison in a most similar research design is consequently to isolate the effect of an independent variable by controlling for the effects of spurious and intervening variables. For an illustration of some of the difficulties of case selection, see Box 3.2.

BOX 3.1 **Research design: most similar electoral systems and gender balances in Parliament (I)**

This type of research design compares two or more cases that have as much in common as possible, except the independent variable. Trying to answer the research question 'Do electoral systems that are more proportional generate a more even gender balance in Parliament?', these variables might be deemed relevant:

Dependent variable: Gender balance in Parliament (%)
Independent variable: Electoral system
Spurious/intervening: Gender balance in the workforce (%)
                          Equality legislation
                          Political culture
                          Parties' candidate selection procedures

The cases in a most similar research design should have different electoral systems but be as similar as possible in all other respects. On this basis it is reasonable to compare the UK to any of the West European proportional representation (PR) countries. However, an even better idea would be to break down the UK case into several cases. General elections in the UK (Westminster elections) use the single member plurality (SMP) system. However, elections to the UK's three devolved assemblies take place by other electoral systems: the Northern Ireland Assembly is elected by proportional representation by the single transferable vote (PR-STV), while the Scottish Parliament and the Welsh Assembly are elected by a mixed electoral system (for example, mixing elements of SMP and PR).

By breaking the UK case into many cases that can be compared with each other, many (if not all) spurious and intervening variables are held constant: the electoral system changes depending on what body is being elected, but political culture, proportions of women in the workforce, and so on, remain relatively unchanged. To test the effect of electoral systems on gender balance, the following four cases could be compared:

- Westminster elections
- Northern Ireland Assembly elections
- Scottish Parliament elections
- Welsh Assembly elections.

If the proportionality of an electoral system really impacts on gender balance, then election results should show the following (if they do not, then there are intervening variables that have not been, but should be, taken into account):

- the Northern Ireland Assembly is the most gender balanced Assembly in the UK
- the Scottish Parliament and the Welsh Assembly have intermediate levels of gender balance
- Westminster is the least gender balanced Assembly in the UK
- within Northern Ireland, Scotland and Wales, the gender balance among returned representatives to the devolved assembly should be greater than among returned representatives to Westminster.

Sources: Farrell (2001); Gallagher, Laver and Mair (2001).

The third point to note is that most different research designs use the logic of comparison in a reverse manner (see Box 3.3). Case selection proceeds on the basis that the cases must not differ from each other with respect to the independent variable, but must be as different as possible from each other in terms of spurious and intervening variables. The logic behind this type of research design is that if the independent variable has an effect on the dependent variable, then it should have the same effect despite the cases being so different when it comes to the spurious and intervening variables. To continue with the example of economic voting, a most different research design would be based on the assumption that if the 'state of the economy' strongly influences 'government popularity', then we would expect to see governments win elections in cases where the economy is doing well, and lose elections in cases with poor economic performance, regardless of all other circumstances. Francis Castles has summarized the 'most different' approach as one that involves 'a comparison on the basis of dissimilarity in as many respects as possible in the hope that after all the differing circumstances have been discounted as explanations, there will remain one alone in which all the instances agree' (cited in Pennings, Keman and Kleinnijenhuis, 1999, p. 12). If the economic voting hypothesis were true, then the variable that remained would be 'the state of the economy'. Comparing two or more cases with similar economic performance in the year or so prior to an election, but which were different in terms of party systems, electoral systems,

single party/coalition government (and so on), should ideally establish whether 'the state of the economy' has a constant effect on 'government popularity'.

Both the 'most different' and the 'most similar' research designs can 'result in the confirmation of theoretical statements' (Przeworski and Teune, 1970, p. 35). In theory both types of comparative research design make the assumption that it is possible to reach the level of experimental control. In actual political science research designs this is highly unusual, but the comparative method remains one of the most widely used forms of research because it is often the best available alternative to experimental research.

Means of overcoming the problems of comparative research – at least partly – are sometimes available. For one, case selection plays a crucial role in determining the extent to which they bedevil our research. This is discussed in more detail in the chapter's final section. Sometimes it is possible to replace (if imperfectly) a strictly experimental setting with statistical analysis, which at least gives us a handle on how probable it is that a purported relationship between some variables exists, but this lies beyond the scope of this chapter.

BOX 3.2 The difficulties of case selection: Almond and Verba's *The Civic Culture*

Gabriel Almond and Sidney Verba's *The Civic Culture: Political Attitudes and Democracy in Five Nations* (1963; later complemented with *The Civic Culture Revisited* first published in 1980) remains a landmark text in the study of political culture. The authors' hypothesis was that democracy is more stable in countries with a particular kind of political culture, which they labelled a 'civic culture'. (Note that the cases were selected on the dependent variable.)

Almond and Verba's criteria for case selection:

- the selected cases had to be democratic countries
- the selected cases had to be countries with different historical experiences with respect to democracy.

Selected cases:

- USA and Britain: examples of stable democracies
- Germany: had a 'broken' democratic record
- Italy and Mexico: had less developed societies whose political systems were in transition to democracy.

Points to note about this case selection:
- selection on the dependent variable: this is suitable for a small-n study
- failure to include any non-democracies: Almond and Verba forwent the possibility of examining their hypothesis as rigorously as possible. If their hypothesis about political culture and stable democracy were true, then the political cultures of non-democracies would presumably be in some key respect fundamentally different from the political cultures of stable democracies. Due to their case selection, however, Almond and Verba were not able to answer this key question.

BOX 3.3 Research design: most different electoral systems and gender balances in Parliament (II)

This type of research design compares two or more cases that are as different as possible except on the independent variable. For simplicity's sake, let's stay with the research question from Box 3.1: 'Is there a connection between a country's electoral system and the gender balance in Parliament?' To recap, all the relevant variables are as follows:

Dependent variable:	Gender balance in Parliament (%)
Independent variable:	Electoral system
Other variables:	Gender balance in the workforce (%)
	Equality legislation
	Political culture
	Parties' candidate selection procedures

In a most different research design the independent variable should be the same across all cases. Referring back to Box 3.1, elections to the Scottish Parliament and Welsh Assembly might seem the ideal cases to compare here, since they use the same mixed electoral system. However, the other requirement of a most different research design is that the cases should be as different as possible in terms of all the other variables. Scotland and Wales are clearly not wisely selected cases from this perspective.

The 'devil is in the detail' when it comes to electoral systems: there is almost endless variation between electoral systems in different countries, even ones that can be broadly categorized as belonging to the same general category of electoral systems (for example, although both France and Australia have majority systems, France operates a

two ballot system, while in Australia the system has a single round of preferential voting where voters are obliged to rank all candidates on the ballot paper, or else their vote is declared invalid).

Since there is so much variety it is often necessary to select the 'least bad' cases: for example, two countries with List-PR electoral systems even if it is well known that they differ in the detail. Sweden and Portugal, for example, have List-PR electoral systems, but have different constituency-level seats allocation formulae; in Sweden there is a higher tier but not in Portugal; and in Sweden voters can choose candidates within a party on the party list, but Portugese voters are not able to do this.

A Swedish–Portugese comparison in a most different research design would also be useful in that these countries are quite different generally in terms of gender roles. Therefore, if it is really the electoral system that determines the gender balance in Parliament, then the Swedish and Portugese Parliaments (and all other Parliaments elected by List-PR) would be expected to have similar gender balances despite all the other differences between these countries.

Why compare? The advantages

The primary advantages of the comparative method can be summarized under four headings: it allows us to contextualize knowledge; to improve classifications; to formulate and test hypotheses; and to make predictions (Hague and Harrop, 2007).

We make comparisons to contextualize knowledge almost without thinking about it in everyday life as well as in more formal political analysis, to integrate and make sense of newly acquired knowledge. Even if the primary interest and concern is with one particular case or event, considering it in the context of other, similar cases or events advances our understanding of the one of primary interest. Dogan and Pelassy (1990) add that this enables us to overcome implicit ethnocentrism, in that comparisons force the recognition that not all countries have the same political system as the one with which we might be most familiar. This point applied more broadly tells us that the comparative method advances a heightened, comprehensive awareness of the diversity of the political world.

A related advantage of approaching a research question comparatively is that doing so has the potential to improve the classifications we use to impose some sort of order on the diversity classifications we use to impose some sort of order on the diversity of the political world. Classification is the 'basic type of concept formation ... neither comparison (non-metric ordering) nor measurement proper can take place without it' (Kalleberg, 1966, p. 73): that is, classification is prior to comparison in the sense that typically we choose cases to compare because they belong to a particular classification. Consequently, a comparative analysis is only ever as good as the classification behind it. However, since the research process is a dynamic between theoretical level (that is, the classification) and the empirical level (the measurements we use), the comparison will then feed back into the classification and improve it by refining it on the basis of additional empirical information. The concept of the nation-state is a good example of how comparisons can improve classifications. Since the Treaty of Westphalia (1648) the sovereign nation-state has been the basic political unit in Europe and, later, elsewhere, too. The concept of the nation-state suggests that a nation and its state coincide with each other perfectly, and that within its state a nation exercises its right to self-determination. This is more or less the case with highly homogeneous countries such as Denmark, where the Danish nation is coterminous with the Danish state. However, it is also routinely assumed that, for example, Ireland is a nation-state, although the Irish nation is not coterminous with the Irish state. A significant proportion of the Irish nation in fact live in Northern Ireland which is in the jurisdiction of the UK. The UK itself provides another example of how mistaken the assumption can be that a nation and a state coincide: here, several nations (English, Irish, Scottish and Welsh) live within the same state. Comparing Finland, Denmark, Ireland, the UK and additional nation-states reveals how varied and layered a seemingly straightforward and commonplace concept such as the nation-state is, and points up the need to refine it into a number of classifications of different types of nation-states on the basis of how nation and state relate to each other.

In a similar vein, taking a comparative perspective on a research question also enables hypothesis testing and development. As we have already seen, in investigating a supposed empirical relationship between two or more variables, a comparative research design can test hypotheses through isolating the effect of one variable on another. Inasmuch as doing so throws up new

ideas and possibilities, it can also suggest how the hypothesis might be usefully refined or reformulated. Having a hypothesis about the effect of the electoral system on the nature of the party system (a much-debated matter: see Duverger, 1954; Harrop and Miller, 1984; Lijphart, 1994; Sartori, 1994; Farrell, 2001) to the effect that a plurality electoral system has a tendency to generate two-party systems and proportional representation electoral systems tend to generate multi-party systems, simply studying one country will not reveal enough. It is only by taking a comparative perspective that the differential effect (or lack thereof) on party systems of different electoral systems may be identified. In the process of testing this hypothesis, some cases might seem to offer confirmation whereas other cases might seem contrary to it (and there is of course the question of how to classify cases such as Germany and Italy, whose electoral systems combine elements of plurality and proportional representation). As an example of a country with a plurality electoral system, the domination of two parties (the Conservative and Labour parties) in British politics ostensibly confirms the hypothesis although there are other parties, too. Similarly, the Netherlands, Belgium, the Scandinavian countries, Spain and many others confirm the hypothesized relationship between proportional representation and multi-party systems. In contrast, Malta has proportional representation but Maltese politics is highly dominated by two parties (the National and Labour parties), possibly even more so than British politics. This might be taken to suggest that the original hypothesis might be too simplistic, and that there are other relevant variables that influence the extent to which electoral systems shape party systems.

Finally, the comparative method can enable predictions about politics. If an empirical relationship between two or more variables has been observed in one temporal or spatial setting, then it can be inferred that the same relationship would hold in another temporal or spatial setting. We can take the example of how EU membership has affected the ability of the four traditionally neutral member states – Austria, Finland, Ireland and Sweden – to make independent foreign policy. Since in the post-Cold War era the EU has started to develop its own foreign, security and defence capability and identity, commensurate with its global economic standing, it might reasonably be assumed that these four member states would find their independent policymaking capacity (a minimum requirement of neutrality)

restricted through EU membership. Meanwhile, Switzerland is an additional neutral country standing outside the EU although membership has been on the Swiss political agenda. Concerned Swiss observers might study Austria, Finland, Ireland and Sweden to find out how Swiss neutrality might be affected by EU membership at some future point in time and on that basis make a prediction based on taking a comparative approach.

These advantages of the comparative method make it a very valuable tool in the political scientist's toolkit. However, assuming that any one of these advantages automatically accrues through comparative research is a flawed assumption: comparative research involves a significant risk of turning into a 'wonderful, creative exercise of comparison that ultimately is meaningless' (Peters, 1998, p. 85).

The limits of comparison

Just as the comparative method can have a number of advantages, there are a number of reasons why comparisons can turn out to be meaningless. Most famously, the condition known as 'too many variables, not enough cases' is the reason why experimental control is rarely an option in political science. Additionally, comparative research is affected by two manifestations of the so-called travelling problem: that is, that neither theoretical concepts nor empirical measurements are consistent (they do not 'travel') across temporal and/or spatial settings. This diminishes the possibility of controlling for the effect of variables other than those that are of primary interest. The comparative method also contends with the issue of value-free interpretations, and with the so-called 'Galton's problem'. All these issues are elaborated below.

The 'too many variables, not enough cases' problem of comparison arises because the political world that is the research environment of political science is too rich and varied (that is, it consists of too many variables) for the researcher to be able to find enough cases to control for all the effects of these variables; it thus becomes impossible to isolate the dynamics of the relationship of primary interest (Ragin, 1987, pp. 23–6). As an illustration, consider the case of even the simplest possible hypothesized empirical relationship: that is, a relationship between one independent and one dependent variable. For example, the emergence of a Green party as a political force might be expected

to affect other parties' positions on environmental policy; the hypothesis might be that other parties will develop their own environmental policies to counter the Green party's electoral appeal. In this bivariate relationship 'presence of Green party' is the independent variable and 'other parties developing their environmental policies' is the dependent variable. Examining the relationship between these two variables comparatively requires at a minimum two cases.

Let's suppose a two-case most similar research design was developed, covering the same party system at two points in time: before and after a Green party became a force in electoral politics. If the hypothesis were correct, then the other parties would pay more attention to environmental issues after the arrival of the Green party than before. However, in addition to the two variables involved in this hypothesized relationship ('presence/absence of Green party'; 'environmental policy of other parties'), other variables in the political environment might impact on this relationship. For example, if the Chernobyl nuclear disaster occurred in between the two points in time of measurement, then this rather than competition from the Green party might have focused party minds on the environment. Of course, environmental disasters of that magnitude might have had the consequence of upgrading the salience of environmental policy generally, as well as generating public support for Green parties. In this case the hypothesized relationship between 'presence/absence of Green party' and 'environmental policy of other parties' is at least partially spurious (that is, the two variables are linked to each other through some third, unidentified variable; here, 'environmental disaster'). But there is no way to test how this additional variable is related to the original two variables without additional cases that have all the attributes of the two original cases but which were somehow not exposed to Chernobyl; given the global repercussions of this event, it is difficult to imagine that such a case exists. Accordingly, there are not enough cases to facilitate controls for all possible variables. As a general rule, a research design requires at least one more case than it needs variables.

The so-called travelling problem is not entirely unrelated to the 'too many variables, not enough cases' problem, insofar as both have a bearing on the possibility of isolating a hypothesized empirical relationship between two or more variables from other variables. The first manifestation of the travelling problem is conceptual: does the meaning of a concept stay constant across time and space? The significance of this was discussed above (definitions of nation-states) in terms of how the comparative method can improve classifications, and it follows from that discussion that mistaken assumptions about concepts and their meanings lead to confusion about what it is that is being compared. The second manifestation of the travelling problem concerns empirical measurement, the concern being that even if the meaning of a concept is constant across cases, if it is operationalized differently for different cases, then there is measurement inconsistency.

A well-known example of this is the measurement of the so-called party identification model of voting behaviour (Budge, Crewe and Fairlie, 1976; Harrop and Miller, 1984; Heath and McDonald, 1988; MacKuen, Erikson and Stimson, 1993; Niemi and Jennings, 1993; Norris, 1997; Todal Jenssen, 1999). This model attempts to explain voting behaviour with reference to voters developing an affinity with a party relatively early in life, and then voting for that party for more or less the rest of their lives. Increased volatility in voting behaviour across the western world (a combination of realignment and dealignment) has diminished the perceived value of this explanation of voting behaviour, but the value of the model here is that it was developed with respect to understanding voting behaviour in the USA, where voters are required to register as supporters of a particular party to vote in primary elections. Political scientists have used this registration as evidence of party identification, and compared it against actual voting behaviour in a presidential race. For a long time they found that few people voted against their party identification, and concluded that party identification was a strongly predictive indicator of voting behaviour. In western Europe, however, there are no primary elections and there is thus no need for voters to register as supporters of any party. Consequently, there was no obvious measurement of party identification against which to compare voting behaviour. Researchers attempted to overcome this problem by formulating survey questions that distinguished between which party a voter or survey respondent usually sympathized with most closely, and which party they had voted for in a recent election. While this was a reasonable solution to a measurement problem, it did mean that transatlantic comparisons were based on different measurements of party identification. The measurement did not travel well (see Box 3.4).

BOX 3.4 **Does the measurement travel? Comparative Manifesto Research**

The so-called Comparative Manifesto Research group has used all post-war election manifestos in 25 democracies around the world to develop comparable measurements of party positions (see, for example, Budge, Robertson and Hearl, 1987; Budge and Laver, 1992; Klingemann et al., 1994; Pennings, et al., 1999; Budge et al., 2001; Laver, 2001). They have coded each sentence in every manifesto to a category or categories to which the sentence refers, such as 'regulate capitalism' or 'law and order'.

The measurement

- interparty comparisons: the coded text makes it possible to compare party positions at a given point in time, on one or more issues in one or more countries.
- intraparty comparisons: the coded text makes it possible to track changes in party positions over time, on one or more issues in one or more countries.

Travel issues

1. Cross-sectional (for example, county, region, unit) comparisons: the same word/phrase can have different connotations in different political systems. 'State' means something different in western Europe and the USA. There are many forms of democracy, so when Dutch parties refer to their consociational democracy they refer to something quite different compared with when Austrian parties refer to their, until recently highly corporatist, democracy.
2. Language issues: a subcategory of cross-country comparisons. Although it is of course possible to translate manifestos, the act of translation may subtly alter the connotations and symbolism of specific words and phrases.
3. Cross-time comparisons: over time, the meaning of a word or phrase within a country can change. Some may disappear altogether while others are new additions to the political discourse. When party positions are tracked over time, the meaning and usage of words and phrases used may subtly change.

The question of whether value-free interpretations are possible is not unique to the comparative method, but it can be particularly troublesome here because this type of research frequently requires researchers to consider unfamiliar political systems or phenomena. The difficulty is that the values of the researcher and the values embodied in the political system under observation may lead the researcher to misinterpret the unfamiliar political system. While complete objectivity is probably never possible, it is worthwhile for comparative researchers to be explicit about how their values might influence the way they approach an unfamiliar political system as a case.

Finally, 'Galton's problem' occurs when the expectation is not met that political outcomes are due to processes internal to each case in the research design. If some hypothesized empirical relationship under examination is really the result of an external or even global process, then studying more than one case will not in truth provide a comparative perspective because the cases will not be independent from each other. In this vein, studying economic policy developments in EU member states represents a clear instance of Galton's problem, since the EU's influence on the economic policy of all member states is undeniable, especially in the era of the single currency and in the period preceding its introduction when member states planning to adopt the euro were also required to adhere to the so-called stability pact. Similarly, the fact that many former British colonies have adopted the Westminster model of government belies any assumption that this choice of government in post-colonial states is the result of internal processes. Comparative researchers are, in other words, stuck between Galton's problem and the need to find cases that are comparable, the problem being that cases that are comparable often are comparable precisely because of being affected by some external process; that is, they are similar because of a Galton's problem. Again, faced with this situation a researcher is normally unable to change the historic reality of the cases he or she is interested in, but should definitely be explicit in his/her analysis about what factors and their effects are judged to be internal to a case and what are external, so that the extent of Galton's problem can be assessed by the readership.

Cases: how many, and which?

Comparative political science is usually concerned with some abstract and generic theory, such as the relationship between

class voting and industrialization, to use Ragin's example (1987, pp. 9–12). However, the research question derived from a theory and asked of a number of cases is typically in itself much more historically and socially circumscribed than the theory. This is at least partly because comparative studies often do not include enough cases to allow the research question to be generically formulated. As indicated in Chapter 2, and as Chapter 6 discusses in detail, sampling occurs when a researcher selects a number of cases for study rather than including the whole universe (or population) of possible cases in a study, and comparative research is typically based on a sample of cases. Quantitative research often deals very explicitly with sampling issues, but in qualitative research it is common that sampling does not receive the attention it deserves (given its potentially crucial impact on the conclusions drawn from the research; see Chapter 6). Typically, qualitative research designs are small-*n* and quantitative research designs large-*n*. Qualitative data is typically too rich and complex to make it possible to manage more than a few cases, whereas contemporary computer technology makes it easy to handle very large amounts of quantitative data.

The number of cases to be included in a comparative research design depends to a large extent on how many suitable cases (given the particular research question) are available. Normally comparative researchers are not lucky enough to find themselves in the 'predicament' of having too many suitable cases; having to make do with what is available is more common, also taking into account one's research resources. Even the study of a single case can be considered implicitly comparative, if it applies some widely used theory or model; at the other end of the scale the demands of inferential statistics can make 1,000 or more cases desirable. 'How many cases do I need for the project I'm planning, then?' is a question that usually comes to mind at this point. This question has no standard answer. The substantive topic of the research project and/or the research question most likely gives strong indications of what kind of data is relevant, and once that has become clear the menu for choice of appropriate methods of analysis is usually quite limited, too. Case studies are a particular case in point that can be elaborated in this chapter, specifically how they, as a genre of research, can be used. On the surface of things, it might seem that case studies per definition do not involve any attempt to compare, that they stand alone, without seeking

to engage or be part of the accumulation of knowledge in the literature. Sometimes a study might contain studies of more than one case (for instance a study of three democratization processes in Africa), which renders the explicit possibility of comparing the cases studied. Sometimes, however, a case study contains only one case (one twentieth-century US presidency) but shows a more or less explicit awareness of other, potentially comparable cases: that is, a study of Lyndon B. Johnson treats him as an example of a broader phenomenon such as democratic presidents in the last century, rather than solely as a figure of interest in his own right (although no one would deny that he is, but that is biographical work rather than political science). Other case studies do not make any effort to relate to existing research, or to add to the cumulative literature in an explicit way. No comparison occurs here, and it is less clear what role such research can play or what value it has in the scientific community.

As this example indicates, sometimes it is not so clear what a case study is. Gerring observes that 'As a general observation we might say that methods, strictly defined, tend to lose their shape as one looks closer at their innards. A study merges into a case study, a single-unit study merges into a study of a sample, a longitudinal study merges into a latitudinal study, informal cases merge into formal cases, and so forth. Methods that seem quite dissimilar in design bleed into one another when put into practice' (Gerring, 2004, p. 346). Upon examination, it certainly seems less clear what a case study is than at first sight. Gerring continues to position the case study vis-à-vis other ways to conduct research in terms of methodological affinities, tendencies that can be more or less strongly felt depending on numerous background factors. Here, researchers need to be aware of the unavoidable methodological trade-offs facing them: for instance, Gerring calls the case study 'a boon to new conceptualizations just as it is a bane to falsification' (2004, p. 350), on the basis of the methodological affinities specific to case studies. The question of 'how many cases?' consequently becomes at least partly a matter of what the researcher wants to achieve: is the objective to falsify? Then a case study is probably not appropriate, following Gerring's argument. Is the objective to take an in-depth look at some known mechanism? Then the case study format would seem a good choice.

Selecting to do a small-*n* study or a case study nevertheless has implications (as does the choice of any particular form of research). Lieberson (1991, 1994) draws our attention to four

implications: he demonstrates why small-*n* and case study work by logical necessity must adopt a deterministic rather than probabilistic notion of causation, why there must be an assumption of no measurement error, that such work must hypothesize only one cause, and that interaction factors are also beyond the scope of research of this kind. To begin with the point about deterministic causation (see Chapter 6 for more on determinism and probabilistic causality), Lieberson argues that the small number of cases that defines this type of research means that it is impossible to examine probabilities. The deterministic approach implies that a cause is only a cause if its presence generates the same effect each time the cause is in some sense 'present'. (The probabilistic approach is less demanding: to define something as a cause, it is sufficient that the cause's presence increases the probability of the effect occurring.) Working with a small number of cases or only one case effectively means, says Lieberson, that the researcher will only be able to identify 'deterministic causes'. It would seem that even a very famous and successful small-*n* researcher, Theda Skocpol, would agree with this point: in 1984 she wrote: 'In contrast to the probabilistic techniques of statistical analysis – techniques that are used when there are very large numbers of cases and continuously quantified variables to analyze – comparative historical analyses proceed through logical juxtapositions of aspects of small numbers of cases. They attempt to identify invariant causal configurations that necessarily (rather than probably) combine to account for outcomes of interest' (1984, p. 378).

Lieberson's three further points are actually all related to the first one: measurement error, for instance, might lead the researcher to conclude that a certain factor does not have the theorized, expected effect. Under the deterministic notion of causation, the consequences are particularly grave, since if one understands causality in a probabilistic way measurement error does not necessarily mean that one rejects a factor as a causal factor, but that one underestimates the probability of its generating a given effect. (That is also problematic, but less grave than rejecting the factor as a cause of out hand.) Moreover, it is a limitation that a small number of cases prevents us from controlling for a large number of 'competing' causal factors. It stands to reason that we cannot observe several different constellations of factors if we only have very few cases. For the same reason small-*n* work cannot cope with interaction effects: they require

too much data. (Interaction effect: the influence of one independent variable on the dependent variable is not independent of the value(s) on another independent variable.)

The number of cases appropriate in a particular research endeavour thus depends on the research question, the data used to answer that question, the methodology appropriate to that data, as well as our general objectives (falsification? generation of new theories? causal inference? exploratory work? and so on). Whatever the number of cases in a comparative study, some criterion for case selection must be employed, and here the 'most similar' and 'most different' research designs provide guidance, but only to a point. In addition to the selection criteria set out by these two types of research design it is also important that cases are not selected on the dependent variable. To select cases on the dependent variable (see Box 3.5) means that cases are chosen because they belong to a specific classification category (for example have a particular value) on the dependent variable, and this can become a particular problem in small-*n* research designs. The problem is that it becomes impossible to find out about the effect of the independent variable on the dependent variable because there is not even a theoretical possibility of variance on the dependent variable. For example, if we want to know about the factors that lead to a revolution, it is necessary to include cases where no revolution occurred (although such cases are trickier to find than cases of revolutions), since otherwise a statement on the effects of purported causes is impossible. As Barbara Geddes points out, 'apparent causes that all the selected cases have in common may turn out to occur just as frequently among cases in which the effect they are supposed to have caused has not occurred. Relationships that seem to exist between causes and effects in a small sample selected on the dependent variable may disappear or be reversed when cases that span the full range of the dependent variable are examined' (Geddes, 2003, p. 129).

BOX 3.5 Selecting cases on the dependent variable: the European Union as an international actor

The member states of the European Union (EU) have been making increasing efforts to act coherently in international politics in general, and in specific conflicts in particular. The rationale behind this is that by

acting in a unified manner the EU member states increase their influence, compared to if they acted as 15 (or 27) individual countries.

Students of European and regional security might wonder under what conditions the EU member states manage to maintain a unified position, and under what conditions unity breaks down. Since all EU member states are also members of the UN, one way to investigate this is to study how EU governments have voted in the UN General Assembly, on resolutions referring to a particular conflict or issue.

Take Kosovo as an example of a conflict in geographical proximity to the EU, and therefore of immediate significance to the EU. In the 1990s the General Assembly adopted six resolutions about Kosovo by voting. The EU15 voted as follows:

Date of Vote	EU Bloc Vote	Comment
23/12/1994	EU14	Abstention (Greece)
22/12/1995	EU14	Abstention (Greece)
12/12/1996	EU14	Non-voting (Greece)
12/12/1997	EU15	
09/12/1998	EU15	
17/12/1999	EU15	

To select *either* exclusively votes where EU bloc voting broke down, *or* exclusively votes where it was maintained, is to select cases on the dependent variable. Selecting on the dependent variable means that there is no variance on the dependent variable. The practical implication of this is that it becomes impossible to reveal under what conditions the EU15 do (or do not) succeed in adopting unified positions on international politics.

For example, if two or more of the successful votes are selected and compared, the researcher might be able to say something about what these cases had in common. Equally, if two or more unsuccessful votes are selected and compared, then the researcher will be able to say something about what those cases had in common.

However, unless cases are included that are *different with respect to the dependent variable,* it will not be possible to draw any conclusions that specify whether particular conditions make it possible/impossible for the EU15 to maintain a unified position.

Source: UN (October 2002).

Whether a study is large- or small-n makes an important difference in terms of how one might select cases for comparison. Specifically, random selection is usually the preferred way of deciding what to observe in large-n studies (especially the selection of respondents in opinion polls). In small-n research cases are typically not randomly selected – and in fact doing so might not be at all appropriate – but are selected precisely because they belong to a particular classification category on the independent variable. We have seen that this is the case in both 'most similar' and 'most different' research designs. In most similar research designs cases are selected according to difference on the independent variable (and similarity on other backdrop variables), and vice versa for most different research designs. However, non-randomly selected cases must not be selected on the dependent variable: that is, if the objective of a research design is to investigate whether mixed-gender units of the armed forces perform differently from all-male units, then 'performance' (poor, medium, good) is the dependent variable. If the units selected for comparison all performed equally (whether they were uniformly poor, or all medium, or good across the board), nothing would be revealed about the research question. The selection on the dependent variable prevents the observation of, even in theory, any difference between mixed-gender and all-male units becoming evident. The consequence of this extreme selection bias is that nothing is learnt about the performance effects of men and women serving alongside each other in the armed forces.

Selecting on the dependent variable may seem like an easy mistake to avoid, but sometimes it is not. Political scientists (especially those researching sensitive topics, and the presence of women in the armed forces may be one such topic) are often forced to use whatever data is available, or made available by 'gatekeepers'. Such gatekeepers may have reasons to give and withhold certain information precisely on the basis of how it relates to the dependent variable. For example, in the above example, a gatekeeper may only provide information on mixed-gender units that performed very poorly; or if the political agenda were different, might withhold that information and only reveal information about excelling mixed-gender units. The consequence for the research project would be a serious source of selection bias that the researcher may only suspect but never overcome. Box 3.6 provides an overview of the points made in this chapter through a checklist.

BOX 3.6 **Research design checklist**

1. The research question
   - What is the research question?
   - What is the dependent variable in this research question?
   - What is the independent variable(s) in this research question?
   - In addition to the independent variable(s), what other variables might also be related to the dependent variable?

2. Case selection
   - Can you list all/several cases that would be appropriate in a study of your research question?
   - Are you sure the cases on the list have not been selected on the dependent variable? *(All comparative research projects need to contain at least a theoretical chance of variance on the dependent variable. Selecting on the dependent variable is a particularly easy mistake to make in small-n projects.)*
   - Are the cases on your list 'typical' examples of the problem/ tension/ issue contained in your research question? *('Typical' examples maximize experimental variance. 'Deviant' cases do not minimize error variance, because it is impossible to extrapolate from unusual cases; see Chapter 6 for more on inference-making.)*

3. Most similar or most different research design
   - Are there cases on the list that would form a solid most similar research design? *(For example cases that are different on the independent variable but 'the same' on other variables. This controls extraneous variance.)*
   - Are there cases on the list that would form a solid most different research design? *(For example cases that are 'the same' on the independent variable but different on other variables? This combination also controls extraneous variance.)*

Conclusion

Most political science is comparative, even if not explicitly so. Comparativists 'examine a case to reveal what it tells us about a larger set of political phenomena' (Lichbach and Zuckerman, 1997, p. 4), and 'perhaps the only circumstance in which a political scientist is not also at least implicitly a comparative political scientist is when he or she remains consistently and exclusively concerned with his or her own national system' (Pennings, Keman

and Kleinnijnehuis, 1999, p. 70). Even then, it is questionable whether such a consistent and exclusive focus on a single system is not at least implicitly comparative because it is likely to make use of the same concepts, models and theories that have been applied elsewhere; if it does not, then it can hardly be said to constitute a part of a wider political science body of knowledge.

This chapter has approached the comparative methodology as the way to obtain as many of the advantages as possible of experimental control in research, but has also acknowledged some common problems that make full experimental control almost impossible to achieve. By careful case selection some of these problems can sometimes be alleviated or even avoided, but even where this is not possible the comparative methodology can be the best alternative available for political scientists.