Research Design

Dimiter Toshkov, Leiden University

(d.d.toshkov@fgga.leidenuniv.nl)

Book chapter written for the new edition of Marsh, Stoker and Lowndes'

'Theory and Methods in Political Science' (Palgrave Macmillan)

[Version 1 (First draft). Last revised: 1 November 2016]

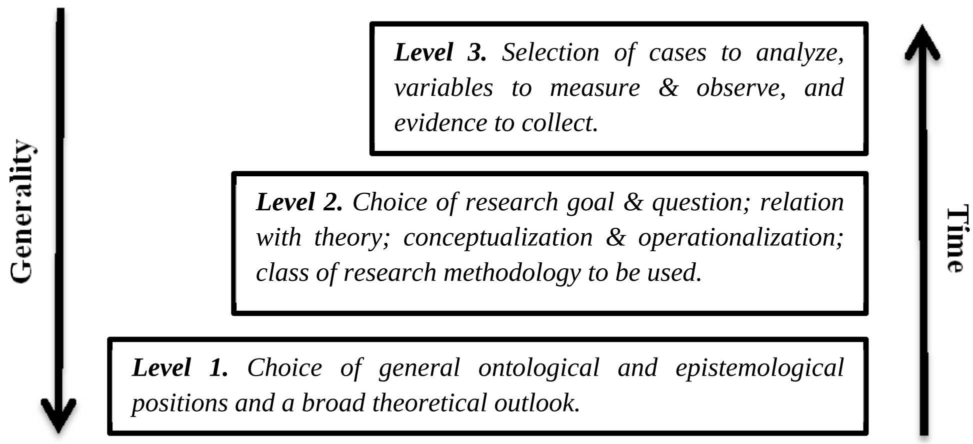
What is research design?

Part I of this book explained that the field of political science hosts a variety of theoretical approaches. Part II introduces a similarly rich set of methods that political scientists use to learn about the social and political worlds. Whatever the preferred theories and methodologies, at some point researchers face choices about how to set up their study. These choices relate to issues of research design. This chapter will present the basic elements of research design for political science projects, and it will discuss how different research designs approach the common challenges of scientific inference.

In essence, research design is about getting valid answers to research questions in a reliable and efficient way. It is about maximizing the validity and scope of application (generalizability) of scientific inferences given the goals of the researcher and constraints of practical and ethical nature. From one perspective, research design can be considered applied epistemology as it deals with the big question `How do we know?' at a more operational level than philosophy. From another perspective, it is a branch of the decision sciences as it is about making optimal choices under constraints. Putting research design in these terms, however, makes it appear far too scientific than it actually is, at least for the time being. Research design remains as much a craft as a science.

In the scientific process, research design choices are made at three levels of generality. At the first, most general, level research design is about the adoption of certain general ontological and epistemological positions and a broad theoretical outlook. For example, it is about whether one approaches the problem of political inequity from an interpretivist or positivist, Marxists or feminist vantage points. (Note that the ontological, epistemological, and theoretical vantage points also *direct* the researcher's attention towards some research questions at the expense of others. Consider that gender-based political inequality is a much more central problem for feminist theory than it is for classic Marxism, for instance.)

Figure X.1 *Three levels of generality of research design considerations*



At the second, less abstract and more operational, level of research design we choose a research goal, formulate a precise research question, clarify the role of theory in the project, conceptualize and operationalize the concepts, and select the class of research methodology to be used. For example, we might (1) choose to research the question 'What is the impact of the electoral system on the political representation of women?', (2) which implies an explanatory goal of identifying a causal effect; (3) specify that the research interest is in *testing* this existing hypothesis derived from institutional theory; (4) settle for an observational (non-experimental) design based on a large number of cases to be analysed using quantitative (statistical) methods; and (5) identify observable variables that capture the conceptually relevant dimensions of electoral systems and representation, such as the level of proportionality of the electoral system and the share of women elected in the national parliament.

Once these choices are made, at the third level research design concerns even more concrete and specific issues, such as the selection of cases to analyse, variables to measure and observe, and evidence to collect. To continue our example, we can decide to study 60 countries, sampled randomly from the list of all sovereign countries in the world, and collect data for each of these countries about the level of proportionality of their electoral system, the share of women in their parliaments, and several additional variables, such as the type of political system and predominant religion, which might be needed in order to answer our original research question.

This chapter focuses on research design at the second and third levels of generality. It assumes that you have already adopted a general ontological, epistemological, and theoretical outlook and explains how one goes about designing a research project once these choices are made (to the extent that one can *choose* his or her own epistemological skin, see Chapter X).

In principle, research design stops where data collection and analysis begin, but one often needs to go back to the design stage in view of problems or opportunities arising during data collection and analysis (see Figure X.2 below). But it is important to emphasize that this chapter does not deal with *methodologies* of data collection and analysis: these are covered in Chapters X to X.

Research design in political science is relevant for all research projects unless they have purely theoretical or normative goals. Normative and positive theory development research have their own methodologies (see Chapter X in particular), but it is hard to speak of their research designs as such, especially when we focus on the more operational issues of research design that we deal with in this chapter.

In any case, research design is relevant for all research projects that engage, one way or another, with the empirical world, irrespective of focus (case or hypothesis), mode (experimental or observational), type of data used (quantitative or qualitative), and method of inference (cross-case or within-case). That being said, it is more common to discuss issues of research design in the context of positivist and realist epistemological traditions, rather than from an interpretivist point of view (for an exception, see Schwartz-Shea and Yanow 2012).

Inference, validity, and research design

We said in the beginning of this chapter that research design aims to provide valid answers to research questions in a reliable and efficient way. But we did not explain yet what we mean by validity. And to get to the concept of validity, we have to introduce the concept of scientific inference first.

Empirical scientific research is about making inferences from something that can be observed to something that cannot be observed directly. In this way science expands what we know about the world beyond what is directly available to our human senses. There are various types of scientific inference. We can mention some of the most important ones (summarized in Box. X.1).

First, we make inferences from observable data to unobservable theoretical concepts. This is the process of scientific measurement, broadly defined to include aspects of conceptualization and operationalization (see below). For example, we can observe what politicians do and what they say in public, but we want to know their ideology or competence – abstract concepts that are not directly available to the senses.

Box X.1 Types of scientific inferences and validity (a non-exhaustive list)

- Measurement *(from observed variables to abstract concepts)*
- Population-level *(from observed cases to broader populations)*
- Statistical *(from observations on sets of variables to systematic associations)*
- Causal *(from established associations to causal relationships)*
- Interpretive *(from observed actions to meanings)*

Second, we make inferences from cases (individuals, organizations, practices, etc.) we observe to broader unobserved populations of which these cases are part. This is a kind of descriptive inference; in particular, from a sample (the cases we observe) to a population. For example, we can interview our friends about their political preferences, but we might be interested in a broader population, such as all young people with voting rights in the country. (There are other kinds of descriptive inference as well, such as from an observed set of characteristics of a unit or a population to the underlying structure of the unit or the population.)

Third, we make inferences from observing associations between concrete phenomena and variables to causal relationships between theoretical constructs. This is the process of causal inference that plays a central role in contemporary social science (King et al. 1994, Morgan and Winship 2015). For example, we might observe that lower-educated people are less likely to vote in parliamentary elections. But we might be interested whether education as such *affects* (rather than merely correlates with) participation in the electoral process, and this requires causal inference. In fact, establishing associations in the first place requires inference (often called statistical inference) in order to separate systematic from chance associations. And then there is the related but separate inference from observations made in the past to the future (predictive inference).

Fourth, we make inferences from observed concrete actions of individuals to the meaning of these actions. For example, we might observe a citizen throwing a Molotov cocktail (an improvised petrol bomb) towards a government building, but to understand the meaning of this action and to make sense of it in its social context, we need to make an inference. This is the kind of inference that interpretivist research is mostly engaged with.

Importantly, in science we pose validity requirements to each kind of inference. Therefore, we speak of the validity of measurement inference, or measurement validity for short; of descriptive validity; of causal validity; and of validity in interpretation. In the methodological literature you can find the popular distinction between internal and external validity (Cook and Campbell 1979, Shadish et al. 2002). Internal validity usually refers to what we just called causal validity *or* to measurement validity, while external validity refers to the validity of any inferences, descriptive *or* causal, to broader populations. Because of these ambiguities, we better avoid the terms internal and external validity and use the more extensive class introduced above, which explicitly evokes the type of scientific inference we want to make. In principle, a research project might target one type of inference only, but in practice it will typically combine several in the process of research.

Coming back to our main topic, we can now say that research design aims to deliver valid inferences at each stage of the research process and with a view of the stated goal(s) of the research project, be that measurement, description, causal explanation, interpretation, or something else, such as prediction or classification.

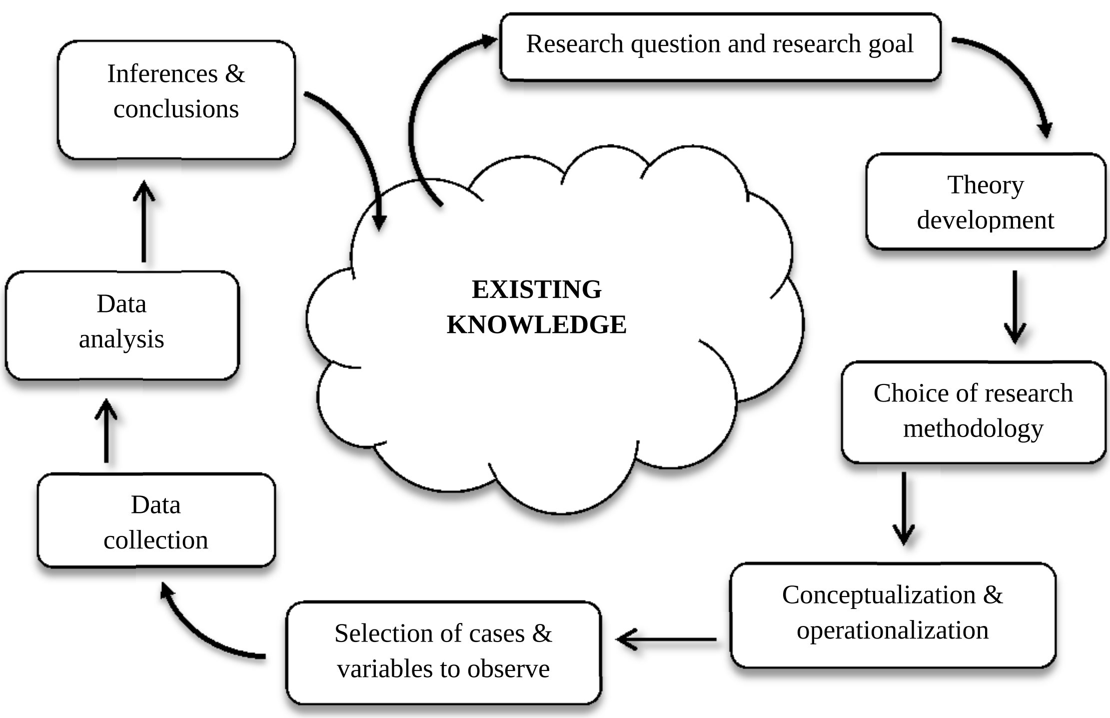
To remind, inferences must have high validity, but inferential validity better be achieved in an efficient way. If the same level of validity and the same scope (generalizability) of inferences can be achieved through different research designs, the one that demands the least amount of resources is to be preferred. In the real world researchers face numerous constraints in doing their work – financial, organizational, technological, ethical, and others. It is important that the design of research projects delivers high validity considering these practical constraints; hence, as efficiently as possible.

The research process

Before we are ready to discuss the design principles of individual stages of the research process, it will be helpful to get a bird’s eye view of the process in its totality. Inevitably, such a view must remain rather abstract in order to fit the multitude of research projects political scientists engage with. So think about the scheme represented in Figure X.2 as an organizing device rather than a faithful description of how political science works. And bear in mind that individual research projects, especially the relatively small-scale projects that students are expected to deliver, need not cover all stages of this research process at once.

In the beginning, there is the research question. Of course, research questions do not pop up into being straight from the void, but are conceived in relation to existing knowledge. In fact, the whole point of research is to expand the boundaries of existing knowledge. The research question already sets the research goal. Once the question and the goal are relatively clear, it is usually time to settle on the broad type of research methodology that is going to be employed. What typically follows is the stage of theory development.

Figure X.2 *A schematic representation of the research process*



The output of this stage is often a set of hypotheses, but it could also be a more unstructured elaboration of the theoretical context of the research question, or even the realization that there is not much theory available to guide the project at all. In any case, theory points to the most important concepts to study, and the next stage engages with the conceptualization and operationalization of these concepts into observable variables.

What follows is the selection of cases to observe, the variables on which to collect data, and the types of evidence to search for. As explained above, all stages up to this point can be considered part of research design. Afterwards, the stage of data collection ensues, and once it is complete, data analysis can begin. The final output of this stage are the inferences and conclusions of the research process that feed back into the body of exiting knowledge.

The links on Figure X.2 represent both logical and temporal relationships. While some projects travel the full cycle from research question to inferences through every stage, other projects might focus exclusively on one stage, such as theory development or conceptualization or data collection. But even when they do, individual projects contribute to the collective scientific enterprise that moves along the stages represented in the figure.

The figure omits a host of links of secondary importance, such as the contributions of existing knowledge to theory development and conceptualization (and vice versa) or the iterative loops between the design sages and data collection, on the one hand, and data collection and data analysis, on the other. Nevertheless, it will do for our purpose of providing a bird's eye overview of the research process. The figure is general enough so that it can serve a variety of research goals, like description of populations or explanation of individual cases, and is consistent with different epistemologies (although derived from a broadly-defined positivist outlook).

Now that we have the general overview of the research process, we can discuss in more detail each stage in the context of research design considerations.

The elements of research design

Research questions and research goals

It is easy to say that research starts with a question. It is more difficult to answer where good research questions come from (for some solid general advice, see Gerring 2012a: 37-57, King et al. 1994: 14-19, and Lehnert et al. 2007: 21-40). Inevitably, to select a research question that can advance the state of existing knowledge, one needs to figure out what we already know about the topic of interest. To push the boundaries of knowledge, we need to know where these boundaries lie.

Interest in a topic can come from a wide variety of sources, including personal dispositions, values, and experiences, academic scholarship and literary fiction, current events, and, lest we forget, direct assignment from someone who commissions the research. Whatever the source of interest in a topic, only a diligent review of the relevant scientific literature would reveal what the current stage of knowledge is (the fabled 'state of the art'). Fortunately, these days it is easier than ever to explore the state of the art in political science using free online bibliographical databases and encyclopaedia, comprehensive literature reviews published in academic journals, or blog posts by researchers active in a field. (At the same time, the amount of research being done on any topic is also growing at an unprecedented rate, which makes staying on top of the existing literature harder.)

Despite the enormous varieties of ways in which a research question can be encountered and phrased, there are several typical constructions. First, it is now popular to frame research questions as puzzles to be solved (Grofman 2001). A puzzle is a pattern of empirical observations that does not make sense in light of existing theory and knowledge; for example human cooperation in light of (a simple version of) rational choice theory (see Chapter X). Puzzles are relevant, but so are substantively important problems, which cannot always be construed as real puzzles. In short, research questions can be motivated by a pattern of empirical observations that cannot be accounted for *or* by theoretical concerns that have not been addressed empirically in a satisfactory way yet.

Second, when it comes to explanatory research (see below), researchers use three typical forms. They ask about the causes of an effect, about the effects of a cause, or about a particular relationship between a pair of concepts. Examples of the three types would be, respectively: ‘What caused the United Kingdom to vote for leaving the European Union in 2016?’, ‘What are the effects of globalization?’, and ‘What is the effect of changing income inequality on support for European integration?’.

The research question is intrinsically connected to the research goal. In its turn, the research goal is one of the major considerations when choosing a research design and methodology. Even when we exclude purely normative and theoretical research, scientific projects can still set a variety of different goals (including predictive, exploratory, interpretive, and problem-solving ones). We can only discuss two of the most prominent in this chapter: description and (causal) explanation.

Description

Etymologically, the word ‘description’ comes from the Latin *describere* with the meaning ‘to write down, sketch, represent’. Scientific description also deals with writing down (listing) features of units (cases) and representing these units in a narrative, a survey, a collections of categories, or a set of numbers. Description can be about a comprehensive (yet never complete) representation of a single unit or event, but it can also be about classifying many cases along a single or a small set of variables. If the classification is into a large number of equally spaced categories, the process is also called ‘measurement’.

What separates casual observation from scientific description? First, scientists strive to use reliable and precise, public and well-documented descriptive procedures and measurement instruments. Second, the value of scientific description is in selecting a set of important analytical dimensions on which to project the empirical world (Toshkov 2016: 123). These analytical dimensions should give insight into the essential nature of a case and its relationships with other cases. Third, scientific description is attentive to the representativity of the cases being described to larger populations of interest. In fact, one of the major purposes of research design for descriptive projects is to make sure that the cases that will be studied will deliver valid inferences about a population from which they have been drawn.

Contrary to what some political scientists believe, description is a worthy scientific goal in its own right (Gerring 2012b). It involves inference (as we explained above), it is linked with theory (which suggests the important case features to be measured and described), and it can have important normative implications as well (for example, when we classify a state as a democracy, or a conflict as a civil war).

There are many, widely different, methods in which to do (collect and analyse data for) scientific description: participant observation, (semi) structured interviews, document analysis, archival research, but also various statistical techniques for surveying, data reduction, and the compact representation of systems of associations. Public opinion surveys (such as the World Values Survey or the European Election Studies), which are perhaps the most recognizable products of political science research, are in essence descriptive projects. (Of course, scholars later use data from these surveys to build and test explanatory models as well.) Examples of other important recent descriptive work in political science includes the mapping of interest group populations across countries (e.g. Gray and Lowery 2000), the collection and classification of public policy outputs related to agenda-setting (Baumgartner et al. 2009), the measurement of political party positions based on a variety of data sources (such as party manifestos, voting records and/or expert surveys) (e.g. Bakker et al. 2012, Budge et al. 2001), the classification of the institutional features of political regimes across the world (see the Polity IV and V-DEM projects), the collection of data on international conflicts (e.g. the international UCDP/PRIO Armed Conflict Dataset), and many more. Dense histories of political events and the evolution of public policies provide research in radically different modes, but one that still remains descriptive in nature.

Explanation

Explanation is the second major research goal we will present in some detail. Explanation is a subject of long-standing and sophisticated debates in philosophy and the sciences, including statistics (see, among others, Salmon 1998, Woodward 2003, Lewis 1986, Pearl 2009). Reviewing these debates is not a purpose of this chapter. Suffice it to note that in contemporary political science the focus is on *causal* explanation (Elster 2007), which is often understood in *counterfactual* terms (Lewis 1986). Causality is taken to be *probabilistic* rather than deterministic (so there are no real laws in social science) (Glymour 1998), but it is still based on causal *mechanisms* (Little 1991). A cause is a ‘difference maker’ (Lewis 1973) so that if the cause would have been different, a different outcome would have occurred (in probability). Such a view of causal explanation is compatible with various research methodologies, from within-case analysis to experiments (see below), but is by no means consensual even within the domain of political science. Constructivists, for example, prefer to focus on constitutive rather than causal arguments altogether (Chapter X), and radical subjectivists question the very existence of causality in the social realm. Other scientists (including many statisticians) are uncomfortable with the idea of causes that cannot be directly manipulated (Holland 1986).

Even within the parameters of our broad definition of causal explanation given above, there is room for different explanatory research goals. We might want to explain a single event (for example, ‘What led to the Cultural Revolution in China in the late 1960s?’) or test for a systematic causal relationship (‘What is the effect of economic prosperity on democratization?’). We can focus on causal mechanisms (‘*How* does international mediation help solve military conflicts between states?’) or on estimating causal effects (‘What is the average effect of facial attractiveness on the electoral success of political candidates?’). We can ask, ‘What would be the effect of lowering the voting age on turnout in Ukraine?’, which would be a *prospective* causal question, and ‘What was the impact of social policy on the integration of demobilized paramilitary and guerrilla fighters in Colombia?‘, which would be a *retrospective* causal question. Further distinctions are possible, such as the ones between questions about the causes of effect versus effects of causes (Gelman 2011) that we already introduced earlier in the chapter.

Interpretation

It is difficult to position interpretation (in the narrow sense of the type of work interpretivist political scientists engage in) between description and explanation. Clifford Geertz notes that (ethnographic) description is interpretive (Geertz 1973: 20), but that still leaves the question whether all interpretation is descriptive open. Bevir and Rhodes (2016) insist that intepretivists reject a ‘scientific concept of causation’, but suggest that we can explain actions as products of subjective reasons, meanings, and beliefs. In addition, intentionalist explanations are to be supported by ‘narrative explanations’. In my view, however, a ‘narrative’ that ‘explains’ by relating actions to beliefs situated in a historical context is conceptually and observationally indistinguishable from a ‘thick description’, and better regarded as such.

It is important to be aware of the different types of descriptive, causal, and interpretive questions because they give freedom to hone your research question to your actual research interests. But also because different goals call for and tend to go together with different research designs and methodologies.

Theory and empirical research

After we choose the research question and clarify the research goal, we have to examine how the proposed research project relates to existing theory. Theory in political science serves many purposes. It collects and integrates what we have learned about the world; it provides the major concepts and ideas that we use to discuss and analyse politics and governance; it specifies the causal mechanisms that we assume to hold; it identifies empirical puzzles, gaps in our knowledge and open questions that need more research; and it provides hypotheses and predictions about the world.

Research projects engage with theory in different ways (see Chapter 3 in Toshkov 2016 on which this section heavily draws). (1) Some projects are interested entirely in theory development without any recourse to the empirical world. One can use a variety of tools, such as game theory, agent-based modelling, simulations, mathematical analysis, formal logic, thought experiments, and more to conduct such deductive theory analysis and development. (2) In projects where theory meets empirics, we are sometimes interested in testing a proposition about the observable world, derived from a theory, in order to see whether it is corroborated by the empirical facts or not. Such a proposition is called a hypothesis. (3) Other projects might be interested in testing a hypothesis not because it is derived from theory, but because it is substantively important. Such projects would still need to elaborate the theoretical context of this hypothesis. (In other words, they would need to reconstruct the causal model of the phenomenon of interest in which the hypothesis is embedded).

Box X.2 ***Types of engagement of research projects with theory in political science***

- Theory development *(purely deductive from premises to propositions;, no input from empirical data during the process, only at start)*
- Theory generation *(inductive, summarizing patterns in empirical data in order to provide theoretical ideas and staring points)*
- Theory testing *(probing theoretically-derived hypotheses against empirical data to examine the veracity and scope of the theory)*
- Hypothesis testing *(matching a substantively-important hypothesis embedded in a theoretical causal model against empirical data)*
- Theory application *(routine application of existing theory to new empirical cases)*
- Abductive explanation of individual cases/events *(iterative testing of competing hypothesis against empirical evidence about a case/event with the aim of identifying the most plausible explanation; theory provides source of hypotheses and knowledge of general causal mechanisms)*

(4) Yet other projects might be interested in producing theoretical ideas through inductive empirical analysis: studying patterns of empirical observations with the aim to generate theoretical ideas that can be later analysed further and eventually tested. (5) Some projects of more modest goals might want only to *apply* an existing theory to a particular, previously unstudied case, without necessarily testing old or generating new theories. (6) When we are interested in explaining the outcome of a single case, theory provides the initial competing hypotheses and the knowledge of causal mechanisms that we investigate in a process called logical abduction (Peirce 1955). Abduction involves the iterative examination of competing hypotheses against empirical evidence about a case or event until only one hypotheses remains as the most plausible explanation, as it is not contradicted by any of the evidence we have assembled. Box X.2 summarizes the ways in which research projects engage with theory in political science.

Whatever the exact terms of engagement between a research project and theory, theoretical reflection is always necessary, even if the goals and ambitions of the project are predominantly empirical and exploratory. This is because theory not only provides hypotheses and elaborates the assumed causal mechanisms, but it also supplies the main concepts of research.

Conceptualization and operationalization

Concepts are the building blocks of scientific arguments and ideas. As they are by definition abstract (and it is their generality that makes them useful), they are not directly observable. It is one of the tasks of research design in empirical political science to translate these concepts into variables that can be measured, indicators that can be detected, and evidence that can be encountered in the real world.

Concepts are typically defined in terms of a set of necessary and sufficient conditions (Braumoeller and Goertz 2000). For example, a *limited access social order* (North et al. 2009) can be defined as a social order in which participation in politics is limited, access to economic resources is severely constrained, and political power is concentrated into a small elite. These three conditions then would be individually necessary and jointly sufficient for an empirical case of a social order to fall into the concept's definition, or not.

Concepts in political science are often complex and hotly contested – think about power, democracy, equality, or ideology, for example. This makes it even more pressing that their intended meaning, attributes (intension), empirical scope (extension), and classification structure (how the various objects that fall under the concept's definition are grouped together, Toshkov 2016: 95) are clarified in a process called conceptualization. Conceptualization prepares the ground for the subsequent step of operationalization, in which the now clearly-defined attributes of concepts are connected with observable variables and indicators (Adcock and Collier 2001, Goertz 2006, Sartori 1975). In the process of actual research, during the data collection phase, the cases will be assigned values on these variables and classed into categories according to the indicators.

Operationalization is also relevant for research that eschews the language of variables and measurement, such as in-depth single-case studies and interpretive work. In these research modes it is more appropriate to think of operationalization linking concepts with pieces of empirical *evidence* that would be searched for and collected in the process of research.

As part of research design, all concepts relevant to the project must be conceptualized and operationalized. Ideally, already in the design stage the researcher should also specify how the necessary data would be discovered and collected.

To recap, so far we discussed the following elements of research design: the choice of research goal and question, the clarification of the relationship with theory, and the double process of conceptualization and operationalization. What remains is a discussion of the type of research methodology that the project will employ and the third-level issues of case and variable selection (see Figure X.1).

Types of research methodologies

The choice of the broad type, or class, of research methodology that a project will use is already more or less clear once its precise question is formulated, its goal is set, and its theoretical ambition is explicated. In principle, there is no logical reason why the choice of methodology must precede conceptualization and operationalization (cf. Figure X.1), but in practice the way we operationalize the concepts is closely linked with the type of data collection that would follow, and, by implication, with the type of research methodology. And, to repeat, researchers go back and forth between the design stages to make sure the elements fit together. So let's just say that the choice of research methodology is made during the design phase of a project and is completed before data collection begins.

This book contains in-depth chapters on the most common research methodology classes in political science, so this chapter will only explain how they relate to each other and highlight their features that are important for the purposes of research design.

Experimental methods (see Chapter X, as well as Druckman et al. 2011, Kittel et al. 2012, and Morton and Williams 2010) are in a class of their own due to the control researchers have over some aspects of the environment during the study. To qualify as a true experiment, or a randomized controlled trial, the researchers must be able to control one crucial design aspect: namely, the assignment of experimental units (the cases) to different groups according to their status in the experiment, typically, 'treatment', which receives the experimental manipulation of interest, and 'control', which does not. Assignment is done randomly in order to ensure the comparability of the groups with respect to any possibly relevant attribute other than the experimental manipulation.

When researchers do not control the assignment of the experimental manipulation, the experiment technically does not qualify as such, even if other aspects of the environment are controlled. Conversely, sometimes nature provides (almost) random assignment of units into different conditions, and in that case we speak of natural experiments (Dunning 2012). From a design perspectives natural experiments are similar to real ones, even though researchers have no control over any aspect of the situation.

Research that is not experimental is called observational because the researchers are left to collecting observations that the world provides in the course of its undisturbed operation. In observational research, which is a vast category with many different types, what the researchers do control is what to *observe*[1] in a particular project. Hence, the major decisions they make relate to the issues which and how many cases to observe, which attributes of these cases to observe, and what evidence about the cases to collect.

When many cases are observed and analyzed, we speak of large-N designs (the 'N' stands for the number of observations) and quantitative methodology. Data gathered in the context of large-N projects is usually analyzed using statistical techniques derived from probability theory. While large-N designs have many cases (the units), they usually have data only about a small number of aspects of these cases (the variables).

Although large-N projects can have millions of observations and use sophisticated statistical methods for data analysis, in essence they rely on comparisons across the cases to derive causal inferences. In that sense, they are not fundamentally different from comparative studies that use a smaller amount of cases (sometimes even as few as two, as in paired comparison designs) and no statistical analysis. With a greater number of observations, large-N designs are better at filtering random unsystematic noise out of the empirical data to detect the systematic 'signal' of interest, for example a descriptive association, a predictive effect, or a causal relationship. But their ability to provide causal inference, in particular, comes from design, and not from the number of cases.

This holds true for big data methodologies as well (see Chapter X), although currently big data projects are usually deployed for purposes other than causal analysis and explanation: classification, measurement, prediction, and other applied goals such as process control.

Comparative research projects (Chapter X) with a smaller number of cases are similarly dependent on their design to derive valid inferences. Their leverage comes the careful selection of the set of cases to study, so that the similarities and differences across the

[1] Note that by observation we mean not passive reception but the active, purposeful, and often laborious measurement, classification, and collection of empirical facts.

cases can support or contradict causal hypotheses about suspected causal effects. Small-N comparative designs typically use a combination of quantitative measures and qualitative information to describe the cases they study. The data they collect for every case is much richer than in large-N studies, not only in terms of scope (more variables per case), but also in terms of depth.

When the number of cases we study goes down to one, we speak of single-case design, which are affiliated with qualitative methodology (Chapter X). (Qualitative methodology is a fuzzy label that covers various modes of research, such as interpretivist and constructivist work, small-N comparisons, broadly positivist but descriptive case studies, and more. The commonality is that they all avoid quantification. But from a design perspective, whether cases are described and compared with numbers or with qualitative categories does not matter too much. What matters is whether inferences are based on comparisons across cases or not.)

By definition single-case designs cannot rely on comparisons across cases for inference. Instead, they rely on within-case evidence only. Within-case evidence can be used to examine competing theoretical hypotheses. It is also useful to trace the process through which a causal effect obtains and the mechanisms behind a causal relationship (Weller and Barnes 2014). Via theory, single-case studies can also contribute to the study of general relationships and their contribution need not be restricted to the understanding of the particular case being observed (Blatter and Haverland 2012).

Research projects in political science also can, and do, *mix* designs and methodologies (Lieberman 2005, Small 2011, Rohlfing 2008). The combination can be done sequentially, for example when a smaller number of people are interviewed in-depth after a large standardized public opinion survey or when a pilot qualitative study identifies appropriate measures that are later collected in a large-N dataset. There are other interesting ways to mix methods and designs to combine their respective strengths and alleviate some of their weaknesses. But it should be noted that mixing only works for methodologies part of the same epistemological paradigm.

Case and variable selection for different types of research

So far we covered issues of research design at the second level of generality (Figure X.1). The next step is to discuss even more operational issues such as case and variable/evidence selection. Because these issues are rather specific, there is little that can be said about them outside the context of particular research goals and methodologies. That is, case and variable selection strategies depend crucially on the goal of the research project and on the methodological design. Accordingly, we will proceed to discuss the issues of case and variable/evidence section in the context of the major types of research methodologies introduced above and with reference to the different research goals outlined earlier in this chapter. To remind, the purpose of case and variable selection, as elements of research design, is to support the reliable and efficient discovery of valid inferences.

Experimental research

In experimental research case selection occurs at two moments. The first moment is when the cases (units, most often individuals) that are going to take part in the study are selected from some wider population. This is the sample selection stage. To make sure that the sample is going to be representative of the population, ideally the sample case should be selected *randomly* from the population, so that each unit in the population has exactly the same chance of ending up in the sample. This is done to ensure that the results of the experiment conducted on the sample will be generalizable to the broader population.

Despite its methodological attractiveness, random selection is not always feasible or available to the researchers. Experiments are often run on ‘convenience’ samples, such as students or panels of online volunteers. This opens the door to potential bias, so that any results from the experiment would not be relevant for the broader population, unless we are willing to assume that all people would react in the same ways to the experimental manipulation as have the students or online volunteers in our samples.

The second moment when case selection issues come into play in experimental research is when the sample is divided into ‘treatment’ and ‘control’ groups. As explained above, the assignment of units to experimental conditions (treatment/control) must be done randomly so that every unit has exactly the same chance of ending up in any of the groups. Random assignment is crucial for valid causal inference because it helps to rule out all alternative explanations of any difference between the groups we might observe after we apply the experimental manipulation.

In principle, in the presence of random assignment researchers need not collect additional data on any variables other than the status of the sample units (treatment or control) and their response to the experimental manipulation. In practice, however, researchers would gather additional data on variables that possibly confound the relationship between the treatment and the response, that can shed light on the mechanisms through which the treatment works (if it does), and that can account for any differences (heterogeneity) in the way the treatment work across subgroups. Some of these variables can be used to ‘assist’ random assignment by forming homogenous groups in which randomization is applied and/or to adjust the results during the analysis stage (Cox and Reid 2000, Imbens 2011).

In the design phase, researchers should also decide on the number of cases that will be recruited to participate in the experiment. This issue is related to the power and sensitivity of our experiment to detect the hypothesized relationships in the data. The required number of cases needed depends on the likely strength of the hypothesized relationship (the stronger the hypothesized effect, the fewer cases needed), the within-group variability of the population (the greater the variability, the more cases needed), and some other factors of lesser importance. Once the researcher can estimate these factors, the numbers can be plugged into a statistical power calculator that would suggest the required number of cases for the experiment (Duflo et al. 2006). Underpowered studies, no matter whether experimental or observational, can produce very biased and misleading inferences (Gelman 2009).

Large-N observational research

When random assignment is not available, political scientists have to use other strategies to derive valid inferences from data. As with the generalizability of experimental research results, in the context of large-N descriptive inference, random selection of the cases to observe is still crucial for the validity of the inferences to a broader population of interest. Even in professional large-scale surveys, however, sample selection is never that simple. Because of practical considerations, such as the need to reach respondents efficiently or the requirement to include a sufficiently large number of respondents from small sub-groups within the populations, sample selection is typically done with a mixture of purposeful and random sampling (Lohr 2009, Valliant et al. 2013). While the details of these procedures are too involved for this chapter, it is important to realize that when sampling is not random and some groups are systematically over- or under-represented in the sample, the results from the research on the sample would not be generalizable to the target population.

Selection bias can also mar causal inferences with observational data. But it is but one possible source of bias and but one manifestation of the general difficulties of deriving causal inferences from observational data. In general, the problem arises from the fact that causality is not directly observable. Associations between variables X and Y that we can observe can arise under a number of different scenarios, including (1) X causes Y, (2) Y causes X, (3) Z causes both X and Y (so Z is a *confounding* variable), (4) conditioning on a variable caused by both X and Y, and (5) chance (see Glymour 2006). That is, all five scenarios would produce observationally equivalent data even when the underlying data-generating mechanisms are very different.

Research design comes to the rescue. There are three different logical strategies one can use to try to solve the general problem of causal inference sketched above (Pearl 2009). First, we can try to measure all confounding variables during the data collection stage, which would allow for estimating the residual 'true' causal effect of X and Y during the data analysis stage. Second, even if we cannot measure all suspected confounders, we can try to identify the effect of X and Y via the mechanisms through which it works (e.g. X leads to M which leads to Y). This is called mediation analysis. Third, if we can find a variable I that strongly influences X, but is not affected by the confounder Z and does not directly affect Y, we can also identify the effect of X on Y by including measures of I in the analysis. In this case I would be called an 'instrumental variable', and the causal identification strategy is one based on instrumental variables.

In the practice of large-N research in political science, researchers most often rely on the first strategy, and implement it in a number of different designs: time series (where a single unit is observed over time), cross-sectional (where a set of units is observed at the same point of time), pooled (which combines variation over time and across units), and more complex multilevel ones.

To recap, depending on the causal identification strategy, large-N projects need to plan to observe different sets of variables in addition to the hypothesized cause and effect of interest. For conditioning strategies, these include all confounders; for mediation analysis – the mediating variables; and, rather obviously, for instrumental variables – the instrumental variable itself. Note that the status of variables as confounders, mediators, or instruments cannot always be verified against empirical data and has to remain to some extent based on theoretical assumptions (Pearl 2009).

Coming back to case selection, when random selection is not available, it is advisable to ensure that the sample of cases has ample variation on both the outcome variable of interest and on the main explanatory variable (King et al. 1994). This advice presumes that the values of the outcome variable are known in advance, before the actual data collection has taken place, which is often not the case. In such situations, cases that exhibit variation on the explanatory variable should be selected. Selecting on the values of the outcome (dependent variable), so that all cases exhibit a single outcome, is not advisable outside the context of preliminary, exploratory studies, because it can lead to biased causal inferences.

When it comes to the number of cases to be observed, the same considerations about design sensitivity and power that we introduced above in the section on experiments apply. Increasing the sample size, however, often comes at the price of increasing its heterogeneity (how different the units in the sample are). Heterogeneity is bad news for estimating causal effects, and the problems it introduces may not be worth the increased statistical power that comes with a larger number of observations (Rosenbaum 2005).

Comparative research

In small-N comparative research the selection of cases can be approached differently depending on the goal. For theory-testing projects, the most straightforward strategy is to keep the cases similar across all theoretically-relevant variables and let only one variable – the one related to the hypothesis of interest – vary. In this way alternative explanations are controlled for by being kept constant across the set of comparison cases. For theory-generation projects, one can adjust this strategy so that a set of similar cases that exhibit *different* outcomes of interest are compared (Przeworski and Teune 1970). (Confusingly, both of these strategies can be referred to as most-similar system designs, but they serve quite different research goals.) One can also find a set of different cases that happen to share an outcome and use them with the intention of identifying a crucial similarity across the cases that can account for the shared outcomes. This is called 'most-different system design', and is also a strategy better suited for exploration and theory generation rather than theory/hypothesis testing.

Single-case studies and within-case analysis

Experimental and large-N observational projects rarely care about the individual cases part of their samples as such. The cases are valuable only to the extent that they can serve to support descriptive or causal inference that are about a general relationship and/or a broader population of interest. But in political science we are often interested in particular cases because of their substantive, real-world importance. For example, we want to know not only about democratic survival in general, but about the chances of India to remain a democracy; not only about government formation in general, but about the duration record-breaking government formation process in Belgium in 2010-2011; not only about the appeal of populism, but about the astonishing campaign of Donald Trump in the 2016 presidential elections in the USA. When we are interested in a case because of its substantive importance, the research design concerns about case selection seems misplaced. We select the case to study, because it is, well, the case we want to study. And that is fine, as long as we do not intend to generalize the findings based on that one case to others.

When we are interested in explaining, rather than just describing, a single case, we have to do that with recourse to more general theories. When such general theories exist, they provide hypotheses about possible explanations of the case. The task of research design is then to anticipate what kinds of empirical evidence will support or contradict these hypotheses, and to devise plans to search for such evidence. For example, if the two main competing hypotheses we have formed, based on prior research, about the onset of the 2011 Jasmine Revolution in Tunisia are that it results from: (1) the deteriorating economic conditions, or (2) increasing political repression, we would want to collect evidence about the evolution of the economic situation and the state of political freedoms in the country, and seek to connect those to the timing of the revolution process. The methodology behind this work is often called 'process tracing' (Bennet and Checkel 2015).

Not all pieces of empirical evidence would carry the same inferential weight. Some evidence would be expected to occur under both hypotheses (for example, increased political polarization), while other would be there only if one of these actually holds (for example, protesters citing economic rather than political grievances as a motivation for their discontent). This idea has been extended and formalized as a typology of the kinds of empirical evidence (in terms of their uniqueness and certitude) in light of competing theories (see Van Evera 1997, Colleir 2011, Mahoney 2012, Rohlfing 2014).

Things get more complicated when we do want to generalize based on a single-case study. For descriptive inference, we would normally want to select a case that is typical of a population. But under some circumstances, actually the description of a truly aberrant case might be more valuable. Descriptive case studies can also be used to check whether measures capture the empirical aspects they should (for example, whether people interpret survey questions in the way the researchers intended, or whether people discuss a political theme in the terms researchers assume). Process-tracing methodology can also help establish and illustrate the mechanisms behind known or suspected causal relationships.

For causal inference, the way single case-studies can contribute to general knowledge is via theory (Blatter and Haverland 2012). If there is very little prior theory on a research question, an in-depth single-case study can be deployed to generate initial theoretical insight, which can be later developed deductively and tested. When a novel theory is proposed, it might be worth testing the theory in a so-called plausibility probe (Eckstein 1992): a pilot case for which the theory is most likely to hold. Conversely, when theory is strong and well-established, we would be more interested in testing it on a least likely case in order to probe the limits of its applicability (see Gerring 2007). Sometime, even single case studies can deal fatal blows to existing theories (an often cited example is Lijphart 1968), but there are not many such examples, because political science theories are typically flexible enough to accommodate a small number of disagreeing observations or predictive failures. Altogether, for the design of single-case projects, the relationship with theory is even more important than for projects based on other types of research methodologies.

Interpretivist research

The discussion of case and variable selection above assumed a broadly positivist epistemological outlook and realist understanding of the political and social worlds. When one starts from an interpretivist point of view, it is less clear what design considerations need to be taken into account. It is not that anything goes in interpretivist research, but the methodological advice for this class of political science research is less structured.

Interpretivist research projects often focus on cases of substantive importance, and in that case their *methodological* concern with case selection is minimal. In fact, one of the criticisms of interpretivists to positivist social research is that the latter is led to study only questions for which systematic data, often collected by government agencies, exists, while these might not be the questions society should care about.

Interpretivists have ambiguous attitude towards the generalizability of research results. Some would say that generalizability is not a goal at all, and that it is not attainable. For others, generalizability is valuable, but is not achieved via random sampling or similar techniques. On the assumption that people are dominated by one or a small number of powerful discourses, conversing with *any* individual can reveal such a dominating discourse, so the exact choice of respondents does not matter too much (but the setting of the interaction does!). Furthermore, the analytical interest is rather in the deconstruction of the genesis and operating mechanisms of these discourses (for the study of which methods related to process tracing can be used, along with the design considerations about the selection of evidence that we discussed above).

But altogether, in the context of interpretivist projects, it is hard to specify in the design stage what kind of evidence should be sought for. In fact, interpretivists favor participant observation, ethnographic fieldwork, and unstructured interviews as methods for data collection *because* these methods are less constrained by prior theoretical considerations, imposed by the researcher, about what is important for a phenomenon of interest. Nevertheless, even the most open-ended data collection cannot capture *all* aspects of reality, so some reflection on what and how should be observed is indispensible during the design stage. That being said, interpretivists definitely cast a wider net when collecting data, and they are more flexible in adjusting their designs during data collection and analysis.

Conclusion: the power and promise of research design

This chapter presented an integrative perspective on research design in political science. Instead of arguing that some research design or methodology is inherently better than others, the chapter argued for a pluralistic and pragmatic approach. Political scientists can use a variety of designs and methodologies that are all legitimate, but some are more appropriate for particular research goals than others. Instead of exaggerating the differences between interpretivist and positivist, or quantitative and qualitative research (cf. Goertz and Mahoney 2012), we showed that they can all be used to answer different research questions and address different research goals.

For example, experiments are often considered the gold standard when it comes to estimating causal effects, testing theoretically-derived hypotheses, and prospectively evaluating the likely impact of social and political interventions. But their application is severely constrained by ethical and practical considerations, and often external validity is suspect. And experiments would only be appropriate when previous studies have generated the theoretical ideas being tested and shed light on the research problem in the first place. And, inevitably, observational research would be needed to probe the generalizability of the experimental results across different contexts.

Attention to research design can unleash the power of political science research by supporting the discovery and testing of valid inferences in a reliable and efficient way. But it also protects against the long list of biases of human cognition (Kahneman and Tverski 2000, Sloman 2005). Many of these biases are particularly troublesome for scientific reasoning in general. But some, in particular the ones related to the covert influence of political ideologies on information-processing and other aspects of cognition and decision making (e.g. Nyhan and Reifler 2010, Kahan et al. 2013), are especially pertinent to political science research. For all their special training, researchers remain human and could benefit from the disciplining role of research design to avoid cognitive and ideological biases.

The empirical social and behavioral sciences, political science very much included, face significant challenges in the form of increased scrutiny to research ethics and integrity.

Research design can help address these concerns as it enhances transparency, supports replicability, and encourages reproduction.

In a recent development, some academic journals in the social sciences, and political science in particular (see the Winter 2013 issue of the journal *Political Analysis*), encourage the public registrations and publication of research designs before data collection and analysis have been conducted. Such preregistration is supposed to ensure that the researchers do not misrepresent the goals of their research projects (for example, selling inductive theory-generating studies as theory testing ones) or 'game' their data analysis to produce 'significant' but unwarranted results. While preregistration is most directly relevant for theory and hypothesis testing designs, it holds a broader promise to improve the research process in the social sciences.

At the same time, research design should not be considered a straightjacket constraining creativity and the discovery spirit of scientific research. It should enhance rather than limit the potential of research projects and give wings rather than weigh down original research ideas.

Further reading

- The themes introduced in this chapter are discussed more comprehensively in the chapters of my book 'Research Design in Political Science' (2016).
- Other recommended general books on research design and methodology include Gerring (2012a) and Gschwend and Schimmelfennig (2007). King, Keohane, and Verba (1994) remains a standard reference, read preferably in conjunction with Brady and Collier (2004) and the 2010 special issue of *Political Analysis*.

- On theory in the social sciences, Elster (2007) is recommended, and on concepts and conceptualization, Goertz (2006).
- On causal inference in large-N research, see Morgan and Winship (2015) and Imbens and Rubin (2015).
- On case study designs, see the books by Rohlfing (2012) and Blatter and Haverland (2012), and on the related subject of process tracing methodology, Bennet and Checkel (2015).

References

Adcock, R. and D. Collier (2001) 'Measurement Validity: A Shared Standard for Qualitative and Quantitative Research', American Political Science Review, 95, 529-46.

Bakker, R., C. d. Vries, E. Edwards, L. Hooghe, S. Jolly, G. Marks, J. Polk, J. Rovny, M. Steenbergen and M. A. Vachudova (2012) 'Measuring Party Positions in Europe: The Chapel Hill Expert Survey Trend File, 1999–2010', Party Politics.

Baumgartner, F., B. Christian, G.-P. Christoffer, D. J. Bryan, B. M. Peter, N. Michiel and W. Stefaan (2009) 'Punctuated Equilibrium in Comparative Perspective', American Journal of Political Science, 53, 603-20.

Bennett, A. and J. Checkel (eds) (2015) Process Tracing. From Metaphor to Analytic Tool. (New York: Cambridge University Press).

Blatter, J. and M. Haverland (2012) Designing Case Studies: Explanatory Approaches in Small-N Research, (Basingstoke: Palgrave Macmillan).

Brady, H. E. and D. Collier (2004) Rethinking Social Inquiry. Diverse Tools, Shared Standards., (Oxford: Rowman and Littlefield).

Braumoeller, B. F. and G. Goertz (2000) 'The Methodology of Necessary Conditions', American Journal of Political Science, 44, 844-58.

Budge, I., H.-D. Klingermann, A. Volkens, J. Bara and E. Tanenbaum (2001) Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945-1998, (Oxford: Oxford University Press).

Collier, D. (2011) 'Understanding Process Tracing', PS: Political Science & Politics, 44, 823-30.

Cox, D. R. and N. Reid (2000) The Theory of the Design of Experiments, (New York: Chapman & Hall).

Druckman, J. N., D. Green, J. Kuklinski and A. Lupia (2011) Cambridge Handbook of Experimental Political Science, (Cambridge: Cambridge University Press).

Duflo, E., R. Glennerster and M. Kremer (2006) 'Using Randomization in Development Economics Research: A Toolkit', NBER Technical Working Paper Series.

Dunning, T. (2012) Natural Experiments in the Social Sciences. A Design-Based Approach, (New York: Cambridge University Press).

Eckstein, H. (1992) Regarding Politics: Essays on Political Theory, Stability, and Change, (Berkeley: University of California Press).

Elster, J. (2007) Explaining Social Behavior. More Nuts and Bolts for the Social Sciences, (Cambridge: Cambridge University Press).

Geertz, C. (1973) The Interpretation of Cultures, (New York: Basic Books).

Gelman, A. (2011) 'Causality and Statistical Learning', American Journal of Sociology, 117, 955-66.

Gelman, A. and D. Weakliem (2009) 'Of Beauty, Sex, and Power', The American Scientist, 97, 310-6.

Gerring, J. (2007) 'Is There a (Viable) Crucial-Case Method?', Comparative Political Studies, 40(3), 231-53.

Gerring, J. (2012) 'Mere Description', British Journal of Political Science, 42, 721-46.

Gerring, J. (2012) Social Science Methodology: A Unified Framework. 2nd Edition: Cambridge Universtiy Press).

Glymour, B. (1998) 'Contrastive, Non-Probabilistic Statistical Explanations', Philosophy of Science, 65, 448-71.

Glymour, M. M. (2006) 'Using Causal Diagrams to Understand Common Prolems in Social Epidemiology', in Oakes, J. M. and J. S. Kaufman (eds) Methods in Social Epidemiology, San Francisco: Jossey Bass),

Goertz, G. (2006) Social Science Concepts. A User's Guide, (Princeton: Princeton University Press).

Goertz, G. and J. Mahoney (2012) A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences, (Princeton: Princeton University Press).

Gray, V. and D. Lowery (2000) The Population Ecology of Interest Representation., (Michigan: Michigan University Press).

Grofman, B. (2001) 'Introduction: The Joy of Puzzle Solving', in Grofman, B. (ed.) Political Science as Puzzle Solving, Ann Arbor: University of Michigan Press), pp. 1-12.

Gschwend, T. and F. Schimmelfennig (eds) (2007) Research Design in Political Science: How to Practice What They Preach. (Basingstoke and New York: Palgrave Macmillan).

Holland, P. (1986) 'Statistics and Causal Inference', Journal of the American Statistical Association, 81, 945-60.

Imbens, G. (2011) 'Experimental Design for Unit and Cluster Randomid Trials', International Initiative for Impact Evaluation Paper.

Imbens, G. and D. Rubin (2015) Causal Inference for Statistics, Social, and Biomedical Sciences. An Introduction., (New York: Cambridge University Press).

Kahan, D. M., E. Peters, E. Dawson and P. Slovic (2013) 'Motivated Numeracy and Enlightened Self-Government', Yale Law School, Public Law Working Paper, 1-35.

Kahneman, D. and A. Tverski (2000) Choices, Values and Frames, (New York: Cambridge University Press).

King, G., R. Keohane and S. Verba (1994) Designing Social Inquiry. Scientific Inference in Qualitative Research, (Princeton: Princeton University Press).

Kittel, B., W. Luhan and R. B. Morton (2012) Experimental Political Science. Principles and Practices, (Basingstoke: Palgrave Macmillan).

Lehnert, M., B. Miller and A. Wonka (2007) 'Increasing the Relevance of Research Questions. Considerations on Theoretical and Social Relevance', in Gschwend, T. and F. Schimmelfennig (eds) Research Design in Political Science. How to Practice What They Preach, Basingstoke: Palgrave Macmillan), pp. 21-40.

Lewis, D. (1973) 'Causation', Journal of Philosophy, 70, 556-67.

Lewis, D. (1986) Causal Explanation, (Oxford: Oxford University Press).

Lieberman, E. S. (2005) 'Nested Analysis as a Mixed-Method Strategy for Comparative Research', American Political Science Review, 99, 435-52.

Lijphart, A. (1968) The Politics of Accommodation. Pluralism and Democracy in the Netherlands, (Berkeley: University of California Press).

Little, D. (1991) Varieties of Social Explanation: An Introduction to the Philosophy of Social Science, (Boulder: Westview Press).

Lohr, S. (2009) Sampling: Design and Analysis. Second Edition: Brooks-Cole).

Mahoney, J. (2012) 'The Logic of Process Tracing Tests in the Social Sciences', Sociological Methods & Research, 41, 570-97.

Morgan, S. and C. Winship (2015) Counterfactuals and Causal Inference. Methods and Principles for Social Research. Second Edition, (Cambridge: Cambridge University Press).

Morton, R. B. and K. Williams (2010) Experimental Political Science and the Study of Causality. From Nature to the Lab, (Cambridge: Cambridge University Press).

North, D., J. Wallis and B. Weingast (2009) Violence and Social Orders. A Conceptual Framework for Interpreting Recorded Human History, (New York: Cambridge University Press).

Nyhan, B. and J. Reifler (2010) 'When Corrections Fail: The Persistence of Political Misperceptions', Political Behavior, 32, 303-30.

Pearl, J. (2009) Causality. Second Edition., (New York: Cambridge University Press).

Peirce, C. S. (1955) Philosophical Writings of Peirce. Ed. By Justus Buchler, (New York: Dover).

Przeworski, A. and H. Teune (1970) The Logic of Comparative Social Inquiry, (New York: Wiley-Interscience).

Rohlfing, I. (2008) 'What You See and What You Get Pitfalls and Principles of Nested Analysis in Comparative Research', Comparative Political Studies, 41, 1492-514.

Rohlfing, I. (2012) Case Studies and Causal Inference an Integrative Framework, (Basingstoke: Palgrave Macmillan).

Rohlfing, I. (2014) 'Comparative Hypothesis Testing Via Process Tracing', Sociological Methods & Research, 43, 606-42.

Rosenbaum, P. R. (2005) 'Heterogeneity and Causality', The American Statistician, 59, 147-52.

Salmon, W. (1998) Causality and Explanation, (New York: Oxford University Press).

Sartori, G., F. Riggs and H. Teune (1975) The Tower of Babel: International Studies Association).

Schwartz-Shea, P. and D. Yanow (2012) Interpretive Research Design. Concepts and Processes, (New York: Routledge).

Sloman, S. (2005) Causal Models: How People Think About the World and Its Alternatives, (New York: Oxford University Press).

Small, M. L. (2011) 'How to Conduct a Mixed Methods Study: Recent Trends in a Rapidly Growing Literature', Sociology, 37, 57.

Toshkov, D. (2016) Research Design in Political Science, (Basingstoke: Palgrave Macmillan).

Valliant, R., J. Dever and F. Kreuter (2013) Practical Tools for Designing and Weighting Survey Samples, (New York: Springer-Verlag).

Van Evera, S. (1997) Guide to Methods for Students of Political Science, (Ithaca: Cornell University Press).

Weller, N. and J. Barnes (2014) Finding Pathways. Mixed-Method Research for Studying Causal Mechanisms, (Cambridge: Cambridge University Press).

Woodward, J. (2003) Making Things Happen: A Theory of Causal Explanation, (New York: Oxford University Press).