

1

Introduction: Designing Research in Political Science – A Dialogue between Theory and Data

Thomas Gschwend and Frank Schimmelfennig

Quick-and-dirty number-crunching ‘quantoids’ face them. Carefully describing and interpreting ‘smooshes’ face them (Hatch, 1985). No matter where they stand on ontological and epistemological grounds and how we stereotype the respective ‘other side’, all researchers face similar challenges posed to core issues of research design. How you deal with these challenges defines the research design for your individual projects. A research design is a plan that specifies how you plan to carry out your research project and, particularly, how you expect to use your evidence to answer your research question.¹

What is a relevant research problem? How can I improve concepts and measurements in my research? Which and how many variables and cases should I select? How can I evaluate rival explanations and which theoretical conclusions can I draw from my research? Which evidence would lead me to reject and reformulate my initial theory? These are central questions political science students inevitably face when they embark on their own research projects in a Master’s or a PhD program.

This book was written to help advanced students of political science think about these issues and come up with solutions for their own research. It has emerged out of a seminar course that we directed for several semesters. As the course united researchers from both the quantitative and qualitative ‘camps’, mutual misunderstandings and heated debates were inevitable. Despite this, seminar discussions also shaped a number of shared beliefs that provide the common ground

2 *Research Design in Political Science*

for this volume:

1. The methodological pluralism in our discipline is a strength rather than a weakness.
2. The basic problems of research design are the same for qualitative and quantitative political science research.
3. The methodological debate in the discipline often remains at an abstract level and does not give sufficient practical guidance for dealing with basic research design problems.
4. The distinction between qualitative and quantitative research is often inadequate. Some solutions to research design problems are common to both types of research; others cross-cut the traditional qualitative-quantitative divide.
5. At any rate, finding solutions to research design problems involves substantial trade-offs along the way. Each solution has its strengths and weaknesses.

Thus, the contributions to this volume do not start with general methodological discussions, but each focuses instead on a specific problem of research design. They explicate the problem, discuss various solutions, emphasize the typical trade-offs involved in choosing one or the other solution, present practical guidelines and illustrate the use of these guidelines in an example taken from their own research. In the remainder of the introduction, we will give an overview of the basic problems and different types of research design that will be taken up in the individual book chapters.

Core issues of research design

At a very general level, scientific research can be conceived of as a dialogue between theory and data. Researchers formulate a theory, analyze data to test it, reformulate their theory in light of the empirical evidence, and then move on to test the reformulated theory with new data. Or, starting at the other end, researchers make observations, develop a theory to explain them, use additional data to test their theory, and possibly reformulate it afterwards. Individual research projects do not necessarily go through the entire cycle. Science is a collective enterprise. Some research projects focus on testing existing hypotheses; others are more concerned with explaining specific observations and generating new hypotheses.

We claim, however, that all research projects that take part in this dialogue between theory and data face the same set of core research design

issues. These are: defining the research question and problem; specifying concepts and theory; operationalization and measurement; selecting cases and observations; controlling for alternative explanations; and drawing theoretical conclusions from the empirical analysis.² In the following paragraphs, we will address these tasks one by one.³

Defining the research problem. First of all, the researcher is faced with the question: 'What should I do research on?' The most general answer to this question is: 'Something relevant.' But relevant to whom and in which way? At this point, we can distinguish between theoretical or scientific relevance on the one hand, and social relevance on the other (see King, Keohane and Verba, 1994, p. 15). Research is relevant to the scientific community if it advances the collective dialogue between theory and data beyond the current state of the discipline – by formulating, testing and improving theory, by generating and improving data, and by describing and explaining observations. To do so, the researcher needs to identify puzzles and problems in the discipline such as a theoretical controversy; imprecise, inconsistent, incomplete or otherwise 'bad' theory; untested theories and unexplained observations; unreliable, invalid or otherwise 'bad' measurement and data. Research is socially relevant if it addresses social problems, improves citizens' and policymakers' understanding of the problem and, possibly, offers solutions. To do so, the researcher needs to clarify the social relevance of her research and demonstrate how it can be used to understand and solve social problems (Gerring and Yesnowitz, 2006). Yet the current state of the discipline leaves considerable room for improvement, and political problems abound. So researchers still have to decide (and justify) which of the numerous problems and puzzles they choose to address.

Specifying concepts. Whether we formulate and test theories or describe and explain observations, we inevitably use concepts such as 'democracy', 'party', 'conflict', and 'peace'. In order to make research relevant, these concepts need to be theoretically and/or socially important. But they also need to be (properly) specified. It must be clear what we mean by a specific concept, that is what its defining attributes are, how attributes and concepts relate to each other, and which empirical phenomena they include and exclude. What attributes define a 'democracy'? Does 'peace' exclude 'conflict'? How do 'parties' differ from other organizations? Clear and unambiguous concepts are not only required for formulating testable theories in the first place. When engaging in a theoretical controversy, the researcher needs to examine the concepts of the competing theories – especially when the theories use the same

4 *Research Design in Political Science*

terms. Starting from the data, descriptive inference requires no less careful concept specification – for instance, if you make statements like ‘The majority of states are democracies’ or ‘The occurrence of war is decreasing’.

Specifying theory. Causal theories formulate cause-effect relationships between concepts. Thus, researchers not only need to specify the concepts themselves but also their relationship. Most basically, theories specify the order of the causal relationship between the concepts: what is the cause, what is the effect? Further specification may concern the form of the relationship (linear or non-linear) and the direction (positive or negative). Theories also need to specify the relationship among various assumed causes. Is it additive, as commonly assumed in linear regression models, or multiplicative, as is the case for interaction effects? Alternatively, causes can be characterized as necessary and/or sufficient conditions of the outcome. For instance, democratic peace theory holds that ‘joint democracy’, the fact that two countries are both democratic, is a sufficient (but not a necessary) condition for durable peace between them.

Furthermore, theories should specify the causal mechanisms that link cause and effect and theorize on the process through which the cause produces the effect. For example, the democratic peace has been explained by the transparency and inertia of political decision-making in democracies, which prevents secret preparations to war, slows down military escalation and gives democracies sufficient time to negotiate and find peaceful solutions to their conflicts (Russett, 1993, pp. 38–40). Advocates of causal mechanism analysis also generally demand that social science theories must specify their ‘microfoundations’ (Coleman, 1990; Hedberg and Swedström, 1998). That is, they must show how social structures and environments translate into individual desires and beliefs (macro-micro), how the actor produces preferences and actions on the basis of these desires and beliefs (micro-micro), and how the actions of many individuals are transformed into a collective, social outcome (micro-macro). The more fully a theory is specified, the more fully it potentially explains observations and the better it can be tested.

Measuring concepts. By specifying concepts and theory, we arrive at testable theoretical propositions. In order to conduct the empirical test, however, the concepts need to be operationalized and measured. Obviously, democracy – even if clearly specified as a concept – cannot be observed directly. This is often also true for the defining attributes. Alvarez and colleagues (1996), for instance, define democracy as a political regime in which offices are filled by contested elections. They then

go on to provide 'operational rules', which specify the offices that need to be included (the chief executive and the legislature) and indicators of 'contestation' (above all that there has to be more than one party). Furthermore, the operationalization would have to include indicators for determining the 'chief executive', the 'legislature' and 'parties'. Even after such a fine-grained operationalization, researchers still have to choose the instruments for measurement, for example expert assessments or legal documents. At any rate, the measurement needs to be both valid (the data needs to correspond to the specifications of the concept) and reliable (repeated measurement of the same phenomenon must produce the same values of the indicator).

Selecting cases. Problems of case selection and selection bias are core issues in both quantitative and qualitative methods textbooks. To be precise, we need to distinguish between units of analysis, cases, and observations. The unit of analysis is the abstract entity that we study (e.g., states, institutions, decisions) which is often given by the theory. 'Case' refers to the specific units of analysis that we choose to analyze. If the unit is 'state', this could be a single-case study of Sweden or a comparative case study of Sweden and Norway. Finally, one case may be equal to one observation if it consists in a single set of values of the independent and dependent variables. A single case, however, can also yield multiple observations. Research on the accession of Scandinavian countries to the European Union might be based on a single set of values for independent variables such as GDP per capita, growth, and export dependency for each Scandinavian state. Alternatively, we can make multiple 'data-set observations' (Collier, Brady and Seawright, 2004b, p. 252) for each case, for instance by observing the values of these economic variables at different points in time. Or we can turn to a series of 'causal-process observations' in order to see how structural economic conditions were transformed into decisions on EU accession (such as lobbying by interest groups and election or referendum outcomes).

For theory testing, the question is how observations can be selected so that the results of the analysis are unbiased and provide a valid assessment of the theory. For the description and explanation of social phenomena, the question arises as to whether the selected observations represent the class of phenomena adequately. Is '9/11' representative of transnational terrorism? Does an analysis of 'Blairism' allow for general conclusions on the tendency toward personalization and media spin in current parliamentary democracies? Sometimes random selection is possible (as in studies of electoral behavior) but even here the selection

6 Research Design in Political Science

procedure may privilege one group of respondents over another – for instance, those people spending a lot of time at home and thus more likely to respond to calls by the polling institute. Sometimes we know the entire population of cases – such as democracies or post-communist revolutions – but empirical analysis of more than a few cases would be too demanding, and the random selection of those cases would most likely lead to bias. Finally, we may not even know the universe of cases. Generally, researchers are therefore confronted with either unintended or intentional non-random selection which must be taken into account in order to arrive at valid generalizations and theoretical conclusions.

Controlling for alternative explanations. In the dialogue between theory and data, we specify a theory in order to test it on the basis of the selected cases and measurements. Alternatively, we draw on or construct a theory in order to explain a set of observations or a specific outcome. Yet even if we find a strong relationship between the theorized causes and the observed effects, how can we be sure that this relationship is not spurious and that other causal factors would not explain the observations just as well if not better? For instance, the ‘democratic peace’ might be attributed to the hegemony of liberal great powers or to high economic interdependence between democratic countries. In other words, we have to address, and control for, alternative causal factors and explanations in our research. But how many and which alternative factors or variables should be included in the analysis, and how do we decide on which of these rival theories or causes provide the best explanation?

Drawing theoretical conclusions. Let us assume we have successfully tested a well-specified theory with valid and reliable measurements on an unbiased selection of cases and that we have been able to reject alternative explanations. In this case, the theory is corroborated and does not need to be revised or rejected. Often, however, we will encounter anomalies such as deviant cases or statistically insignificant relationships. What if, for instance, we find a single instance of two democratic countries waging war against each other? Could one deviant case simply be ignored or would this mean the democratic peace theory is flawed and should be dumped in the junkyard of falsified hypotheses? Could the theory be saved by respecification or by limiting its scope? At any rate, empirical research results do not speak for themselves. The conclusions we draw from them need to be well considered so that knowledge is improved rather than prematurely destroyed or falsely preserved. At

Table 1.1 Research design tasks and problems

Research design issue	Challenge
Research problem	Relevance
Concepts and theory	Clear specification
Measurement	Validity and reliability
Case selection	Valid and general inferences
Control	Valid and best explanation
Theoretical conclusions	Scientific progress

first glance, these considerations do not seem to be part of the research design, because they only come up after research has been concluded. However, research should be designed from the start in a way that allows us to draw the right conclusions for theory.

Table 1.1 sums up the main challenges posed by the various issues of research design. We need to define a relevant research problem, clearly specify our concepts and theory, provide for valid and reliable measurement, select cases that allow for the formulation of valid inferences and generalizing our results, control for alternative explanations to demonstrate the validity and superiority of the proposed theory, and advance scientific progress in drawing our theoretical conclusions from the findings. How we get there or, more modestly, how we get closer to meeting these challenges, will be the subject of the chapters in this volume.

Basic types of research design

In general, research designs can be individually tailored to the concrete research problem at hand. However, the literature suggests that there are a few basic types of research design that researchers can opt for and that differ with regard to, for instance, the selection of variables and cases, the choice of data and methods, and their implications for theory. In the following, we will provide an overview of the different types that will be taken up in the individual chapters. The one basic dichotomy is that of factor-centric versus outcome-centric research designs; the other one is large-n versus small-n designs.

Factor-centric versus outcome-centric research design. George and Bennett (1997) originally introduced the difference between factor-centric and outcome-centric research to describe alternative inference processes in case studies. There is no need, however, to restrict this terminology

8 Research Design in Political Science

merely to case study research. In fact, we find it very helpful when evaluating potential research designs more broadly.⁴ Research designs can be distinguished by the type of causal inference a researcher is trying to make in order to answer a research question. In planning to make a causal inference a researcher might be either interested in providing evidence for one or more particular causal mechanisms and effects or, instead, wants to account for specific outcomes as completely as possible. For instance, you could be interested in the mechanisms of how voters' preferences for particular parties facilitate their decision in the voting booth or, instead, try to predict their voting behavior.

We call a research-design *factor-centric* if one is primarily interested in the explanatory power of causal factors. The goal is to estimate the direction and size of a particular causal effect of one or a few independent variables, X_i ($i = 1, \dots, n$), on a dependent variable, Y , and to assess their robustness. Independent variables are either explanatory or test variables, which are of key interest for the causal effects and mechanisms you are after, or mere control variables, which are included to make sure that the causal effects can really be attributed to the explanatory or test variable rather than to alternative causal factors. Typical research questions of factor-centric research designs are: Does X_i cause Y or what effect does a X_i have on Y and how much? Thus if you are interested in how partisan preferences anchor a voter's decision-formation process, you might want to allow for alternative ways in which vote-choice decisions can be rooted – such as ideological or candidate preferences – in order to disentangle their potential impact from that of your major explanatory variable of interest.

A research design is *outcome-centric*, however, if one is primarily interested in explaining outcomes. The goal is to comprehensively assess potential and alternative explanations by considering many independent variables, X_i , that *in toto* try to account for variance in the dependent variable, Y , as completely as possible. Examples are explanations of the varying success of UN peacekeeping operations or the differential impact of EU law on the member states. Outcome-centric research might also be interested in explaining specific single events (the Iranian revolution or the end of the Cold War – in other words, a dependent variable without variance). The typical research question of outcome-centric research designs is: What causes Y or why Y ? Thus if you are interested in predicting individual voting behavior you might want to choose an outcome-centric research design and consequently include additional independent variables (e.g., contextual or media effects) that help you better predict behavior in the voting booth, even though the omission of those

variables does not have the potential to distort the individual-level relationships that factor-centric researchers might focus on.

What reasons might there be to choose one design rather than the other? We suggest that the choice is mainly up to the researcher's interest and considerations of relevance. If you are mainly interested in explaining important events in politics (such as wars or revolutions) or predicting the outcomes of specific political decisions (such as the formation of a government coalition), the obvious choice is an outcome-centric design. This often requires an in-depth knowledge of phenomena in which you are interested.

If, however, your research is driven by a theoretical interest in causal factors (such as resources or institutions) or mechanisms (such as political socialization or political dilemmas), factor-centric designs are the most suitable. Researchers opting for a factor-centric research design have to 'control for', 'account for' or 'hold constant' the influence of all potential confounding factors in order to separate out those effects from the causal relationship in which they are primarily interested. This is the central aspect for making valid inferences based on factor-centric research designs. There are various strategies that facilitate researchers in disentangling the causal net. Including control variables in regression equations, matching methods or laboratory and field experiments are potential solutions that require many observations. Yet there also exist strategies that allow for distinguishing the hypothesized from confounding effects without leveraging many observations. One strategy is to systematically compare only a few carefully matched cases. Another strategy is the quasi-experiment, where one compares the very same case before and after the 'treatment' such as an institutional change or a policy intervention (George and Bennett, 2005, ch. 8).

The choice between factor-centric and outcome-centric research designs is not necessarily tied to the state of theory development in a given field, although in theoretically less advanced fields researchers often opt for an outcome-centric research design. Such researchers try to explore new phenomena by focusing directly on the variance of the dependent variable. Nevertheless, good arguments have been made that focusing on a comprehensive explanation of a phenomenon by maximizing the accounted variance of a dependent variable may not be the most promising first step to develop new theoretical and empirical insights (e.g., Geddes, 2003, ch. 2, King, Keohane and Verba, 1994, p. 169, note 8). Such a strategy may simply not be feasible due to data collection problems, may make estimates of all causal effects more

10 *Research Design in Political Science*

uncertain, or may be a hindrance to the accumulation of knowledge based on a common theoretical framework.

Alternatively, following a factor-centric research design strategy, one could break up a comprehensive explanation into more manageable building blocks of a theory, identify relevant variables to describe the causal mechanisms involved in these blocks, and afterwards, piece those building blocks together. There is, however, considerable skepticism as to whether such a 'lego' strategy (Pierson and Skocpol, 2002, p. 717) actually facilitates the accumulation of knowledge. Critics of this approach point out that, rather than answering big relevant questions in broad contexts, this strategy leads to robust answers of small and potentially trivial questions (Pierson and Skocpol, 2002, pp. 713–18).

While theory development of a given field might predispose researchers to employ a particular research design, it in no way determines the choice between factor-centric or outcome-centric research designs. Even in theoretically more advanced fields, researchers do not only opt for a factor-centric research design, although it might be easier to isolate a particular causal factor and focus on the direction and size of its effect, given the advanced state of theory development. For instance, researchers may be primarily interested in forecasting future outcomes such as elections or state failures. Then, of course, outcome-centric research designs are essentially required to answer this kind of research question.

Large-n versus small-n research design. What's your 'n'? One of the most often applied dichotomies to classify research designs refers to the number of cases and observations you study. Large-n and small-n research designs differ in the way in which they leverage available empirical information. Large-n studies are commonly associated with statistical tests of correlation-based inferences following a probabilistic model of causation and leveraging 'data-set observations', that is, 'observations [that] are collected as an array of scores on specific variables for a designated sample of cases ...' (Brady, Collier and Seawright, 2004, p. 12).

Small-n studies, however, are commonly associated with either within-case analysis or cross-case comparisons (George and Bennett, 2005) and with leveraging multiple 'causal-process observations' for a single case (Brady, Collier and Seawright, 2004, p. 252). Case studies rely on process-tracing in order to better understand the causal mechanisms of the relationships and phenomena of interest (e.g., see George and Bennett, 2005, pp. 147–9). Such inferences can be made by closely tracing hypothesized causal processes either within a particular case or by a

systematic (controlled) comparison across a small number of cases (such as George and Bennett, 2005, ch. 8).

In other words, large-n studies seek to achieve and increase the validity of causal inferences by increasing the number of cases and data-set observations, whereas small-n studies seek to attain the same goal by carefully matching a limited number of cases and increasing the number of causal-process observations. Small-n research prefers depth to breadth, whereas large-n research prioritizes breadth. As a result, small-n research potentially leads to very precise causal stories for one or a few cases at the expense of generality, whereas large-n research strengthens our belief in the generality and average strength of causal effects at the expense of rendering individual cases largely 'invisible' (Ragin, 2000, p. 31) and by being unable to explain any single case precisely.

How should one choose between a small-n and a large-n research design? A fundamental principle is that better data collection methods are preferable to better data analytical methods. Thus, it is the art-part of designing your research in cleverly using available information, or gathering new information, and thinking hard about alternative sources of information and how they can be leveraged. Whenever sufficiently quantifiable and comparable information is available, large-n research designs are typically used. But buyer beware! Increasing the number of observations, even if potentially available, is no free lunch. Is the new information really comparable to the original? Do I have to stretch concepts in order to derive comparability? Do the indicators fit the new cases? The leverage obtained by adding observations might be reduced. Alternatively, no harm is done in adding causal process information to bolster causal claims based on the original data set. In fact, this perception is also shared by hard-core large-n statistical wizards (for instance, see Beck, 2006; Goldthorpe, 2001).

The division of published research into small-n and large-n is not only conceptual but shows up in actual research practice as well (see Bollen *et al.*, 1993, p. 327; Ragin, 2000, p. 25). Apparently there is a divide between small-n and large-n research designs, but what is small and what is large in that regard? On the lower end there are many single-case studies or studies with a handful of cases, while on the upper end there are also many studies that employ 50 and more, and in case of survey data, thousands of observations. Given that we all have finite time horizons and eventually need to produce some research output, researchers typically focus on the depth of their case knowledge when employing a small-n research design while they focus on the breadth of their findings when employing a large-n research design. Studies between those two

12 *Research Design in Political Science*

poles have 10 to 50 observations. For such a study, it becomes less clear whether it should leverage on in-depth knowledge or on its breadth. On the one hand, there are quantitative electoral forecasting models containing less than 15 observations which nevertheless employ the method of statistical control common in large-n research designs (e.g., Bartels and Zaller 2001; Lewis-Beck and Rice, 1992; Norpoth and Gschwend, 2003). On the other hand, qualitative comparative methods such as QCA or fuzzy-set analysis (Ragin, 1987; Ragin, 2000) can accommodate dozens of observations.

At the end of the day, we are interested in why stuff happens in order to provide explanations and improve our understanding of cause-effect relations in the social world. There is, however, a considerable controversy in the literature about how to conceptualize causality. Small-n research tends to be framed as the analysis of necessary and sufficient causal conditions. This entails at least implicitly a rather deterministic (and nonlinear) view of causality. Large-n (but also some small-n) research, however, has it the opposite way around and adopts a probabilistic view of causality, according to which ‘...“causes” are factors that raise the (prior) probabilities of an event occurring ...’ (Gerring, 2001, p. 129). In general, deterministic causes are helpful if we can assume that the relationship between independent variables and our dependent variable is in fact deterministic. They can give us clear guidelines as to what we should be seeing empirically, if they were really true, and help us disentangle the causal net. But when can we really be sure about deterministic causes in political science? On the one hand, nature might be random to some extent. Thus, even if we were able to measure our concepts precisely, we would never be able to completely explain variation in our dependent variables. This is still true even if we were to include all the variables we can ever dream of and specify the potentially non-linear model correctly. In other words, we not only assume that we included, but also modeled, all contingent causal factors correctly. On the other hand there is also the problem that all measures are imperfect. Thus the very act of measuring a theoretical concept, even if we tend to believe in a deterministic causal world, does always introduce some randomness in the analysis.

This controversy is not only relevant from a philosophy of science perspective. It also has important implications for your research design and the interpretation of your results. Think about it this way: How do we deal with a single case or observation that deviates considerably from an otherwise nice causal pattern? If you believe in a deterministic causal world with perfect measures and correctly specified models of

relationships, you will have a serious problem. The single deviant case is evidence against your theory and must lead to its reconsideration. For believers in probabilistic causes, be they small-*n* as well as large-*n* researchers, a deviant case is simply an outlier. However, even if you believe in deterministic causes and perfectly specified theoretical explanations, outlying observations can happen and do not by themselves invalidate your hypothesis simply due to less-than-perfect conceptualizations of your theoretical building blocks and measurement error.

There are several ways to deal with outlying observations independent from the number of observations available. One strategy is to argue that the model is correctly specified and observations deviate from the general pattern because of noisy measures. Another way is to account for outlying observations directly by rethinking your theory. Following this strategy, the deviations from an expected general pattern are of substantive interest rather than produced by our inability to measure precisely. When rethinking your theory, one conceivable strategy would be to try specifying 'scope conditions' (Ragin, 2000, pp. 61–2; Walker and Cohen, 1985) and make explicit under what circumstances we expect certain relationships to hold. Maybe the theorized causal structure does really only hold for a sub-sample of all available observations – given the unit homogeneity assumption (Achen, 2002, pp. 446–7). Another strategy would be to keep all observations but reformulate the expected universal causal relationships by considering interaction effects or non-linear transformations among independent variables. This would allow you to stipulate conditional or non-linear effects of several explanatory or test variables on your dependent variable. In addition, including new independent variables might prove helpful to better account for outlying observations. Probably due to the economy of scale – that is, a single outlier does seem to matter more for a proposed explanation if the number of observations is five rather than 5000 – small-*n* researchers are likely to jump at deviant cases while large-*n* researchers look rather for a quick statistical fix if they care at all about a few outlying observations (e.g., Western, 1995).

Nevertheless, many important theories are framed in terms of necessary and sufficient conditions (see, for example, Dion, 1998; Goertz and Starr, 2003; Seawright, 2002). In trying to bridge the gap between deterministic and probabilistic causal worlds, new methodological approaches develop tests and estimate models of necessary and sufficient conditions, partly within a Bayesian framework in order to avoid falling into a small-*n* trap (Braumoeller, 2003; Braumoeller and Goertz, 2000; Clark, Gilligan and Golder, 2006; Seawright, 2002). Thus the

14 *Research Design in Political Science*

Table 1.2 Typology of research designs

		Type of causal inference	
		Factor-centric	Outcome-centric
Number of observations	Large n	Statistical control, (field) experiments	Forecasting, qualitative comparative analysis
	Small n	Cross-case comparisons, quasi-experiments	Case studies

choice between small-n and large-n research designs is partially independent of whether statistical tests are used, whether correlation- or process-based inferences with data-set or causal-process observations are employed, and of whether one believes in probabilistic or deterministic models of causation.

We thus arrive at a two-dimensional conceptualization of research designs represented by the cells on the main diagonal in Table 1.2 that goes beyond the widely used dichotomy of qualitative and quantitative research. Factor-centric research designs employ the method of statistical control or (field) experiments to disentangle a key causal factor in the causal net if many observations can be leveraged for inferential purposes. Outcome-centric small-n researchers provide an in-depth, within-case study of potential factors and causal processes that explain the occurrence of single events as comprehensively as possible. The off-diagonal cells are not empty, however. On the one hand, focused cross-case comparisons or quasi-experiments can be used in factor-centric research designs if only a few observations are available. On the other hand, there are large-n outcome-centric research designs, which have the potential to describe a phenomenon and forecast future occurrences of this phenomenon using statistical as well as qualitative comparative methods.

An overview of the chapters

In the previous sections we gave an overview of the core problems and different types of research design. In order to get to know our tool-box for developing a well-designed research project, we need to know to what extent the core problems of research design are the same for all of them. The literature is spread along two extremes here. On the one hand, some argue that qualitative research should merely follow a quantitative template of how to do good research as closely as possible and

everything will be all right (King, Keohane and Verba, 1994). This is 'quantitative imperialism'. At the other extreme, there are scholars who portray qualitative and quantitative research as having entirely different logics (see, for instance, McKeown, 1999; Thomas, 2005). Consequently, qualitative and quantitative research designs – be they factor-centric or outcome-centric, small-n or large-n – cannot talk to or learn from one another. This is 'qualitative separatism'. (One could imagine qualitative imperialism as well as quantitative separatism, but these positions are rather rare in contemporary political science.)

In this book, we start from the assumption that research generally consists of a dialogue between theory and data and that all types of research at least face the same problems and challenges. Whether they also lend themselves to the same solutions, however, is an open question that will be taken up in the individual chapters, each of which focuses on one research design issue. The chapters follow a common template. They first start with a specific problem of research design in political science. Second, they explicate the problem, discuss various solutions, and emphasize the typical trade-offs involved in choosing one or another solution. Third, each chapter presents practical guidelines on how to deal with this particular research design issue in actual research. Fourth, they illustrate the use of these guidelines in an example taken from the authors' own research.

In Chapter 2 on 'increasing the relevance of research questions', *Matthias Lehnert*, *Bernhard Miller*, and *Arndt Wonka* define and distinguish theoretical and social relevance. The chapter then focuses on the widely neglected social relevance of research designs. Lehnert, Miller and Wonka deny that there is an inherent trade-off between theoretical and social relevance and show how researchers can generally improve the relevance of their research projects by responding to three questions. Who is affected by what? How can the results be evaluated? Which advice can be offered?

Do you really know what you are talking about? This is *Arndt Wonka's* central question in Chapter 3 on concept specification as a central issue for research design in political science. Although it is not uncommon in the literature that different defining attributes are used to refer to the same concept, Wonka maintains that ambiguous concepts are not helpful in generating research which is expected to yield relevant results. After formulating some hands-on advice on how to avoid conceptual ambiguity, Wonka puts these suggestions to work and applies them to the concept of 'supranationality' as is used in his and other scholars' research on the European Union.

16 *Research Design in Political Science*

In Chapter 4 on ‘typologies in social inquiry’, *Mathias Lehnert* deals with a special case of concept specification. He critically evaluates whether and how typologies can be used for either description or explanation of social phenomena. He thereby develops three criteria by which different ‘types of typologies’ can be distinguished in order to confine the use of typologies to particular purposes. Typologies provide simplified accounts of complex phenomena and can help establish unit homogeneity in both factor-centric and outcome-centric research designs. In addition to providing advice on when and how to use typologies in political science, Lehnert illustrates how typologies can be used fruitfully by referring to his own and other scholars’ work on political institutions and their effects on political outcomes.

Measurement, *Bernhard Miller’s* topic in Chapter 5, is the next logical step following concept specification. Miller highlights both the challenges when devising measures, couched in issues of reliability and validity, as well as the tools that can be employed to address those challenges; here, he focuses particularly on the efficient use of indices as composite measures. He emphasizes the universal role of theory and concepts in any measurement process for any research design. Besides explicitly considering typical trade-offs one faces in everyday research practice, Miller also provides clear advice on how to devise new measures and illustrates them using his research on coalition committees.

In Chapter 6 *Julia Rathke* deals with the problem of comparability and equivalence of measurements when relying on secondary data sources. She argues that increasing the number of observations is no free lunch but requires at least conceptually equivalent measures. Rathke distinguishes between two different strategies in the measurement process to make sure that we arrive at conceptually equivalent measures: increasing the level of abstraction and establishing functional equivalence. After providing some practical advice on how to make data and indicators comparable, Rathke draws on her research on the effects of social capital on political orientations in Germany to demonstrate the practicability of her advice.

Two chapters deal with selection and selection bias – one in quantitative research, the other in qualitative research. In her discussion of selection bias in large-n research, *Janina Thiem* in Chapter 7 deals with challenges that quantitative research often encounters: the universe of cases – that is, the population of interest – is quite large and theoretically well-defined but only partly observable. If those unobservable observations of the realized sample are not randomly distributed, every inference drawn from this sample will be biased. While well-known statistical

fixes – which may or may not be helpful for a given research problem – exist for such situations, Thiem argues that successfully dealing with selection bias is foremost a theoretical problem. She then provides practical guidelines on how to identify and deal with potential selection biases theoretically as well as statistically and finally applies these guidelines to potential selection effects in the analyses of roll call votes in the European Parliament.

In his discussion of case selection and selection bias in small-n research, *Dirk Leuffen* in Chapter 8 focuses on a situation that qualitative research often encounters: the universe of cases is quite large but not well-known or not well-defined. After reviewing Mill's classical methods of agreement and difference and the equally well-known most-similar systems or most-dissimilar systems designs, Leuffen presents theory-guided typologies as a strategy for case selection. He specifically argues that leverage can be increased by narrowing down the domain, focusing on a small set of theoretically interesting cells of the typology, and concentrating on 'hard cases'. He illustrates this strategy with an example from his research on French divided government.

Chapter 9 by *Ulrich Sieberer* focuses on the research design issue of control and discusses some basic theoretical and methodological choices when selecting independent variables and the trade-offs that come with them. He argues that the status of independent variables to control for the influence of alternative factors differs greatly depending on whether factor-centric or outcome-centric research designs are employed, while it makes no difference whether you employ small-n or large-n designs. Sieberer derives a number of practical guidelines and illustrates them using his own work on explaining party unity in legislative voting behavior.

Andreas Dür tackles the challenge of discriminating between rival explanations in outcome-centric qualitative research. In Chapter 10 Dür distinguishes three problems – omitted variable bias, explanatory overdeterminacy, and indeterminacy – and suggests various strategies to meet this challenge successfully: uncovering logical inconsistencies in alternative explanations, increasing the number of observable implications of one's own and rival theories, examining causal mechanisms through process tracing, and selecting additional 'most likely' or 'least likely' cases. After discussing their strengths and weaknesses, he illustrates the use of these strategies in his own research area: the analysis of trade liberalization.

Dirk De Bièvre in Chapter 11 is concerned with the final phase of the research process: what to do with a theory after it has been tested

18 *Research Design in Political Science*

empirically and found wanting? He puts forward a theoretical understanding of falsification – entailing the replacement of faulty hypotheses with new, presumably better ones – and presents guidelines for formulating hypotheses and conducting research so that researchers can make the most of theoretical falsification. De Bièvre draws on a current research project on the effects of judicialization in the WTO to illustrate the use of these guidelines.

Finally, in the concluding chapter we are concerned with the lessons that can be learned for improving the dialogue between theory and data. While all types of research face the same challenges there is no cookie-cutter approach to help us dealing with them. Instead different research designs offer and require different solutions, each of which produces specific trade-offs. How you evaluate these trade-offs should determine how you carry out your research project and, consequently, the type of research design you choose.

Notes

1. For similar definitions, see Brady, Collier and Seawright (2004, p. 302); de Vaus (2001, p. 9); or King, Keohane and Verba (1994, p. 118).
2. For a similar list drawn from King, Keohane and Verba (1994), see Collier, Mahoney and Seawright (2004, pp. 36–7).
3. In reality, however, designing research is rarely so neatly ordered. It does not always start at the beginning of the process or finish at its end, and it involves a lot of going back and forth between the design problems.
4. For similar distinctions see Ganghof (2005a); Gerring (2001, p. 137); Scharpf (1997, pp. 24–7).