Case selection and selection bias in small-n research

Chapter · October 2007

DOI: 10.1057/9780230598881

CITATIONS
30

READS
3,650

1 author:

Dirk Leuffen
University of Konstanz
89 PUBLICATIONS **2,081** CITATIONS

SEE PROFILE

8

Case Selection and Selection Bias in Small-n Research*

Dirk Leuffen

Introduction

Designing social research is often a blood, toil, sweat and tears experience, with the road to publication usually long and winding. Constantly, the researcher has to weigh different options, and case selection is often considered a particularly delicate and demanding step. For King and colleagues (1994, p. 115), 'poor case selection can vitiate even the most ingenious attempts, at a later stage, to make valid causal inferences.' In small-n as well as in large-n approaches 'the cases you choose affect the answers you get' (Geddes, 1990). However, case selection usually differs between those two approaches – and for good reasons. Whilst large-n studies generally seek representativeness, for example by random sampling, case selection in small-n research usually follows an intentional logic. Intentional does not, however, mean arbitrary. In the end, the types of cases you select determine which inferences you can draw.

In an idealized research cycle, case selection usually takes place after the formulation of the research question, elaboration or compilation of theories and concept specification. Case selection thus links theory development and the empirical testing of these theories. As they usually select their cases non-randomly, small-n researchers are particularly in jeopardy of introducing selection bias. A selection bias results from a faulty inference that wrongly attributes the properties of the scrutinized cases to the larger universe of cases. In this chapter, I will first identify different types of selection bias. I will then introduce some strategies for case selection that are commonly applied in small-n research. I will argue that case selection in small-n research should be considered a theory-guided iterative process. Theory defines the variables that are to be included in the research design. On the basis of these variables we can

construct multi-dimensional classification schemes that structure the possible universe of cases on theoretical grounds. Such typologies help in selecting cases as well as discussing the generalizability of one's findings. In this chapter, I will illustrate some methods of case selection by referring to my own research on the consequences of French divided government. Referring to my own research allows me to point out some problems and trade-offs around the issue of case selection that can arise during the research process. Methodological treatises often tend to merge the steps of case selection, data collection and data analysis. This, however, is not the practitioner's view. In practice, for example, the universe of cases is often not known right from the start, and case selection must take place behind a veil of ignorance. But how should one proceed when the universe of cases is 'clouded in mist'?

In general, when it comes to case-selection there is not one road to salvation, rather 'earthly sinners' must find their own paths. Some paths, however, seem more appropriate than others. Researchers should first of all be conscious about the pitfalls of case selection and be as transparent as possible when describing their case selection strategies. Verbalizing the problems around case selection and discussing the principal trade-offs is already a step that helps the reader gauge the impact of possible bias.

Design problem

This chapter analyzes case selection and selection bias in the social sciences. A case here is considered a unit or an object of comparison. It takes a particular value on each dimension that is submitted to the comparison (Eckstein, 1992, p. 125). As stated above, a selection bias is a systematic error that results from improper inferences drawn from a sample (Collier and Mahoney, 1996, p. 59). Accordingly, in 'configurative-ideographic studies' (Eckstein, 1992, p. 136), 'atheoretical' or 'interpretative case studies' (Lijphart, 1971, p. 691) selection bias is not a big problem since those kinds of research focus on cases *per se*. However, to therefore conclude that we should all refrain from drawing generalizations and inferences altogether is certainly not a satisfying response. In fact, such a strategy would correspond to suicide from fear of death.

In theory, we can distinguish between different types of bias. In a real world or contingency bias, the universe of available cases is biased by historical contingencies. For instance, in comparative country studies the number of available cases is clearly determined by the historical development of nation-states (Ebbinghaus, 2005, p. 138f.). If, however,

nation-building is linked to what we wish to explain, such a contingency bias may raise some delicate problems of endogeneity (Hug, 2003, p. 257). In this chapter we will, however, mostly focus on researcher-induced bias. Such bias results from improper measurement or non-random case selection strategies. Measurement bias can be attributed to unreliable indicators or biased sources. Hug (2003, p. 258), for instance, shows that an analysis of social movements that primarily relies on media reports is likely to suffer from a bias. The media pre-select cases according to their own logic. For example, newspapers might over-report violent forms of participation. Similarly, legislative datasets can be biased. Mayhew's (1991) study on divided government, for example, is criticized by Fiorina (1996, p. 90) for only analyzing the production of legislation and neglecting its demand. Edwards and colleagues (1997) and Binder (1999) in response use data-sets that include unsuccessful proposals as well. This, indeed, brings forth more nuanced findings about the consequences of divided government.

Self-selection, volunteer or participation bias is closely related to the measurement process. Pre-selection mechanisms such as response or non-response non-randomly determine who participates in a survey or experimental sample. In the case of such bias the sample does not mirror the target population properly. Whereas real world and measurement bias equally applies to large-n and small-n research the bias resulting from an intentional or non-random selection of cases is more of a problem for small-n research (Achen and Snidal, 1989, pp. 160–1; King, Keohane and Verba, 1994, p. 140). In small-n research, cases are generally selected intentionally. An intentional selection of cases, however, leaves room for manipulation. A rather obvious selection bias occurs if a researcher only selects cases that confirm the initial theory. Conspiracy theories most clearly display such confirmation bias. Conspiracy designers only collect information that support their theory and do not report any opposing evidence. Almost anything can be claimed by such a method; however, such practice does not at all meet the standards of social scientific research. A more common type of confirmation bias, however, consists in selecting cases that share the same value on the dependent variable (Achen and Snidal, 1989; Geddes, 2003; King, Keohane and Verba, 1994, p. 129). When interested in revolutions, researchers study revolutions (Skocpol, 1979), when interested in prosperity they analyze economically successful countries. Not including 'negative' cases, however, can lead to the false conclusion that any characteristic that these cases share should be considered a cause. Finally, selection bias does not only occur in cross sectional analysis.

Also in longitudinal designs there is a risk of leaving out important information. For example, when complaining about the present without comparing the past there is a selection bias. If the previous values of our variables remain unknown, how can we judge a causal effect?

Understanding what to avoid is one thing. Knowing how to proceed is a different matter altogether. How shall cases be selected in small-n research? Before introducing different techniques for case selection in small-n research, I will shortly address the question of why small-n in the first place. Given the enormous challenges of small-n, why do researchers still engage in such dangerous research strategies?

Why small-n?

If case selection is such a crucial research step that bears such enormous risks in small-n research, why then choose a small-n design in the first place? It is generally accepted that the formulation of research questions and interests should precede the elaboration of the research design. If you are interested in a particular topic you should work with the best methods available to solve your research puzzle. Subject to the research question there are basically two reasons for choosing a small-n design. First, data properties might restrict the number of available cases. Second, methodological objectives such as getting a better grip of causal processes or mechanisms can incite a researcher to choose a small-n approach. As a matter of fact, for numerous research questions, just a few cases exist out there. Think, for example, of revolutions, international monetary systems or interim Presidents in the Fifth Republic. This is not to say that such topics can never be related to broader concepts; for example, when interested in revolutions one could also include mutinies in the analysis. This, however, often comes at the price of *concept stretching* (on concept specification and its pitfalls see Wonka, Chapter 3). When interested in the extinction of dinosaurs it might not be advisable to study today's endangered species since the particular conditions and potential causes such as pollution might differ a lot.

There are other research questions that relate to larger populations of cases in the real world. However, systematic data-sets do not exist. Can the researcher afford to collect new data? Data collection always comes at a high price in an imperfect world with limited resources. In addition, external restrictions such as data secrecy, for example, often hinder the analysis of political decision-making. Political processes often lack transparency, and decision-makers try to hide their strategies and preferences from the public. This applies equally to international relations as well as

to domestic and comparative politics. For example, in my own research on French European policy-making in the context of divided government, both sides of the split-executive generally emphasize their influence in the executive decision-making process. When there is reason to doubt available information – for example, when you are confronted with severe inconsistencies – the researcher has to collect additional data. The application of different data sources to one issue – methodologists speak of triangulation – is, however, time-consuming. In such a situation opting for a small-n design might therefore be the only viable solution.

On the other hand, the choice for small-n can be methodologically motivated. For example, small-n research is often employed to uncover causal mechanisms. Techniques such as process-tracing (George and Bennett, 2005) enable us to at least theoretically include a great number of variables in a within-case analysis. By testing necessary causes we can narrow down the scope of potentially relevant variables and thereby foster theory development. This particularly holds for research programs that are still in their infancy. For Goldthorpe (2000, p. 59) detailed case studies can play a *heuristic* role in the 'context of discovery', prior to the testing of any resulting theory against further, independent cases in the 'context of validation'. More specific theories can then be submitted to more rigorous tests in a following step.

In the remaining sections of this chapter, I will concentrate on data property motivated small-n designs. That means that the universe of cases is actually quite large, yet accurate information is not available to the researcher. If the universe of cases is clouded in mist, how should one proceed?

Practical guidelines

In 1843 John Stuart Mill introduced the methods of difference and agreement. Originally designed for experimental research – that means research in which all relevant variables can be manipulated while everything else is held constant – these methods of inference have frequently been referred to in non-experimental case selection as well (Meckstroth, 1975). The methods of agreement and difference are useful tools to test necessary conditions (Ebbinghaus, 2005, p. 143). In the method of difference you select two or more cases that differ on the key independent variable while otherwise being as similar as possible. This is why Przeworski and Teune (1982, p. 32) speak of a 'most similar systems' design. If the dependent variable – *ceteris paribus* – varies in correspondence with the key independent variable, we detect a causal effect. King

and colleagues (1994) basically recommend this method for case selection. These authors suggest selecting observations 'according to the categories of the key causal explanatory variable' (King, Keohane and Verba, 1994, p. 137).[1] They find that this reduces the risk of introducing a bias since the selection procedure does not predetermine the study's outcome. Once the categories of the independent variable are defined, the cases should carefully be matched across the different categories. For example, when comparing decision-making in a coalition or single party government we should select similar types of decisions for both categories of this independent variable.

Fine examples of an application of the method of difference are Dreze and Sen (1995) as well as Haverland (2000). In order to test the 'goodness of fit' hypotheses for implementation records in the European Union, Haverland studies the implementation of the European Packaging and Packaging Waste Directive in Germany, the Netherlands and the United Kingdom. These cases are carefully selected along the categories of the independent variable. The 'goodness of fit' theory predicts that conformity between European provisions and national rules and practices can best explain the degree of national adjustments to European law. In this particular case Germany shows the best fit, the United Kingdom the least best, and the Netherlands scores in between. Haverland's in-depth case studies, however, cannot confirm the causal relationship that was hypothesized by the 'goodness of fit' hypothesis. In fact, the adaptation of the directives follows a different logic altogether. Whereas the UK did well in implementing the directives, Germany showed great difficulties. Haverland therefore comes up with an alternative explanation. He claims that institutional veto points can better explain national adjustments to European provisions.

In the method of agreement, the cases should agree on the dependent variable as well as on the key independent variable. The other variables should, ideally, differ between the cases. In the method of agreement, 'the investigator employs the logic of elimination to exclude as a candidate cause (independent variable) for the common outcome (dependent variable) in two or more cases those conditions that are not present in both cases' (George and Bennett, 2005, p. 155). The method of agreement is often linked to Przeworski and Teune's (1982, p. 34) 'most different systems' design. However, the selection of cases that agree on the dependent variable has lately been criticized by many methodologists (Achen and Snidal, 1989; Geddes, 2003; King, Keohane and Verba, 1994, p. 129). In fact, the serious shortcomings of Mill's methods in non-experimental social research have extensively been addressed in the

literature (George and Bennett, 2005, pp. 153–60; Goldthorpe, 2000, pp. 49–50; Lieberson, 1992). First of all, these methods rely on the rather weak assumption that all candidate causes can first be identified and then included in a small-n research design. In social reality, cases are rarely as neatly ordered as demanded by this theory. They score more or less similarly on the different variables of interest. In addition, neither the method of agreement nor the method of difference can account for equifinality or multiple causation (King, Keohane and Verba, 1994, p. 87).[2] If a variable exerts its causal power only under specific conditions or in combination with other variables it will nevertheless be eliminated by Mill's methods. Methodologists therefore speak of 'false negatives' (George and Bennett, 2005, p. 156).

Mill's most important contribution, on the other hand, probably consists of drawing our attention to the importance of introducing variance in our designs. The cases we compare should somehow differ. In general terms, causal analysis needs variance.

How to make use of theory-guided typologies

Constructing multidimensional classification schemes or typologies (George and Bennett, 2005, p. 235; Lehnert, Chapter 4) can be a useful starting point for case selection. In order to do so the researcher should first identify the independent variables. This is a theoretical task. Of course, a review of the existing literature is generally the first step. Once the independent variables are identified, they constitute the different dimensions of our typology. This gives us a clearer picture of the potential universe of cases. We can now classify the different cases and derive theoretical predictions about the values of the dependent variable. Our classification scheme also helps us relate our cases to other cases as well as to the universe of cases. This helps in judging the direction and magnitude of possible biases. Although typologies are primarily descriptive in character, they are a valuable starting point for causal analyses. In the end, when it comes to generalizing findings, qualifying one's cases and relating them to the larger universe of cases is also facilitated by typologies.

Pliny's strategy: not many things but much!

In addition, theory-guided typologies allow researchers to control for alternative explanations (Diesing, 1972, p. 189; George and Bennett, 2005, p. 237; George and McKeown, 1985, p. 45). Control is a central element of case selection. Whereas George and McKeown (1985, p. 28) suggest filling as many cells of a typology as possible, I take here a

different stance. Following Pliny the Younger's dictum '*multum non multa*' – not many things (*multa*) but much (*multum*) – I propose to start off by consciously narrowing down the domain and focusing on a set of theoretically interesting cells. Only after some certainty is established in these clearly delineated domains, for example, by analyzing more than just a single case per cell, we should move on to other cells and/or dimensions. When theories are weak – and that is often precisely the *raison d'être* of small-n social science – it is often unclear which variables should be taken into account in the first place. This can be a particular challenge for small-n research in which we can only test a very limited number of variables due to a lack of degrees of freedom (Goldthorpe, 2000, p. 49; Lijphart, 1971, p. 686; King, Keohane and Verba, 1994, pp. 118–124).[3] The *multum non multa* approach, however, averts the degree of freedom problem by limiting the number of variables being tested. Of course, if a theory is very sophisticated – for example, if we can derive precise expectations for each and every cell – and data on the entire range of cases can possibly be collected, there is no reason not to follow George and McKeown's suggestions. However, when confronted with the trade-off between scope and certainty it often makes sense to opt for certainty given the restrictions of small-n research. More robust findings about particular types of cases clearly contribute in a building block fashion to the cumulative development of the social sciences (Geddes, 2003).

Sinatra's strategy: select 'hard' cases!

Once a theory-guided typology is constructed we can select cases that promise the best possible answer to our research question – always given the restrictions of a non-ideal world. In general, it is always good to submit our theories to hard tests. Accordingly, a useful small-n strategy can consist in selecting particularly 'hard' cases. Hard cases are cases that represent a tough test for a theory. A typology is a useful tool for identifying such cases on theoretical grounds and relating them to the larger universe of cases. If a theory holds in a particularly hard case it is more generally supported. The underlying logic could be called the 'New York inference' following Sinatra's song line 'if I can make it there, I'll make it anywhere.' Such 'hard' cases are also called '*a fortiori*' cases: if a theory beats alternative theories in a particularly difficult context it should all the more hold under more favorable conditions. For example, I relied on a New York inference when generalizing from my findings on 'high politics' in my analysis of French European policy-making during divided government. I argue that if the Prime minister can increase his influence on European Council decision-making it is very likely that he, equally,

should play an important role in 'low politics.' An equivalent logic respectively applies to the selection of most likely or easy cases. A theory can best be called into question when it fails to explain cases that should be an easy game on theoretical grounds. Turning Sinatra upside down we could respond: 'if the theory can't make it there, it won't make it anywhere!'[4] Designs that rely on New York inferences can thus test theories on the basis of just a few cases. However, the researcher must assure that she is not focusing on outliers.

Case selection as a theory-guided, iterative process

How many cases should be included in a small-n research project? The number of cases generally depends on the number of variables that you want to investigate. However, the Sinatra inference underlines that there are exceptions to this general rule. In general, even in small-n studies there usually is a diminishing marginal utility in terms of depth of analysis. At some point the additional utility gained by deeper investigation into a single case can become very small. In very general terms, case analysis should therefore be as deep as necessary – the most important characteristics of a case, of course, need to be grasped – and as wide as possible as to the number of cases. If cases are easily accessible or data is available why should we not include such information in our research?

Small-n case selection should be considered a theory-guided iterative process. Knowing about the trade-offs in the process of selecting one's cases is important. The researcher should be aware of the advantages of the different approaches when designing social inquiry. In the following I will illustrate some techniques and trade-offs around the issue of case selection by referring to my own research on the consequences of French divided government.

An application: gauging the effects of French divided government

In my PhD-thesis I analyzed the consequences of divided government in France (Leuffen 2007). The so-called cohabitation is a split-executive government in which the French president stands in opposition to the parliamentary majority and to the Prime minister (PM). So far, the Fifth Republic has seen three cohabitations: in the first cohabitation President Mitterrand had to govern with PM Jacques Chirac (1986–88), in the second cohabitation President Mitterrand had to govern with PM Balladur (1993–95) and in the third cohabitation President Chirac had to cope with a government of the *gauche plurielle* under PM Lionel Jospin (1997–2002).

The literature was divided about the effects of cohabitation. When compared to the sophisticated literature on the consequences of the US-divided government, the French case clearly seemed understudied.

In a first step, the research question had to be narrowed down. Since the effects of cohabitation should vary considerably across different policy areas, I decided to restrict my analysis to one field, European policy-making. This was a theory-based decision. The effects of divided government depend on the decision-making procedures and on the policy-preferences of the actors (Milner, 1998, p. 774). In a rational institutionalist perspective, if the actors share the same preferences, divided government does not matter since the outputs of the political game do not change. In addition, decision-making should be shared between the divided actors. Only if there is some cooperation between the branches of government can divided government have a systematic effect on 'system production' (Mayhew, 1991, p. 35). These conditions clearly apply to the field of European policy-making. In France, the President has only limited formal power in the legislative process. On the other hand, the French Constitution is ambiguous as to who – the President or the PM – should be responsible for foreign policy-making. This ambiguity also covers European policy-making since this domain can be situated between the domestic and international realms. I also assumed that this proximity to domestic politics makes Europe a more controversial and partisan subject as compared to more traditional foreign policy issues. Thus, in European policy-making the two conditions for an effect of cohabitation were met.

After narrowing down the research question how did I proceed? The theory that cohabitation matters, expects that the vote-maximizing PM during cohabitation strives for utmost influence on formulating the positions that France holds in the European arena. In contrast, during unified government the President dominates French European policy-making. The PM thus challenges this supremacy during cohabitation. The literature is divided about how the split-executive formulates France's European policies. Some authors find that the PM, indeed, dominates the process; others, however, maintain that the President is still the most important actor. A third branch argues that European policy-making should be considered a *domaine partagé*. Based on an account of the actors' resources and the institutional setting, I follow this third read and propose to test a veto-player theory of French European policy-making during cohabitation (Tsebelis, 2002). Veto-player theory generates a precise hypothesis on the effects of cohabitation. This theory claims that the number of veto-players rises during

divided government since the PM should be added as an additional veto-player. On average, the size of the French acceptance sets thus shrinks. Following Schelling (1960, p. 28) this should also affect decision-making at the European level (Putnam, 1988). In such a view, cohabitation should thus systematically slow down European integration.

In order to test this theory I introduced some simple models on core executive decision-making (Dunleavy and Rhodes, 1990; Laver and Shepsle, 1994). The models generate different predictions about the positions France should hold in the European arena. Which cases did I select in order to analyze the effects of French divided government? Most studies on the consequences of US divided government rely on statistical techniques (Cameron, 2000; Mayhew, 1991). In the US, the relations between the President and Congress are extremely well documented. This, however, does not apply to the intra-executive French decision-making processes. The government archives are closed for 25 years and since both sides had a strategic interest in exaggerating their respective influence during cohabitation, the available data on the formulation of French positions and the initial preferences of the actors must be handled with great care. The available sources are thus weak, which is why I opted for a triangulation of the data. However, triangulation, as data collection in general, is time-consuming. Therefore the number of cases I was able to study in-depth was clearly restricted. Of course, an in-depth study of only a few cases promised to provide a clearer picture of the decision-making mechanisms.[5] Unveiling the mechanisms is a general advantage of comparative case-study approaches.

Which cases did I select for in-depth analysis? The research question on the effects of cohabitation is factor centric. The logic underlying this question relates to Mill's method of difference. I argue that – all other factors being equal – the switch from unified to divided government changes executive decision-making. This, in turn, affects the positions France holds in the European game. If there are systematic differences between unified and divided government, cohabitation matters. If we do not find any systematic differences there is no causal connection between unified or divided government and the European positions that France holds. King and colleagues (1994) have argued that cases should be chosen according to the categories of the key causal variable. In my research the key causal variable comprised the two categories of divided and unified government. Therefore decision-making processes during divided as well as during unified government should be analyzed. To every rule, however, there is an exception. Such an exception is also

acknowledged by King and colleagues:

> However, most research is part of a literature or research tradition … , and so some useful prior information is likely to be known. For example, the usual range of the dependent variable might be very well known when the explanatory variable takes on, for instance, one particular value. The researcher who conducts a study to find out the range of the dependent variable for one other different value of the explanatory variable can be the first to estimate the causal effect. (1994, p. 146)

This precisely fits the case of cohabitation. Unified government decision-making has been analyzed extensively in the past and, in fact, there is little doubt about the dominance of the French President. In contrast, our knowledge about decision-making during divided government is very limited. I therefore decided to focus my in-depth case studies on decisions that were taken under the conditions of cohabitation. My analysis of unified government, on the other hand, is based on the literature.[6] Since in the end I finally compare European policy-making during divided and unified government, there is no selection bias (Gerring, 2004, p. 347; King, Keohane and Verba, 1994, p. 137). Because I concentrate my in-depth analysis on decisions taken under divided government, I am able to better understand the decision-making processes during cohabitation. Had I decided to carefully match cases – for example, by adding the Maastricht negotiations or another budgetary round as, for example, the Delors-II-package – I would have only been able to study three or four cases from each of the two categories of the independent variable. This is thus where a trade-off comes into play. I opted for increasing the variance of cases during divided government since this category of the independent variable had been understudied in the past.

But which of the numerous possible cases should be selected from divided government? This is now where I rely on a theory-guided typology. Whereas George and McKeown (1985) suggest filling as many cells as possible, I decided to concentrate on a few important cells and leave it to future research to fill the rest. For example, I decided to restrict my analysis to 'high politics', that means policies negotiated at the level of the European Council. The analysis thus controls for 'range of decision'. As I did not include 'low politics' in my in-depth analysis, I can only make preliminary statements as to how cohabitation matters in such decision types. However, there are theoretical reasons that allow for formulating such expectations. If I am able to show that the President lost

power in 'high politics' this should all the more apply to 'low politics'. Given the President's limited administrative resources, his influence in such decisions should clearly be restricted. In an ideal world, I certainly should have tested this expectation. This, however, would have exhausted my resources which is why I kept this question for future research.

While I restricted my analysis to 'high politics' European policy-making, I decided to maximize the variance between the remaining cases by including other control variables. These control variables were 'issue area', 'decisional scope' and 'personalities'. Accordingly, the cases that were finally chosen covered different issues; they contained intergovernmental conferences as well as 'normal' European council decisions and were selected from all three cohabitations.

A final criterion for my case selection was a logical one. In order to test my models I needed to analyze some kind of conflict amongst the actors (Stokman und Thomson, 2004, p. 10; Thomson, Stokman, Achen and König, 2006). Without any initial differences between the actors we can hardly measure the impact the actors have on the decision-making processes. From a large-n view such a restriction should bias my findings. However, this research does not establish frequency distributions. I do not make precise claims on how often there is conflict amongst the actors of cohabitation, but rather I analyze how decisions are made during unified and divided government. In my research on cohabitation I finally analyzed seven cases in depth and added four mini-case-studies.[7] The mini-cases are less detailed but support the results of the other cases. They thus increase the robustness of my findings.

Typological theorizing can finally be useful when it comes to generalizing one's findings (George and Bennett, 2005, pp. 233–62). The reader has to know which cells have been filled by your research and what remains to be done. The typology helps to discuss the conditions under which your theories seem to hold. You have to be clear about whether there are potentially relevant conditions that you were unable to test. Another useful technique for assessing the generalizability of one's findings consists in establishing links to other similar data-sets or causal process observations. How does my data, for example, correspond to other data sets on governance in the Fifth Republic and beyond? How do the findings relate to the case of US divided government and the conflict between the President and Congress?[8] When generalizing findings, one should always come to nuanced assessments and inform the reader about the remaining uncertainties. Careful formulations, such as '[those] preliminary observations indicate' (Haverland, 2000, p. 100), help to better assess the scope of your findings. A modest stance on one's

findings is a virtue not just in terms of professional ethics but also in terms of efficiency. It contributes to the cumulative development of the social sciences.

Conclusion

In the beginning social scientific research often resembles a roller coaster ride. The researcher moves up and down the ladders of abstraction, the various levels of analysis, between data and theory fascinated by a theoretical or empirical puzzle. When streamlining one's research there are, however, many trade-offs the researcher must face. This also applies to case selection. Since the empirical findings of a study depend enormously on the cases studied, the importance of case selection can hardly be overrated. This is why case selection is considered a major step in designing social research.

In this chapter, I analyzed case selection and selection biases in small-n research. After introducing standard procedures of case selection, some applications of these methods were highlighted by my own research on divided government in France. In general, case selection in small-n research can be considered a theory-guided iterative process. There is no uniform rule as to how to proceed in case selection. Of course, the cases you select first of all depend on the questions that you ask. There are, however, some strategies that might help to avoid the most pertinent pitfalls of case selection. After carefully narrowing down the research question and formulating theories, one should select cases that guarantee sufficient variance in terms of our research question. However, given the restrictions of small-n research it is sometimes difficult to maximize the variance between the cases of our analysis. Here typologies can help to select theoretically interesting cases. Sometimes it can be an asset to restrict one's analysis to just a few interesting cells and relate one's findings to the literature. In general, the smaller the domain, the more certain one's findings are likely to be. However, we lose out in terms of scope.

In general, our case selection mechanisms should be made as transparent as possible so that the reader can estimate the magnitude and directions of possible bias (Geddes, 2003, p. 25). While designing social inquiry, you are constantly confronted with possible trade-offs. Such trade-offs should explicitly be addressed when publishing one's findings: which options did I face? Why did I finally choose this particular strategy? Discussing the pros and cons of your case selection is a sign of maturity and methodological awareness. If you enable the reader to

gauge possible bias and restrictions of your own analysis, you take an important step towards improving the quality of social scientific research.

Notes

* The author thanks the participants of the MZES working group on research design for many vivid and thought-provoking discussions. Many thanks also go to the editors of this volume as well as to Markus Haverland, Guido Schwellnus, Stefan Seidendorf and Stefanie Walter for helpful comments and suggestions. The original manuscript of this article was drafted in 2005 and completed in 2006. Gerring's (2007) seminal contribution on case study research is therefore not taken into account. The interested reader is strongly advised to also read Gerring's text book.

1. Their suggestion also relates to Mill's method of 'concomitant variation'. This method relies on at least ordinally scaled key independent and dependent variables and assumes that there is a proportional causal relation. Luebbert's (1987) study on the development of interwar regimes in Europe is a good example of an application of the method of concomitant variation (cf. Mahoney, 2000, pp. 403–4).
2. We speak of equifinality if different causes or combinations of causes bring about the same effect.
3. However, the careful collection and analysis of causal process observations can counterbalance some of these effects.
4. A good example for such logic is Lijphart (1968). Studies that rely on such reasoning additionally employ qualitative techniques in order to investigate the mechanisms that bring about the observed results (King, Keohane and Verba, 1994, p. 477; McKeown, 1999, p. 173; Rogowski, 1995). Within-case analyses can thus support a theory in terms of causal mechanisms (George and Bennett, 2005, p. 206; McKeown, 1999, pp. 173–4).
5. Static comparisons of preferences and negotiation outcomes often neglect how certain results come about. For example, in the case of my research on cohabitation I was able to show by careful process-tracing that President Mitterrand during the first cohabitation had less influence on the internal French decision-making processes than commonly believed. Since the results of the European bargains were closer to Mitterrand's than to Chirac's initial positions many commentators had falsely inferred that Mitterrand was able to impose his will on Chirac. The process-tracing, however, revealed that Chirac – for example in the budgetary negotiations of the Delors-I-package – finally accepted the European agreement in order to gain utmost financial support for French farmers. Thus, Mitterrand's influence was not decisive.
6. In addition, I analyzed two cases that started during unified government and ended during divided government. One case is on the GATT-negotiations. These negotiations started under the socialist government but were only concluded in 1993 and 1994 under the Balladur government. The second case is on the Amsterdam treaty. In this case the Juppé government was replaced by

the government of the *gauche plurielle* shortly before the final negotiations at Amsterdam. Such case-studies include two sub-cases and have a quasi-experimental structure (George and Bennett, 2005, p. 81). Most other factors do not vary but there is a change from unified to divided government.

7. The case studies cover, for example, the budgetary Delors-I-package and the Amsterdam intergovernmental conference. Such decisions usually combine various issues. Therefore the number of observations on which my results are based is actually higher than the number of case studies (cf. Gerring, 2004).
8. A similar technique is applied by Gschwend and Leuffen (2005) when these authors back their quota-sample data by comparing it to other sources (cf. Gschwend, 2005).