CHAPTER 1

Introduction

> Singular causal claims are primary. This is true in two senses. First, they are a necessary ingredient in the methods we use to establish generic causal claims. Even the methods that test causal laws by looking for regularities will not work unless some singular causal information is filled in first. Second, the regularities themselves play a secondary role in establishing a causal law. They are just evidence—and only one kind of evidence at that— that certain kinds of singular causal fact have happened.
>
> —*Nancy Cartwright*

THIS book addresses some of the core issues in doing (1) multimethod research, (2) causal mechanism analysis, and (3) case studies. Multimethod research has become very popular and almost a requirement for book-length studies. Multimethod can mean many things, but here it means combining case studies with statistics, qualitative comparative analysis (QCA), experiments, or game theory models. The purpose of case studies is to explore causal mechanisms at the heart of theories. One does case studies because cross-case methods give little purchase on the causal mechanisms (M_i) by which X produces Y. For multimethod researchers, showing a significant causal effect in a cross-case analysis is not sufficient; one needs to provide a causal mechanism and evidence for it. Demonstrating a causal effect is only half the job; the second half involves specifying the causal mechanism and empirically examining it, usually through case studies.

Causal mechanisms, cross-case analyses, and case studies form the *research triad*; see figure 1.1. This volume rests on the proposition that commitment to multimethod research is commitment to the research triad. Multimethod research

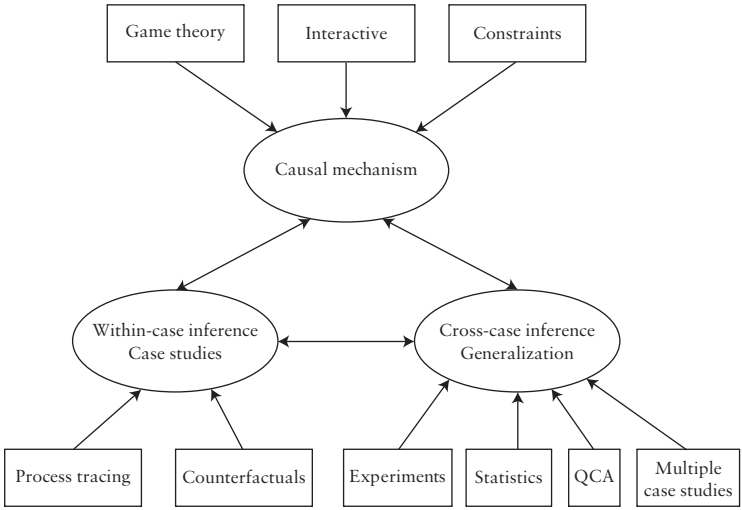

Figure 1.1: The research triad: causal mechanism, cross-case inference, and within-case causal inference.

typically is conceptualized as qualitative—within-case inference—along with quantitative cross-case inference. The research triad adds a third dimension to that, i.e., causal mechanisms. The research triad is an integrated approach because a commitment to multimethod research is also a commitment to a causal mechanism approach to explanation and social science research.

One way to see the integration is by looking at research that focuses, e.g., 90 percent of the research effort, on just one corner of the triad. For many experimentalists most of the effort is devoted to determining *the treatment effect*. They usually do not talk about causal mechanisms per se. While they usually have theories and hypotheses, these all boil down to the one treatment effect. Game theorists provide an nice example of those who focus a lot of attention on the model, i.e., causal mechanism. It is quite possible to publish articles where essentially the whole article is the model (this is very true in economics). Often historians focus on single events and the whole focus of the article is explaining some individual historical event. They are

not interested in generalization and maybe only very implicitly interested in causal mechanisms.

The integrated approach rests on a central claim:

> As you move away from one-corner-only research you embrace the research triad.

For example, as a game theorist moves away from exclusive interest in the model, she begins to be involved in multimethod research, cross-case analyses, and within-case analyses. As soon as a case study researcher moves from one case to more cases, he is asking about generalization. Finally, as the Cartwright epigraph forcefully states, all statistical, experimental analyses and generalization imply individual case causal inference.

Good multimethod and causal mechanism research means a relative balance between the three corners of the research triad. If 90 percent of the effort is in demonstrating significant causal effects via cross-case analysis then the research is not serious multimethod or serious about the investigation of the causal mechanism. If the case studies are only "illustrations," then there is little commitment to multimethod research. Conversely, research using case studies—e.g., the popular paired comparison—is weak on generalization. Doing five or six case studies—as is common in security studies books—deals poorly with cross-case analysis and generalization. The ideal is a fairly balanced effort on all three points of the triangle.

Case studies are often considered of questionable inferential value. Clarke and Primo illustrate this view of case studies. When they refer to case studies they almost always describe them as "exploratory": "one can also design exploratory models with an eye toward explaining the events surrounding a specific case" (2012, 92). By "exploratory" they strongly imply the endeavor is not about causal inference. For example, their discussion of the analytic narratives project (Bates et al. 1998) falls into this category. When they talk about "empirical models" they mean statistical models. So a model of the US Senate—statistical or formal—is not exploratory, but a case study of the American Civil War in analytic narratives is exploratory. In contrast, the research triad emphasizes that case studies are about causal inference.

In everyday life—along with virtually all natural sciences—people successfully make individual-case causal inferences, e.g., origins of the universe, origins of the human species, why a given person died, why the Challenger shuttle exploded. None of these inferential successes relies on randomized treatments assigned to subjects, nor do they depend on conditional probabilities.

A core philosophy motivating this study is that we want to explain individual outcomes. Statistical analyses do not provide explanations: "There is little argument in political science that statistical models cannot serve as explanations in and of themselves. This belief manifests itself in the relegation of statistical models to devices for *testing* explanations" (Clarke and Primo 2012, 154).

The research triad assumes that one accepts the importance and value of causal mechanism analysis. One can find statistical methodologists who do not believe this is possible or important: "The importance of searching for causal mechanisms is often overestimated by political scientists, and this sometimes leads to an underestimate of the importance of comparing conditional probabilities. We do not need to have much or any knowledge about mechanisms in order to know that a causal relationship exists. . . . In general, as our understanding of an issue improves, studying individual cases becomes less important" (Sekhon 2004, 288–89; last sentence of the article). Gerring illustrates the skepticism about whether causal mechanism analysis is essential: "To clarify, this is not a polemic against mechanisms. It is a polemic against a dogmatic interpretation of the mechanismic mission. I argue that the analysis of causal mechanisms is best regarded as an important, but secondary, element of causal assessment—not a necessary condition" (Gerring 2010, 1500).

I do *not* survey or discuss how one goes about doing within-case causal inference in this book. There is a booming literature on the topic, e.g., Bennett and Checkel (2014), and Beach and Pedersen (2012; 2016) on process tracing, Mahoney (2012) on hoop and smoking gun tests, Goertz and Levy (2007), Levy (2008), and Harvey (2011) on counterfactuals. One can do statistical within-case causal inference. I take no position on how one does within-case inference. Similarly I do not cover how to do observational statistics, experiments, or QCA. As illustrated

in figure 1.1 these methodologies provide input for the research triad but are not covered here.

"Qualitative" and "quantitative" are not very useful in describing or analyzing multimethod research. Instead of multimethod research as qualitative and quantitative, the research triad contrasts *within-case* causal inference (case studies) with *cross-case* causal inference (comparative case studies, statistical models, experiments, or QCA). This produces some surprising methodological bedfellows. Standard usage puts statistics in the quantitative category and set-theoretic approaches (e.g., QCA) in the qualitative. Experimenters spend a lot of time stressing the differences between experiments and observational research: here they are both cross-case methodologies. Similarly *comparative* case studies are cross-case analyses. I consider these all as versions of cross-case causal inference. In contrast, single case studies by their very nature are about what happens in individual cases. Case studies are fundamentally about within-case causal inference. The research triad means that multimethod research is multicausal inference analysis.[1] The causal inference techniques, procedures, and methodology of each type, cross-case and within-case, serve different but complementary goals.

The research triad works from a basic principle:

> Multimethod work involves cross-case causal inference AND within-case causal inference.

Multimethod in this book means complementary causal inference methodologies. How one does cross-case inference or within-case inference is less important than the causal inference goals.[2]

[1] Thanks to Hillel Soifer for this point.

[2] Multimethod research can have other roles. Within the potential outcomes framework many of the core assumptions about treatments, randomization, selection, effect heterogeneity, noncontagion, etc. could be examined by looking closely at some individual cases. For example, Harding and Seefeldt (2013) stress that case studies can be extremely useful in understanding selection processes. Particularly, in the context of field experiments where "nature" does randomization, one needs to closely investigate via case studies the extent to which this assumption holds. For example, core to John Snow's famous natural experiment about cholera was an extensive analysis of who got water from the

To connect cross-case and within-case analysis means having a methodology for choosing cases for causal mechanism analysis. The practical problem of choosing cases runs as a bright red thread throughout the book. These decisions face all who connect case studies to other methodologies. In particular, I offer much specific guidance about case selection. This means a systematic set of guidelines for case selection including a list of criteria for getting to a final decision about which cases to choose.

McGuire (2010) illustrates typical multimethod research involving a statistical analysis and case studies. His dependent variable is health outcomes. The second chapter of his book is a large-N cross-national statistical analysis, which is followed by eight country case studies. These case studies focus on explaining outcomes in those countries, for example, the following:

> What needs to be explained, then, is why Costa Rica's infant mortality rate was so low in 2005 (why it attained a certain level); why it fell so fast from 1960 to 2005 (why it achieved a certain amount of progress), and why it fell faster during the 1970s than at other times within this period (why it evolved at a certain tempo). The sustained and effective public provision of basic health services to the poor, this chapter finds, goes a long way toward explaining why Costa Rica from 1960 to 2005 achieved a rapid decline (and, eventually, a low level) of infant mortality. (McGuire 2010, 66)

McGuire is making claims about what happened in Costa Rica and why it happened, i.e., within-case causal inference.

The standard rationale for multimethod work involves looking at causal mechanisms via case studies: "Despite some claims to the contrary in the qualitative methods literature, case studies are not designed to discover or confirm empirical regularities.

---

two different water companies. He showed that there was no obvious bias since both companies delivered to the same neighborhoods and there seemed to be no particular bias (e.g., wealth) among their clients. This volume does not deny the usefulness of case studies for these purposes. I am not trying to survey all possible uses of case studies or qualitative information in multimethod research. Rather I focus on using case studies to explore causal mechanisms. However, in my extensive survey of existing practice—articles and books—I have almost never seen case studies or qualitative methods used to check core assumptions of statistical models (this might be more common in sociology, which has a longer multimethod tradition).

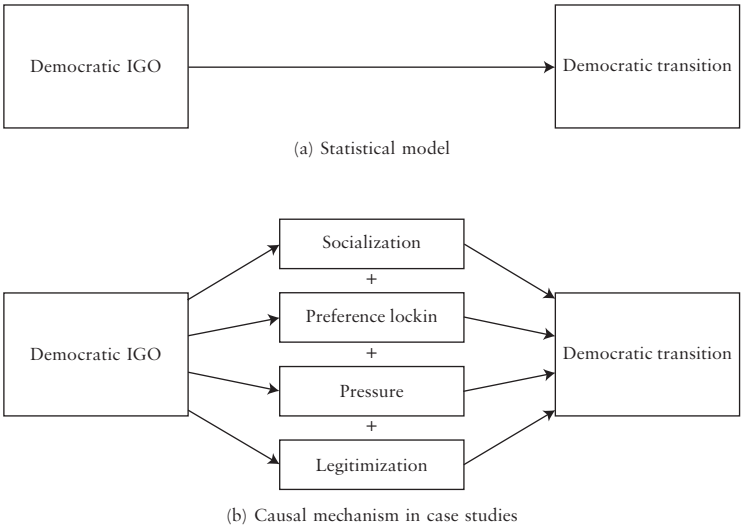

Figure 1.2: Causal mechanisms and statistical multimethod research: democratic IGOs and democratic stability. Source: based on Pevehouse (2005), table 5.1.

However they can be quite useful—indeed, essential—for ascertaining and assessing the causal mechanisms that give rise to empirical regularities in politics" (Fearon and Laitin 2008, 773). This is exactly what the research triad proposes: cross-case analyses for "empirical regularities" and case studies for causal mechanisms.

McAdam and Boudet give a similar rationale in their study of environmental social movements: "We conceived of the project as an attempt to develop an alternative to the methodological conventions of social movement research. Equally dissatisfied with 'thin,' large-N studies of protest events and rich but nongeneralizable case studies of this or that movement, we sought a middle ground between these two modal 'poles' of social movement scholarship" (2012, 52).

Pevehouse (2005) provides a nice example of why people want to do multimethod research. He argues that democratic IGOs (intergovermental organizations) can help establish democracy, make it more robust, and encourage transitions to democracy. As illustrated in figure 1.2a, there is a causal connection proposed between democratic IGOs and democracy in states.

He shows that there is a significant correlation between democraticness of the IGO and democracy in its member states. Multimethod research comes into play because he thinks there are multiple—and not mutually exclusive—causal mechanisms that explain this significant correlation. Figure 1.2b adds the causal mechanisms that produce this statistical effect: (1) acquiescence effect, (2) legitimization, (3) pressure, and (4) financial assistance (Pevehouse 2005, table 5.2, 153). Here one sees Pevehouse going around the research triad, from the statistical analyses to case studies to causal mechanism analysis.

Many hypotheses, experiments, and the like propose multiple causal mechanisms connecting the treatment to the outcome, as illustrated in figure 1.2. For example, Helfer and Voeten list three causal mechanisms whereby the European Court of Human Rights influences state policy: (1) preempting future international court litigation, (2) persuasive authority, and (3) agenda-setting at the national level. They argue—like Pevehouse—that "these three mechanisms may work separately or in tandem" (Helfer and Voeten 2014, 82). Hence, one role of case studies is to explore which of these mechanisms is actually at work.

A central role of case studies is combining within-case causal inference with analyses of causal mechanisms. These two in fact cannot be separated. To do process tracing, for example, means to make causal claims about the case. Usually it means that the causal mechanism in question can explain or contributes to the outcome in the individual case. Some researchers offer case studies as "illustrations"; this is disingenuous. Good multimethod research means that one must be serious about the within-case causal claims and make sure that they stand up to scrutiny.

Why after all is one doing a case study in multimethod research? The answer is almost always "to explore causal mechanisms." This is why the research triad is integrated and tightly linked. Doing multimethod research means doing case studies, which means exploring causal mechanisms. Once you have started down the multimethod path you have to pass by causal mechanisms.

Causal mechanism by definition means causal complexity (see chapter 2). A commitment to multimethod research and the research triad is a commitment to causal complexity. Causal

mechanisms involve causal complexity. The contrast is with those who are almost exclusively interested in the effect of X_1 on Y. As the various figures in chapter 2 illustrate, causal mechanisms are complex entities. Most often the easiest—and maybe the only practical—way to empirically explore mechanisms is process tracing in individual cases.

One variable, single treatment analyses are by contrast causally simple. The classic examples are single-variable hypotheses empirically investigated via statistical analyses or experiments. At the other extreme are game-theoretic models, which often have numerous assumptions and complex features. Even simple game-theoretic models involve multiple assumptions about individuals, their beliefs, rules of the game, etc.

A quite popular move for game theorists is to move to the cross-case corner of the research triad. There have been extensive efforts—with substantial NSF (National Science Foundation) funding—to connect game theory and statistics, hence game theory and statistical cross-case inference. Notably the EITM program (Granato and Scioli 2004) has held summer workshops for a number of years where the central component has been connecting statistical methods with game-theoretic models.[3] However, there has been very little done to connect game-theoretic models with qualitative methods and case studies (though see Lorentzen, Fravel, and Paine 2016), hence game theory and within-case causal inference.

Chapter 6 argues that there is a very natural connection between game theory and within-case causal inference. In addition to the cross-case analyses, one moves in the within-case inference direction. Many—if not almost all—of the crucial theoretical entities in game-theoretic models are hard to observe and measure in large-N settings. Factors like beliefs, information, uncertainty, and preferences are hard to determine even in one case, not to mention dozens or hundreds. As a result, statistical tests are virtually always indirect. For example, in the audience costs literature and debate (discussed in some detail in chapter 7) the usual proxy of audience costs is democracy. "Democracy" is

[3] Other modeling techniques are presented, such as agent-based computer models, but the core is game theory and statistics.

some distance away from the theoretical mechanism developed by Fearon (1994). In contrast, there is some hope of assessing core model features in one or a few individual cases.

Case studies naturally link up with game theory because the game-theoretic model itself is a causal mechanism. A game-theoretic model thus calls out for an empirical analysis in individual cases to see if the causal mechanism works in real life as advertised in the model. In practice many formal game theory articles—where the focus is on the model—include historical examples. Fortunately for this volume, Peter Lorentzen and his colleagues have surveyed the use and nonuse of case studies in formal work in international relations and comparative politics.

While formal modelers may be less interested in testing their models against individual cases, their critics have not been hesitant in recent years to do so. A number of the prominent critiques of game-theoretic work are discussed in chapter 7 including critiques of audience costs theories and Acemoglu and Robinson's work (2006). These examples involve using case studies to evaluate the empirical usefulness of formal models.

Causal complexity can take other forms. What one might call "simple complexity" involves causal heterogeneity and interaction theories and hypotheses. In figure 1.1, above the causal mechanism ellipse, I have "game theory" as well as "interactive" causal mechanisms. This volume explores some general kinds of causal mechanisms. Particularly of interest are those that can also appear in cross-case analyses. Interaction terms are not uncommon in statistical analyses. QCA is built from complex interaction terms. To develop hypotheses about interactions of various sorts is to start down the road from the cross-case analysis toward the causal-mechanism corner. Much of this volume then is how to connect these two corners to the case studies in the third corner.

Within the potential outcomes framework there is tension between the individual-level counterfactual, which allows for a great deal of causal heterogeneity, and the overall goal, which is the estimation of average treatment effects (ATE). The ATE can contain a great deal of heterogeneity. Much of current methodological research involves looking at this causal heterogeneity.

For example, a famous experiment—a HUD-funded study (Department of Housing and Urban Development, USA)—looked at the hypothesis that the neighborhood in which an individual was raised (poor versus wealthy) has important impacts on social mobility, educational achievement, etc. Households were assigned randomly to poor versus wealthy neighborhoods (obviously the design was more complicated than this). These poor households contain heterogeneity of subjects: one of the findings was that girls benefited from better neighborhoods, but boys were perhaps worse off. In this case the average causal effect is meaningless, since it depended on the gender of the person in interaction with neighborhood.

By interviewing boys, girls, and parents in the housing study—i.e., doing multimethod research—the reasons for the differences between boys and girls became clearer:

> Girls in more advantaged neighborhoods also made friends at school and work rather than in the community, exposing them to a different set of peers. Among boys, a different set of social processes occurred. Boys who moved to more advantaged neighborhoods were separated from male role models left behind in their former neighborhoods, engaged in public leisure activities like hanging out and playing sports that put them at greater risk of contact with police, and lost the opportunity to develop "street smarts" that they would be more likely to need to navigate dangerous streets when they returned to poor neighborhoods. (Harding and Seefeldt 2013, 98)

This means that there is an interaction between the treatment and gender.

Interaction terms or mediating variables are a simple kind of complexity. QCA paths typically involve two to four factors and thus embody, as a matter of course, causal complexity. Cross-case methods can incorporate some modest complexity, but then one needs to think about how that links up with case studies and causal mechanisms. For example, how does one do case selection when there is an interaction term or a path with INUS (Insufficient but Nonredundant parts of a condition which is itself Unnecessary but Sufficient for the occurrence of the effect) variables?

Chapters 2 and 4 explore constraint causal mechanisms. These are inherently complex because constraints have causal effects only when there is motivation to violate the constraint (this is the core idea of Starr's 1978 opportunity and willingness framework). Nuclear deterrence does not explain the peace between the USA and Canada. For example, the impact of veto players (Tsebelis 2002) depends on the degree of ideological divergence among the veto players. Case studies must be chosen so that the causal mechanisms embodied in these constraint theories can be seen, i.e., not USA–Canada for the mechanism of nuclear deterrence.

Chapter 5 explores multimethod in the context of statistical interaction hypotheses or set-theoretic models. The X_1 and X_2 interaction raises the question about how they work together in a causal mechanism. In the QCA context there is the additional claim that within the path, X_1 and X_2 are necessary conditions. This means that the material in chapter 4 is critical. The causal mechanism analysis must focus on how X_1 and X_2 together produce Y. At the same time one needs to look at how the absence of X_1 or X_2 prevents Y from occurring.

One can start at the within-case case study corner of the research triad. This links easily to the causal mechanism corner because that is typically why one is doing the case study to begin with. Much more problematic is the link to the generalization ellipse. From the causal mechanism corner one asks about the scope and generalizability of the causal mechanisms.

Salmon (1998) refers to two grand traditions within philosophy of science, the first related to the hypothetico-deductive approach. The second tradition is that of causal mechanisms. The hypothetico-deductive approach is based in philosophy of science where the science in question is physics. The issue of scope and degree of generalization does not really arise. In contrast, in the philosophical literature on causal mechanisms, scope is a core consideration. Biological sciences—particularly biochemistry—become the sciences of analysis. In these sciences the scope of the causal mechanisms and theories is typically open and up for discussion. Causal mechanism ideas describe much better what biologists do than the covering laws of physics.

So instead of talk of testing causal mechanisms, one can reformulate the question in terms of the scope of the causal mechanism. Causal mechanisms with wide scope pass tests of importance and significance, while causal mechanisms of limited scope are empirically less important.[4]

In principle, theoretical and empirical scope are important research questions. However, they are rarely discussed in applied research and rarely appear in methods and research design textbooks. For statistical work, empirical scope is implicitly defined by the limits of the data set. The scope of game theory models is typically unclear since it is rarely made explicit.

Usually, case studies of the classic qualitative sort are not seen as good tests of a theory, hypothesis, or causal mechanism.[5] Chapter 7 discusses how a variety of scholars are beginning to do what I call "large-N qualitative testing." "Large-N" here is somewhat ironic because the testing uses many instances of within-case causal analyses to explore the scope and validity of some prominent game theory models or statistical analyses.

In large-N qualitative testing one starts with, say, a game-theoretic model. Then one moves within the research triad to individual cases and within-case causal inference and then to the generalization part of the triad. Finally, one draws some general conclusions about the empirical validity of the game-theoretic model.

Particularly in the case of game theory models, large-N qualitative testing begins by determining the scope of the case study population. For game theory models, constructing the case study population is a challenge (see the discussion of the audience costs debate). As discussed in chapter 7, a key part of the methodology is constituting the population.

The large-N qualitative critics of statistical studies almost never use *all* the cases in a statistical analysis for their within-case analyses, in fact the cases chosen form a relatively small

[4] As we shall see, however, they might apply to historically important cases, so in that sense they are substantively important.

[5] Sometimes an author will claim that a case study can disconfirm a deterministic hypothesis (e.g., Gerring 2012). Such strict falsificationism has long been discredited in the philosophy of science literature, and one can find little evidence for it in scientific practice.

subset of the statistical population of observations. At the same time, researchers claim that they have looked at all the relevant cases. Chapter 7 analyses this tension in some detail. All this is part of determining how empirically generalizable the model is.

Critics sometimes find that the game-theoretic analysis has narrow empirical scope. In discussions of statistical analyses, they find that while there might be a statistically significant correlation in the large-N analyses, there is little evidence for the causal mechanism when doing within-case causal inference.

Chapter 8 proposes rethinking how scholars do *multiple* case studies (i.e., not comparative case studies) and the connection between statistical analyses and within-case causal inference. It attacks directly the standard question:

> How generalizable is a successful within-case causal inference of a causal mechanism?

This means thinking about moving from the case study ellipse in figure 1.1 to the generalization ellipse.

In the "medium-N paradigm" outlined in chapter 8, additional case studies are designed and implemented in order to evaluate how generalizable the causal mechanism is. To investigate the empirical scope of a causal mechanism means deciding how many cases to explore and which specific cases to include.

The medium-N paradigm has important implications for case study research generally. For example, it implies that popular designs like paired comparisons have little to say for them. Similarly, the popular book format of one theory chapter and five or six equal case studies can be improved upon significantly.

Hence, chapters 7 and 8 form a pair. Consistent with the importance of exploring research practice in this book, chapter 7 looks at influential and controversial tests of prominent theories, such as Cusack, Iversen and Soskice (2007), Acemoglu and Robinson (2006), and Fearon (1994), by qualitative scholars, using the cumulation of within-case causal inference. Given that these large-N qualitative tests were published in major journals, e.g., *American Political Science Review*, the implicit methodology of these articles has convinced editors and reviewers.

Chapter 8 then systematizes and presents the logic of the medium-N paradigm. For example, most of the studies surveyed in chapter 7 attempt to look at *all* relevant cases. This might not be practical and might not be an efficient use of resources.

The relevance of the medium-N paradigm extends to experimental research as well. In both experiments and the medium-N paradigm, the emphasis is on high-quality within-case causal inference. For much of the discussion in chapter 8 one can substitute "experiment" for "within-case causal inference." For example, if you can do, say, four case studies of the same causal mechanism, which ones would you choose? Replace "case study" with "experiment" and you have the same basic methodological issue.

Jim Mahoney and I (Goertz and Mahoney 2012) argued that there were two cultures of methods research, one based on set theory and mathematical logic and the other based on statistical methods. Comparative case study and statistical multimethod research then lies at the intersection of the two cultures. Comparative-case-study methodology was, and is, strongly influenced by cross-case comparison ideas from statistics (see appendix A). Much case study methodology—notably Gerring (2006; 2017)—is explicitly based on statistical models.

One motivation for this volume is *integrating* and bridging these cultural divides. Often there is relatively easy and common linkage between two corners of the research triad, but a large gap to the third. This is because the two cultures cut through the research triad. For example, EITM builds strong linkages between game-theoretic models and cross-case analysis, but completely ignores the empirical exploration of game-theoretic causal mechanisms, which can only really be done in case studies. Classic qualitative research is the common link between case studies, within-case causal inference, and the causal complexity of causal mechanisms, but it has difficulties with cross-case analysis and generalization.

This tension appears in the statistical multimethod literature, which is strongly located in the cross-case statistical corner. Statistical concerns drive case study selection and at the same time downplay the importance of within-case causal inference and causal mechanism analysis.

One can see this cultural tension in Gerring's (2006) list of nine different case selection criteria. The causal mechanism type is number (7), the pathway case. This causal mechanism type is buried in the middle of Gerring's list. While he does not explicitly consider this list as ordinal, there is no doubt that the dominant and privileged types are those at the top: (1) typical and (2) diverse. The typical case is the average, representative case in the context of some population of cases.

Here we see the tension with practice. Since most researchers are interested in multimethod research for causal mechanism reasons, they rarely use Gerring's preferred case study designs.

Similarly, in recent years there has been a surge of articles, papers, and books looking at case selection for case studies from a potential outcomes perspective (Glynn and Ichino 2015; Herron and Quinn 2016; Seawright 2016; Weller and Barnes 2014). Since matching (e.g., Nielsen 2016) is almost always used for confounders, this leads naturally to pairs of comparative cases. This potential outcomes literature is an updating of the classic paired comparison–similar system design of qualitative methods.

In this book I think it is quite important to have a clear unified account of what case study research looks like when approached via statistical models and ways of thinking. In other words, if the cross-case statistical analysis were driving everything what would the methodology look like? In contrast to the integrated approach where each corner has its role to play, the statistical approach runs everything from the cross-case, statistical corner. Appendix A provides an account of a statistics-based methodology of case- and comparative-case-study research. For example, I discuss matching methods as the standard approach to confounders.

Appendix A is not a summary or a recap of this recent work, but offers its own account. It differs in important respects from the current literature. For example, it provides a pathway case selection procedure that is closer to the potential outcomes philosophy than Gerring (2017). I rely on the basic counterfactual of the potential outcomes approach while Gerring uses comparative statistical models to choose cases.

QCA provides a good example of an approach that stresses the corners of the research triangle. Ragin's groundbreaking

book, *The comparative method*, proposed Boolean algebra as a means of doing cross-case comparisons. At the same time, he has continually emphasized that QCA is a *case-based* methodology. QCA results are closely connected to cases (unlike statistical multimethod research in general; see appendix A). Good QCA research always connects the paths of the cross-case analysis to the cases on those paths. The interactive nature of the paths in QCA points the researcher to a potential causal mechanism.

Individual scholars can maintain a balance between the corners of the triad. Bueno de Mesquita's research (see chapter 6) provides a good example. He develops game-theoretic models, does large-N statistical tests, applies the model to cases, and does predictions. Another example comes from the Robbers Cave experiment of Sherif et al. (1961), a classic in social psychology. Sherif and his colleagues explicitly discussed their combining of methods as a strength of the study; Donald Campbell in his 1988 introduction stressed its multimethod nature:

> One of the valuable slogans of the new emphasis on qualitative, contextual methodology is "thick description" (Geertz 1973). The Robbers Cave study provides such thick description. Moreover, the many ingenious subexperiments that are introduced, with their "natural" opportunities for quantitative measurement, add greatly to the "thickness," creating opportunities for participant action and qualitative observation that would not otherwise have existed, as well as providing quantitative measures. In this study, better than anywhere else I can think of, the proper synthesis of the qualitative-versus-quantitative dialectic is achieved. (Campbell in his introduction to Sherif et al. 1988, xxi)

It is common in sociology and psychology to do large-N analyses combined with interviews of individuals. For example, Louise Roth's book (2006) on the gender bias of Wall Street illustrates balanced qualitative and quantitative research. She had a clear large-N sampling strategy, did questionnaires and regression analyses. Yet much of the book comes from her interviews, which helped her understand how the structures and practices of Wall Street produced significant gender bias.

John Snow, Cholera, and the Research Triad

In his methods debate with Brady and Collier, Neal Beck asked the question of who owned the Snow–cholera example: "Who gets to claim John Snow?" (Beck 2010, 500).

In many respects, John Snow is a model of the research triad. He conducted a famous natural experiment, hence he is firmly located in the cross-case, empirical corner. At the same time, a lot of his research fits with the within-case and causal mechanism emphasis of the other two corners, as illustrated by his analysis of the Broad Street pump.

Snow is also a good choice because medical and drug examples are common in the social science methodology literature, and in particular, the multimethod literature. The well-known statistician David Freedman introduced the Snow–cholera example to social scientists in an often-cited chapter with the nice metaphor of "shoe leather" in the title. Freedman's Berkeley colleagues in political science picked up this example and it appears regularly in Brady–Collier–Dunning–Seawright publications, hence it is well known in the qualitative as well as statistical methods literature. Snow is a major figure in the history of epidemiology, sometimes considered the founding father for his various innovative methods for studying the causal mechanism whereby cholera is transmitted (Hempel 2007; Vinten-Johansen et al. 2003).[6]

It is useful to think of medical research in the context of the research triad. Modern drug developments start in the causal mechanism corner. For example, depression drugs are based on biochemical analyses of how the brain works. The cross-case analysis comes much further downstream in terms of testing whether a drug actually works. Medical doctors are located at the end; they hope that the drug works on individual patients and they have to decide whether generalizations about the effect of the drug apply to patient X.

Traditional medicines start from the individual case corner. Over time, societies discover that some plant works for some

[6] When considering this cholera example, it is critical to understand that Snow could not really get at the causal mechanism of cholera. This would require virus theory, Pasteur, and developments in biochemistry.

disease: this is generalization from individual cases. Modern medicine kicks in when it tries to find the "active ingredient" and mechanism. The history of medicine is full of examples of things that worked (e.g., penicillin). Often the elaboration of the causal mechanism came much later. Here the route is thus from individual case to generalization to causal mechanism.

As is often the case, these examples, such as Snow–cholera, get a life of their own. It is useful to go back to the sources and look at what kinds of analyses Snow conducted to convince himself—at the time he was not too successful at convincing others—that cholera was transmitted via drinking water.

Snow is justly famous for his natural experiment involving two water companies that competed in the same neighborhoods to deliver water to residents. The shoe leather Freedman referred to was Snow's extensive work going through the neighborhoods to show that the treatment—which water company used—was an "as if" randomized treatment. This involved showing that there was no bias by income, profession, education, etc. in the choice of water source. Hence, it is not surprising that in Dunning's (2012) book on natural experiments this example receives an extended discussion.

However, Snow conducted two other extensive shoe leather analyses. The one most interesting for this volume is "the case of the Broad Street pump," which illustrates many features of the arguments made in the chapters to come.[7] Most discussions of process tracing methodology employ the detective metaphor. Finding the cause is like Sherlock Holmes discovering the murderer. Collier (2011) has pushed this to its natural conclusion by publishing a methods article using an extensive analysis of a Holmes story. The Broad Street pump case has many features that illustrate Snow's detective work at its best.

In 1854 there was a very severe, but localized, outbreak of cholera in London. Within the area of a few square kilometers there were numerous cases of cholera. Snow as detective set out to discover the cause of this outbreak. It occurred at a time when he had convinced himself that infected drinking water was

[7] The third involves an important rule for dealing with confounders or alternative explanations and will be discussed in chapter 3.

the source. With his causal mechanism in mind he looked to link the cases of $Y = 1$ (cholera) with an $X = 1$—closeness to the source of bad water. Eventually his attention focused on the Broad Street pump, which seemed to be at the geographic epicenter of the outbreak; almost all victims were closer to the suspect pump than any other public water source.[8] He paid particular attention to the $X = 1$ cases of people who lived close to the pump but who did not get cholera ($Y = 0$). These cases could seriously challenge or disconfirm his causal mechanism. He tried to verify as much as possible his hunch that they got their water elsewhere, hence in fact they were $X = 0$ cases. Snow was eventually able to convince local authorities to shut down the pump.

In her biography of Snow and cholera, Hempel (2007) entitles the chapter devoted to the Broad Street pump "Proof definitive," and entitles her book *The strange case of the Broad Street pump: John Snow and the mystery of cholera*. In her analysis of Snow, this seems to be more or as important as the "Grand experiment" (one of her chapters) in demonstrating the mechanism of cholera transmission. In many ways the Broad Street pump became the symbol of Snow's work. For example, during the annual Pumphandle Lecture in England, members of the John Snow Society remove and replace a pump handle to symbolize the continuing challenges for advances in public health. At the US Centers for Disease Control in Atlanta, when an epidemiological problem requires a rapid, straightforward solution, staff have been heard to ask, "Where's the handle to this Broad Street pump?" (Vinten-Johansen et al. 2003, 392).

Snow's natural experiment and his detective work on the Broad Street pump were *both* brilliant pieces of scientific research. This volume argues we have much to learn from the Broad Street pump example. In fact, most multimethod work involving case studies fits with the Broad Street pump model. We shall see, for example, that most multimethod scholars when doing within-case causal inference regarding a causal mechanism

[8] The idea of drawing a map based on distance to the nearest water source eventually became Voronoi tilings, which are part of the mathematical theory of tessellation (how to fill up a space with tiles).

focus on what I call the (1, 1) cases. This is a natural choice if you want to connect a causal mechanism $X = 1$ to the outcome $Y = 1$. The key to the Broad Street pump analysis was showing that almost all cases of cholera ($Y = 1$) were within the area where the pump was the closest public water source ($X = 1$).

In short, this book claims Snow by exploring his Broad Street pump methodology, how and why he selected the cases he did to explore a particular causal mechanism.[9]

Research Practice, Exercises, and Bibliographies

The approach developed in chapters 3 to 5 is essential to understanding multimethod *practice* in statistical, QCA, and game theory multimethod research. This volume features a strong interest in research practice. Most statistical research cites methods articles as justification. For research involving case studies this is far from standard practice. It is quite common for statistical multimethod articles and books to choose cases and do case studies without any reference to the methodology literature. Apparently, these practices are not too objectionable since they appear in major journals—e.g., *American Political Science Review*, *International Organization*, and *World Politics*. One goal of this volume is to make explicit the implicit methodology used and give it a critical examination. All of the chapters have significant pages devoted to research practice, as evidenced by publication in top-ranked journals and university presses.

Steve Samford and I have conducted an extensive and systematic survey of case study research practice in comparative politics (his field) and international relations (my field). We have systematically examined all articles that involve case studies—single case studies, comparative case studies, game theory, QCA,

[9] Dunning (2012, 18) appears to claim the Broad Street pump as well by using its famous map of cholera occurrence as the cover of the book. However he says that "Snow's strongest piece of evidence, however, came from a natural experiment that he studied during the epidemic of 1853–54." Dunning's main discussion of the pump case appears in the chapter "The central role of qualitative evidence" in the context of a discussion of causal mechanisms. In short, Dunning himself sees the pump analysis as a causal mechanism one.

and statistical multimethod in the top journals that publish articles using this kind of research (i.e., not *American Political Science Review*, *American Journal of Political Science*, and *Journal of Politics*, which only rarely publish research with a significant case study component). To explore the book context I have surveyed three top publishers: Cambridge, Princeton, and Cornell University Presses. This is particularly important because books give the opportunity to select more than two or three cases, which is critical for chapters 7 and 8.

This volume describes quite well most statistical multimethod research practice. Game theory multimethod practice—chapter 6—follows the logic outlined in this volume. Appendix A shows that very few scholars actually use the estimated statistical model to choose cases; instead they follow the logic of chapter 3. For example, in multimethod statistical work people almost never choose cases where $X = 0$ and $Y = 0$. This is puzzling from a statistical perspective but makes complete sense when the focus is on causal mechanisms via within-case causal inference.

For all my methods books (2005; Goertz and Levy 2007; Goertz and Mazur 2008; Goertz and Mahoney 2012) I have provided exercises for classroom and individual use. This book is no exception. I have gathered the exercises for all my books together in one file and about once a year I email an updated set of exercises (email ggoertz@nd.edu to be put on the list). The exercises are divided by topic, which means that they might cover more than one book, e.g., counterfactuals. The 2015 edition has about 350 exercises.

Some of the exercises have answers (provided for instructors), some are for discussion. Often they include extensions of the books into new but related areas. Typically the exercises refer to existing research that I am reading or teaching. They often contain the seeds of future papers and books.

I have found that they are very useful for students looking for paper topics for methods classes. Since they rely on published research, they have a strong applied flavor.

Because understanding research practice is core to this book, I have created several bibliographies. These will also be updated and available from me on request or via the Princeton University Press web page for the book.

The discussions of research practice are based on the following bibliographies:

1. Case study article bibliography. Established by Steven Samford, this includes all articles in *World Politics*, *International Organization*, *Comparative Political Studies*, and *Perspectives on Politics* in the period 2006–2015 that include one or more case studies. This includes game theory or statistical multimethod work with one or more case studies.
2. Game theory multimethod. Based on the larger bibliography by Lorentzen, Fravel, and Paine (2016), these are articles and books that have formal or game-theoretic models and one or more case studies. This bibliography is not complete or systematic but starts from Lorentzen, Fravel, and Paine and adds references as I discover them.
3. Medium-N designs. This includes books or articles with 10 or more case studies. These may or may not include other methodologies such as statistical analyses or game theory.

Statistical multimethod—statistical analyses with case studies—are very common, so there is little need for a bibliography. For example, most books with statistical analyses also include case studies (based on my incomplete survey of publications in international relations and comparative politics by Cambridge, Princeton, and Cornell University Presses, 2006–2015).

Hopscotch

Just as multiple causal paths may lead to an outcome, so too are there multiple paths through this book. Julio Cortazar published a famous novel entitled *Hopscotch*. While a traditional novel has a linear structure of chapters that are read in order, in *Hopscotch* the reader can jump around, reading the chapters as in the game hopscotch. This volume is a hopscotch methods book.

In this section I give an overview of each of the chapters, followed by some hopscotch suggestions.

Chapter 2 explores the concept of a causal mechanism and its connection to multimethod and case study research. It presents

the standard motivation for multimethod research where there is significant cross-case evidence for $X \rightarrow Y$ and the researcher wants to explore the causal mechanisms by which X produces Y, i.e., $X \rightarrow M \rightarrow Y$.

The standard $X \rightarrow M \rightarrow Y$ view of causal mechanism significantly misses the way many researchers think about causal mechanisms. First of all, typically it is not one M that is present but rather multiple M_i (see figure 1.2). Second, these multiple M_i are occurring at the same time. Third, these M_i are substitutable and combinable. Hence, to think of causal mechanisms in terms of simple causal chains in time does not capture many theories and hypotheses.

Another weakness of the $X \rightarrow M \rightarrow Y$ view is that many mechanisms involve interaction terms, such as INUS causes and constraint causal mechanisms. Minimally, we have $X_1 * X_2 \rightarrow Y$. The theory of the interaction term must be at least a sketch of a causal mechanism. Chapter 2, as well as chapters 4 and 5, explores in some detail more complex causal mechanisms involving interaction terms.

So while the standard $X \rightarrow M_1 \rightarrow M_2 \rightarrow \cdots \rightarrow M_n \rightarrow Y$ captures some theories and causal mechanisms, there are other popular setups. This becomes clearer once one draws figures representing the theories (e.g., Waldner 2015).

I use examples from international relations and comparative politics to illustrate and discuss the connection between what constitutes causal mechanisms and cross-case and within-case analyses. They involve game-theoretic, statistical multimethod, and comparative case studies. They are prominent works that have defined debates in their fields.

Examples of actual research are useful in exploring some basic varieties of causal mechanisms. They help, in particular, to illustrate the constraint causal mechanism, which focuses on explaining why the outcome did not occur. Constraint causal mechanisms almost always imply an interactive causal mechanism between motivations and constraints, i.e., constraints have no causal effect absent motivations to be constrained.

Chapter 3 argues that the central goal of case studies is analyzing causal mechanisms. Does the theorist's causal mechanism work in individual cases? How does the causal mechanism

produce the outcome? This chapter explores the basic logic of case selection when one wants to investigate how and if a causal mechanism produces the outcome.

Much work on case selection in qualitative work (e.g., King, Keohane, and Verba 1994) stresses the importance of variation on the independent and dependent variables. In contrast I focus on the various combinations of the two, e.g., $(X = 1,\ Y = 0)$ or $(1, 0)$. Some combinations are central to the analysis of causal mechanisms, e.g., $(1, 1)$, while others such as $(0, 1)$ are not very relevant. As such it is not a question of variation on X or Y separately as usually argued, but the various roles that the *combinations* of X and Y play in the multimethod enterprise.

The second half of the chapter adds complications in the form of confounding variables (or control variables) and alternative explanations. Analyses of causation have always stressed the importance of confounders; how is one to incorporate these concerns into the choice and analysis of individual cases?

Chapter 4 discusses constraint causal mechanisms. Here causal mechanisms are not about how X *produces* Y, but rather how X prevents Y from occurring or *constrains* Y. In many causal mechanism diagrams the $\rightarrow$ needs interpretation. When the arrow means "produces Y" is different from when the arrow means "prevents Y": they are different causal mechanisms with different methodological properties.

Constraint causal mechanisms are closely associated with necessary conditions. A strong constraint must be satisfied—i.e., is a necessary condition—for the outcome to occur. Necessary conditions are most useful in explaining why the outcome *did not occur*. At the same time, necessary conditions form part of the explanation for why events do occur, i.e., part of sufficient condition causal mechanisms.

Chapter 5 focuses on simple causal complexity in the form of statistical interaction terms or set-theoretic ANDs. It follows naturally from chapter 4 on constraint mechanisms. For constraints to have a causal effect there must be a motivation to break the constraint. Glass ceilings do not have an impact unless women want to move up the professional ladder. This means that constraint causal mechanisms must involve interaction

terms. This chapter also contrasts interaction-term analyses with simple additive or contributing factors models. It addresses the relatively common situation where there are two or more causal variables in the cross-case analyses.

Outside of a couple of QCA articles (Rohlfing and Schneider 2013; Schneider and Rohlfing 2013), there is almost nothing on multimethod research for more complex hypotheses. As such, chapter 5 takes a look at these issues in the simplest complex QCA model, X_1 AND $X_2 \rightarrow Y$. In addition, it focuses on various QCA-related issues such as multimethod research in the context of fuzzy logic variables and fuzzy logic falsifying cases.

Chapter 6 applies the methodology developed in chapter 3 to "game theory multimethod" research, defined as a formal model combined with case studies. The game-theoretic model provides a causal mechanism that can be empirically explored via case studies or cross-case analysis. The EITM project has focused on the game theory–statistics connection; chapter 6 looks at the game theory–case study connection. The game theory–statistics multimethod research explores the generalization question, but fails to deal with the extent to which one can find the causal mechanism of the model in individual cases. As such, game theory–statistics focuses on two corners of the research triad but not the empirical analysis of the causal mechanism corner. In terms of practice, the methodology in chapter 3 describes well what game theorists do when they include case studies in their books and articles.

Chapters 7 and 8 flip this way of thinking on its head by supposing one starts with successful causal inference in a case study and then worries about how generalizable the causal mechanism is. These chapters discuss core parts of the *multiple case study* methodology that scholars have already begun to implement. There are some central ideas: (1) Do we find the proposed causal mechanism in individual cases? (2) How often, or within what scope, do we find the proposed causal mechanism? These are questions about generalization. For example, Copeland (2015) explores individually all major power crises and wars 1790–1991 to examine how generalizable his trade expectations theory is.

Chapter 7 explores some influential articles that have used case studies and within-case causal inference to test some prominent theories or to reexamine statistical tests of theories. These have appeared in top journals and have provoked a wide-ranging debate (for example, see the special issue of *Security Studies* (2012) devoted to audience costs theories of international conflict). These critiques have relied fundamentally on a set of within-case causal analyses that often contrast with findings of cross-case statistical analyses or challenge the empirical relevance of game-theoretic models.

Chapter 8 looks at what happens when one begins with intensive causal mechanism analysis and then moves to the selection and analysis of more cases. One might start with a convincing causal mechanism analysis in one or two cases. How should one move to include additional cases in the analysis? One answer—the classic one—is to do a statistical analysis. But what about doing more case studies? How can selecting cases lead to making a strong argument about the generalizability of the causal mechanism?

Regarding hopscotch, since you are reading these lines you have begun on square 1, with chapter 1.

One possibility would be to continue with chapter 3, which outlines core arguments of the book.

The causal mechanism chapter (chapter 2) can be read at any time.

For those particularly interested in causal complexity, the causal mechanism chapter (chapter 2) followed by chapters 4 and 5 is recommended.

For those interested in two cultures arguments, read chapter 3 followed by appendix A.

Chapter 6 on game theory multimethod is available for a hop anytime after chapter 3.

One could hop to chapters 7 and 8, which form a pair, to see the new medium-N paradigm proposed in this volume.

Those interested in QCA should jump at some point to chapters 4 and 5, where I discuss necessary conditions, fuzzy logic, and interaction terms.

As in the hopscotch game and novel, there are multiple paths through this volume. Some chapters can comfortably be

skipped if not of interest, and the order can easily vary from person to person. Beyond reading chapter 3 very early on, the sequence can vary. While not a writer of Cortazar's skill, I have tried to make the various hopscotch readings of this volume work.