and Bennett discuss case studies and their relationship to typological theories. Dul and Hak talk about case studies and their relationship to linear probabilistic versus necessary and sufficient condition hypotheses. All of these books also express in various ways the rapidly growing interest in mixed and multiple methods and at the same time the need to connect methodology more closely to theoretical concerns.

The diversity of approaches to case studies means that there will be disagreements about core issues. The Lieshout contribution to this newsletter illustrates a natural and positive consequence of the flowering of work on case studies. King, Keohane and Verba devoted basically one chapter to philosophy of science and causation issues; George and Bennett make this topic central to their volume. Lieshout raises important concerns about the nature of causal mechanisms

and causation in George and Bennett. In the Gerring symposium, one point raised by several contributors is the nature of "single-outcome studies," i.e., studies that focus on explaining just one case. This raises the core issue of the role of case studies in causal generalizations and the importance of this as a goal in case study research. I suspect that this will be a continuing topic of conversation among qualitative methodologists. The Casellas essay discusses the concept of representation and its relationship to case selection and typologies. It thus also illustrates how critical issues arise at the intersection of different methodological approaches.

Finally, I am still planning to have a review of qualitative methods and research design syllabi for the next issue so please email me your syllabi or the syllabi in use at your university if you have not done so. Thanks.

---

Symposium: John Gerring, Case Study Research: Principles and Practice (Cambridge, 2007)

Case Studies Are for Intensive Testing and Theory Development, Not Extensive Testing

Michael Coppedge
University of Notre Dame

coppedge.1@nd.edu

Case Study Research is a landmark book. This culmination of years of careful thought by John Gerring is by far the best dissection of case studies in the literature, in several ways. First, it is the most comprehensive discussion. It looks at case studies from every possible angle, and in a penetrating way that exposes the term "case study" as a handy label for what is actually a great variety of methods. It also examines case studies broadly, going beyond political science to describe variants of case studies that are done in economics, psychology, and medicine. The breadth of Gerring's reading about this family of methods is extremely impressive. Second, it is clearly thought through and clearly explained. It corrects several mistaken notions about case studies. Third, chapter 7 is the most sensible and clear assessment of process-tracing that I have yet read. Fourth, because it is comprehensive and clear, it offers a new set of concepts for the different types of case studies and their goals and procedures, which could become a standard set of concepts that will make it easier for us all to debate these claims without getting tangled up in definitional issues. So it is a very important book. It's probably a bit too technical for most undergraduates (although I am assigning chapter 3 to my undergrads this semester), but it should be required reading for graduate students, especially those in comparative politics.

I have only a few outright disagreements with Gerring's arguments, and they are all about minor points. However, I do have a more significant disagreement on matters of emphasis. If I were writing this book (which probably violates the "minimal rewrite rule" [206] because I am far less well-read than Gerring is on this topic), I would want to be more categorical in my judgments. It often seems that Gerring is trying too hard to find something nice to say about every possible kind of case study. (One exception is the "most-different cases" method, which he effectively dismisses.) I would want to state outright that some kinds of case study or cross-case analysis are very useful for certain purposes but not at all for others, and some are just not worth doing.

In particular, I would make a more rigid distinction between theory development and hypothesis testing. Gerring recognizes this distinction but does not make it stick everywhere that it should. This problem arose, I think, because he chose to define "case studies" in a way that makes generalization one of their inherent purposes. A case is an element in a sample, which is drawn from a population, he reasons, so by definition, there is no point in doing a case study unless it generalizes to the population in some way. Maybe the problem is that there is an unnoticed ambiguity in the term "generalization." It can mean using a case to test whether a hypothesis is generally true, as Harry Eckstein and Douglas Dion have advocated doing. This, in my opinion, is impossible. There are no truly crucial cases in political science due to the multicausal and probabilistic nature of political phenomena, and our priors are not strong enough to support Dion's prescription. There is a kind of testing we can do with a single case, which I will discuss below. But usually the kind of generalization that one does in a case study is not testing generalizations, but hypothesizing them. It is true that the case must relate to the population to be relevant, but it relates by proposing relationships that might be generally true. But a case study cannot tell us whether they really are generally true; that requires large-sample testing within the whole domain in which the theory applies.

That kind of testing could be called "extensive testing." There is a different kind of testing, which is sometimes called
“intensive testing,” which is ideal for case studies, but it has a very different purpose and logic of inference. The goal of intensive testing is to judge which of several competing hypotheses does the best job of explaining a single case. It is therefore what Gerring discusses in the epilogue as “single-case outcome studies,” and here and there as “internal validation,” but it doesn’t get the emphasis it deserves, because it constitutes at least half of the justification for doing case studies. Unlike extensive testing, which tests the same propositions in a large number of cases, intensive testing tests a large number of propositions in a single case. The logic is, “if my theory is true, then I would expect to observe these 20 things in this case. If the alternative theory is true, then I would expect to observe these 20 different things. Using Bayesian logic, if the 20 predictions of my theory are confirmed and the 20 alternative predictions are not, there is only a very low probability that my theory is wrong, and it becomes the better explanation for this case.” It is usually impossible to quantify these probabilities, but the logic behind them is very strong, and it makes case studies a very powerful method for explaining single outcomes.

This different emphasis would alter a few of the book’s passages. For example, I endorse Gerring’s conclusion on p. 147 that “Case studies...rest upon an assumed synecdoche: the case should stand for the population. If this is not true, or if there is reason to doubt this assumption, then the utility of the case study is brought severely into question.” I think there are always reasons to doubt this assumption, so it is never safe to generalize from one or a few cases. That’s why we should use them for theory development and intensive testing rather than for any attempt at extensive testing.

Another example: In his interesting discussion of matching as a promising alternative to specifying control variables in a regression, Gerring states that simply asserting that two cases are more or less the same for the purpose of matching “can be a huge advantage over large-N cross-case methods, where each case must be assigned a specific score on all relevant control variables—often a highly questionable procedure, and one that must impose strong assumptions about the shape of the underlying causal relationship.” (133–34). Yet it is always possible to specify at least a subjective dummy variable as a control, which would be exactly as accurate as asserting that two cases match, and it is often possible to assign more precise scores for regression variables. If assumptions about the linearity of a relationship are false, they can be modified and tested. I come away convinced that matching, which Gerring explains very clearly, is a method worth trying, but I suspect, as I think he does, that it will not be as useful in practice as it sounds in principle.

A final example concerns scope conditions. I love Gerring’s call in chapter 4 (76–85) for making scope conditions explicit and non-arbitrary; this is essential. But its implications are ambiguous unless we make it clear what the scope conditions demarcate. If it is tested propositions, there is little room for arbitrariness: the scope of tested propositions is exactly as large as the sample or the case used in the test; we can’t generalize beyond it, unless it was a random sample of sufficient size, in which case we can generalize to the population. But if we are talking about how far a hunch might travel, then the scope of the hypothesis is hypothetical. It is essential to speculate about what the scope conditions may be, but we won’t really know until some extensive testing is done.

I also have one question that is unrelated to any of this. In chapter 6 (with Rose McDermott), which makes a beautiful, concise argument that an experimental logic undergirds all case studies, the most rigorous category, “Dynamic Comparison,” is defined as having both spatial and temporal variation. I wonder whether cross-sectional time-series analysis meets this criterion.

In conclusion, I think that in reality I agree with Gerring on just about everything and he agrees with me. I have quoted some passages in which he seems to have an opinion different from mine, but they are balanced by other passages that sound very close to what I have said on these issues. If we have differences, I believe they are only differences of emphasis.

---

Moving the Doormat to the Main Menu: Case Study Research Methods in the Social Science Toolkit

Evan S. Lieberman
Princeton University
est@princeton.edu

John Gerring’s motivation for his book, Case Study Research, is the same as Harry Eckstein’s writing on the same subject three decades ago: He points out that case studies are much maligned—the methodological doormat if you will—despite their recurrence in so many influential works in our field and throughout the social sciences. To address this conundrum, Gerring hopes to “restore a sense of meaning, purpose and integrity to the case study method” (66).

And I think he largely does just that. He gives scholars the potential to do case studies in such a way that any social scientist could clearly see the logic through which the analysis could generate strong causal inferences.

It is a vital and lucid work that ought to appear on any graduate research methods syllabus. As much as it is a book about case studies, it is a treatise on research design and logical thinking that updates and integrates many classic and more recent contributions. The book keeps its feet on the ground by examining a rich array of examples of completed work in political science, often with a healthy dose of pragmatism.

In my comments, I will highlight some of the novel insights found within various chapters in the book, and also raise some issues that I think warrant some additional attention, either by Gerring, today, or by him or other scholars in the future.

**Definitional Issues**

First is the question of defining the case study. If the quest is to dignify case studies, then it is necessary to know
what we are dignifying. There is a lot of ambiguity in the conventional use of the term, and in Gerring-style, we are afforded a careful, well-thought-out definitional discussion in chapter 2.

He writes that the case study is "the intensive study of a single case where the purpose of that study is—at least in part—to shed light on a larger class of cases (a population)... at the point where the emphasis of a study shifts from the individual case to a sample of cases, we shall say that a study is cross-case" (20).

By definitional fiat, he declares that case studies should be theoretically oriented, and not purely idiographic. Of course, others, particularly in other disciplines, may use the label “case study” for other purposes, but he makes clear that this is an enterprise that is social scientific. If one cites this book when doing a case study, this is a clear signal: the case study will be used to explore, illuminate, probe, or test broader propositions about specified causes and effects, even as it uses proper names and particularities in the discussion and analysis.

What Can Case Studies Do? How Are They Used?

In Chapter 3, Gerring puts his best foot forward and highlights the options and comparative advantages of this method relative to other methods. In attempting a similar task, Eckstein may have done more harm then good for the method by over-selling the inferential possibilities with case studies. Gerring both tempers and expands the use of case studies by customizing the opportunities for case study work within the research cycle, and according to the type of theory one is dealing with, particularly whether the theory is of a more deterministic or probabilistic variety.

In many ways, Gerring’s book shares a vision of what constitutes convincing evidence that is more similar to King, Keohane, and Verba’s Designing Social Inquiry than many other recent works by qualitative methodologists. While it is true that Gerring has written extensively on the subject, he is clearly not arguing that qualitative research or case studies are in any way the “superior” empirical strategies. He is measured about what he promotes. As compared with Designing Social Inquiry, however, one does not finish this book thinking, "well, really the best thing to do is to avoid case studies, and to redesign a social inquiry such that it will produce many observations leading to a dataset amenable to quantitative analysis." Gerring stresses that there are often key observations that shed disproportionate light on the connections between causes and effects, and that these must frequently be combined across different types of units in order to understand political dynamics. However, these connections are not made within the context of conventional datasets, which, by contrast, require unit homogeneity.

In this sense, the book really champions the idea of a larger division of labor—one that might be integrated within studies, or across studies, which is a view that many give lip service to but this book really defends in logical terms. A handful of key interviews or historical records may reveal the plausible links between cause and effect that complement correlational research of different types of observations such that we can be convinced about the validity of a particular proposition. Brady and Collier refer to this as causal process observations, and Gerring also identifies a complementarity with dataset observations. In any case, this is a book that could be called “Mixed-Method” research as much as it is “Case Study Research.”

Issues of Case Selection

The largest chapter of the book is chapter 5: “Techniques for Choosing Cases.” I think any researcher on the verge of doing case study research ought to go through this list of nine selection strategies and make sure they can self-consciously recognize one or more of these types as characterizing their own, and if not, that should indicate that the study could probably be re-framed in a more crisp manner.

However, I do have some questions to raise. First is the attention to the “pathway” case, as a separate type of case study. A pathway case, according to the author, is one in which the causal effect of one factor can be isolated from other potentially confounding factors. As a pragmatic concern, I don’t understand how one can choose a particular case for its potential to illuminate causal mechanisms prior to having done the research. And how is this different from some of the other strategies, such as the “crucial case” strategy or even the “typical” case strategy, which are also used for hypothesis testing given particular scores on independent and dependent variables? I suppose the identification of this strategy as just one of nine makes me wonder what we are doing in the other hypothesis-testing case study strategies.

I really appreciate the attention to deliberate case selection, and I know Gerring realizes that many case studies are done with less pre-study consideration than he recommends. But I do think more needs to be said about the role of area expertise and personal experience. Traditionally, a clear justification for the choice of cases has been that some investigators develop internal databases of contextual knowledge and measurement skills that increase the reliability and validity of the study. If true, perhaps this strategy deserves its own place— it is probably the most practiced strategy in any case, and it is hard to imagine this changing.

Relatedly, there are the examples of what I might call the “convenient” case study. This is akin to available non-probability sampling, but it is also the reality of being a social researcher. Sometimes case studies find us, we don’t look for them. We may be working or traveling somewhere or reading something, and our interest is piqued, perhaps because what we observe confirms or contradicts some prevailing theory. One might respond, “well, then you did not use a case ‘selection’ strategy per se.” Perhaps, but I am going to guess that a sizeable portion of case studies are generated this way—they are the product of life circumstances and interests, personal and professional. Thus, it may be hyperbole to speak always of case selection when we talk about case studies. What we might be able to do is to engage in case justification and use the strategies that Gerring identifies to describe what it is we have a case of.

Also, I am not sure where we would put Michael Burawoy’s
“extended case study” into this typology of case studies. In various works, Burawoy has sought to use extended and deep case studies to discover flaws in and then modify existing case studies. While it is true that in some sense the goal is slightly different from the stated one of Gerring’s book, which is to make inferences about a larger population of cases, it does share an orientation toward theoretical conclusions. But again, for Burawoy, the selection of the particular case tends to be based more on the prospects for depth of study. This approach is intended to highlight when variables were simply conceptualized poorly or relationships misunderstood. It strikes me that by depriving this type of case a real label, it is devalued in principle, while in practice, it provides potentially enormous contributions to knowledge. But to my knowledge one could select such a case without any prior knowledge of the case’s place in the distribution of explanatory or outcome variables.

Notwithstanding, the chapter does an excellent job of bringing together the diverse forms of case selection strategies that have been used by scholars, including the logic of inference associated with Mill’s methods, and integrating them within a single, comprehensible framework.

Process Tracing

In chapter 7, Gerring argues that process tracing is usually a component of case study research; it usually relies heavily on contextual evidence. He states, “the hallmark of process tracing, in my view, is that multiple types of evidence are employed for the verification of a single inference—bits and pieces of evidence that embody different units of analysis...individual observations are therefore non-comparable” (173). I like this characterization. But I wonder—and I am really putting this out here for discussion—should the properties identified for process tracing be definitional for what we mean by case studies in political science? Should we reserve the term “case study” for those studies that employ such heterogeneous evidence? Otherwise, one might call, for example, every single lab experiment a case study.

The chapter on process tracing highlights clearly some of the strong empirical findings using this approach. He makes a nice recommendation—that we ought to be able to graphically diagram an argument in a series of steps, if even in the somewhat frighteningly complex manner that Mahoney does in the case of Skocpol’s States and Social Revolutions. I agree that diagramming is a good heuristic technique and probably ought to be used as a standard.

But the chapter on process tracing is short. A mere 13 pages, one-fifth the page-matter afforded the chapter on case selection. I don’t want to overinterpret, but this brevity, I think, reveals some of Gerring’s own apprehension with the case study method as it is generally practiced or understood in political science. Gerring writes, “process tracing evidence is, almost by definition, difficult to verify, for it extends to evidence that is nonexperimental and cannot be analyzed in a sample-based format...” (184). He says that the mitigating factors for process tracing are that it is (1) supplemental; and (2) can be vetted by “experts.” He concludes, “despite its apparently mysterious qualities, process tracing has an important role to play in case-based social science...it deserves an honored place in the toolkit of social science” (185).

Well, I detect a touch of inner conflict in that last sentence. And I think much more needs to be said about process tracing. Doing case studies well is doing this kind of analytical detective work. It is hard. I don’t know if we can develop general rules and strategies. But I think we can try, and the George and Bennett volume on case studies offers some additional discussion of process tracing.

My own suggestion for advancing the technique of process tracing is to identify more tailored sets of guidelines according to theoretical content and the level of analysis under investigation, whether it be the mobilization of collective actors, the making of policy, or the development of institutions. Political scientists would benefit a great deal by breaking down a set of criteria which they believe would establish a reasonable baseline for convincing or at least acceptable process-tracing evidence, including, say, temporal proximity of links in a causal chain; the explicitness of actor intentions; and/or the types of sources used. These benchmarks would not be ironclad rules, but might provide some standard for how we could evaluate the robustness of a qualitative result, just as conventional statistical analysis has measures of statistical significance. Just as a 95 percent confidence interval is arbitrary, so would these standards be, but without them, we have no reference line for discussing the content of evidence, except for completely useless metrics like number of months spent in the field.

Single-Outcome Studies

Finally, there is a concluding chapter on what Gerring calls, “Single-Outcome Studies.” A single-outcome study is when a researcher seeks to explain a single outcome for a single case. This is an incredibly important chapter—it takes on the elephant in the room of much social science research: that research agendas inevitably get driven by real-world, often catastrophic, events, such as the 9/11 terrorist attacks. I offer Gerring the acronym SOS for single-outcome study.

But I must admit, I leave this chapter a bit confused. Are SOS’es case studies, or not? Often times the language of the chapter contrasts the SOS with the case study, but in describing the studies associated with nested analysis (e.g., Lieberman 2005) as types of single-outcome studies, well, that seems to me to clearly meet the criteria for a general case study. Indeed, the very terms used to describe those studies were “model-building” and “model-testing,” suggesting a direct engagement with the types of cross-case claims identified as central to Gerring’s ideal of the case study method. More generally, it is hard to imagine a political scientist studying a “single outcome” without some view of a larger universe of cases, so I require greater elaboration of what this type of study actually is, and how it is distinct from others.

Descriptive Inference

Finally, I want to discuss, if briefly, the explicit omission from the book, which is the task of descriptive inference. It is important to recognize that in this book, most of the science or
method of case study research comes from the strategies of case selection.

At the outset of his book, Gerring highlights that he focuses on causal inference because treatment of the "descriptive task of gathering evidence is well covered by other authors" (9). Actually, I disagree. I think that political scientists have very, very few good references on gathering appropriate data, particularly for the type of enterprise Gerring describes. The collection of observations that come from heterogeneous sources and unit of actors remains fairly ad hoc, and the task of summarizing accounts has received scanty treatment in political science.

Imagine a hypothetical study of ethnic conflict: someone is doing a case study in county X, and a survey reveals no hostile negative attitudes, but five in-depth insider accounts relate a mix of characterizations, and there is a riot in one province in which various ethnic slurs were shouted. As far as I know, the methodological literature tells us very little about how to score this case. And yet, this is the ever-present dilemma for the case study researcher working with multiple sources of data.

I don't blame Gerring for omitting this type of discussion from the book because it is really a different kettle of fish, but I did want to highlight my belief that the integrity of case study research will rest on principles of descriptive inference at least as much as on principles of causal inference. More generally, I think that the discipline has devalued case studies for the very reason that we have emphasized the value of causal inference to a much greater extent than good measurement and descriptive inference, even though we know you can't do the former without the latter.

Going Forward

To conclude, Case Study Research: Principles and Practices is at its very best in relating the possibilities for case study work in logical juxtaposition to other types of inferential strategies. It is smart, and provides study analytical scaffolding for the development of new case studies. I think it should encourage us to do more case study work with our heads held high. But we will still need to be explicit and self-conscious about how those studies get done so that we can have an even better handle on what it is about intensive study of a case that convinces us of the strength of a general proposition. Gerring has made a major contribution to social science by helping to systematize this genre of research.

Note

1 Such as Campbell and Stanley (1966) on quasi-experimental research design; Eckstein (1975) on case studies; Collier's (1991) and Sartori's (1970) statements on the comparative method and the relationship to statistical methods; and much of the more recent qualitative methods research carried out by other scholars such as Mahoney (1999); George and Bennett (2004); Brady and Collier (2004); and King, Keohane, and Verba (1994).
Among those case studies that pursue causal inference, Gerring further focuses on those that seek to generalize from the case to a larger population. He in fact defines a case study "as the intensive study of a single case where the purpose of the study is—at least in part—to shed light on a larger class of cases (a population)" (20). One implication of this definition is that work that seeks to explain a single outcome in a single case is not defined as a case study. Gerring instead considers this work to be a "single-outcome study," and he briefly considers it in the epilogue.

Should studies that primarily try to explain particular outcomes be left out of the definition of case study? I think the answer is "no." Indeed, I will argue that Gerring overemphasizes the generalizing aspects of case study research. And I will suggest that this overemphasis derives from his conviction that the merits of case studies are best evaluated according to the extent to which they approximate an experimental design. The overall consequence, I will suggest, is that Case Study Research does not address important methodological issues that apply to a significant strand of case study research in political science—namely, case studies in which the main goal of analysis is to identify the causes of a specific outcome in a particular case.

Contributions

Let me first applaud John Gerring for clarifying much of the general logic of all case studies, and especially generalizing case studies. The book includes a superb discussion of the relationship between observations, cases, samples, and population—which allows Gerring (23–25) to vividly illustrate the differences between a case study dataset, a cross-case cross-sectional dataset, and a time-series cross-sectional dataset. These distinctions, in turn, motivate an excellent discussion of different types of research design, including three types of case studies: diachronic, synchronic, and diachronic-synchronic.

The important chapter 3, "What is a Case Study Good For?," offers great insight into the comparative strengths and weaknesses of case study research. With cross-case research as the comparison group, Gerring concludes that the case study is strong at hypothesis generating and weak at hypothesis testing, strong on internal validity and weak on external validity, strong on locating causal mechanisms and weak on specifying causal effects, and strong on working with deep propositions and weak on working with broad propositions. I agree with the general thrust of these conclusions, though I want to argue below that: (1) the relative strengths of case study research derive mainly from its effort to explain particular outcomes in specific cases, and (2) the relative weaknesses of case study research are less of a concern when we realize that the goal is often mainly to explain the particular outcome.

Chapters 4, 5, and 6 are co-authored with Jason Seawright, Rose McDermott, and Craig Thomas, respectively. The chapter with Seawright offers an excellent discussion of different techniques for choosing cases. Although researchers often have theoretical reasons for looking at certain cases, this chapter addresses the methodological issues that should drive case selection. The typology of types of cases: typical, diverse, extreme, deviant, influential, crucial, pathway, most-similar, and most-different is the most sophisticated and comprehensive of its kind (of which I am aware). To boot, the chapter includes an interesting discussion of the appropriate cross-case technique for locating different kinds of cases. Likewise, the chapters with McDermott and Thomas add distinctive contributions concerning, respectively, the application of an experimental template and the use of process-tracing evidence for generating valid causal inferences. Again, they are written more with the goal of uncovering the logic of causal inference in case study research than offering a single set of procedures that analysts can or should try to follow in their research.

Concerns

My fundamental concern with the book is that it underappreciates the extent to which the primary goal of many case studies is to explain a particular outcome in a specific case. Of course, all case study researchers must draw on general knowledge, broad theory, and insights from a larger range of cases. In that sense, even the most particularizing case studies are very centrally engaged with generalization. However, the goal of the analysis is often to use general insights and individual case knowledge to explain the particular. Whether and how an explanation of the particular sheds direct light on a broader class of cases is a secondary issue.

The idea that many case studies seek to explain particular outcomes should not be controversial. What caused World War I? What caused the French Revolution? What caused sustained high growth in Korea? What caused the breakdown of military rule in Argentina? These kinds of questions are familiar in political science, and I suggest that they animate much of the case-study research in the discipline. I want to suggest that these sorts of questions are probably as common as alternatives: what does World War I teach us about the causes of war in general? What does the French Revolution teach us about the causes of revolutions in general? What does high growth in Korea teach us about the causes of high growth in general? What does the breakdown of military rule in Argentina teach us about the breakdown of military rule in general? I think Gerring favors the latter kinds of questions. But, as point of fact, I want to argue that many or most case-study scholars primarily address the former questions, turning to the latter briefly and often inconclusively (e.g., as speculative observations at the end of their studies).

Political scientists who favor general knowledge and who dismiss particular knowledge may find my argument to be discouraging for the case study method. Yet I think such a reaction is inappropriate. The social sciences must be oriented toward explaining the particular as well as making generalizations that apply to broader populations. Both kinds of explanations contribute knowledge. We cannot look to historians to develop valid causal inferences about particular outcomes. Many historians (though certainly not all) lack the training in theory and method to carry out this kind of research—their distinctive contribution to causal analysis rests
with the discovery and use of novel sources of data. The achievement of valid causal inference, even when particular outcomes are under analysis, requires the toolkit of social scientists.

If explaining a single outcome is one’s goal, then one will naturally be concerned with internal validity, mechanisms, rich explanation, and the possibility of novel hypotheses—the strengths of case study research identified by Gerring. By the same token, the comparative weaknesses associated with case study research—e.g., testing general hypotheses, external validity, insight about average causal effects, and generalization—seem less troubling when we realize the particularizing goals of much of this research. Hence, appreciating the centrality of explaining the particular allows us to better see the source of strengths case study research that Gerring himself identifies. And it makes the weaknesses of this method appear as less problematic.

However, if much case study research has as its goal the explanation of the particular, then some of the orienting assumptions of Gerring’s book need to be rethought. Most basically, questions arise about Gerring’s argument (12) that “the characteristic virtues and flaws of case study research designs can be understood according to the degree to which they conform to, or deviate from, the true experiment.” An experiment, after all, is intended to assess the effect (if any) of a given intervention on a dependent variable of interest. It is not designed to necessarily offer any kind of complete explanation of an outcome. Experiments teach about the average effects of interventions for populations, not about all of the factors that explain the outcome of interest in particular cases. In this sense, the experimental method is thoroughly predicated on the “effects of causes” approach. The same is true of statistical research, which tries to mimic an experiment in the context of an observational study. By contrast, case study researchers are often not concerned with the average effects of their causes across a large population. They have a very different research goal: to use reigning theoretical orientations, general knowledge, and novel inductive discoveries to explain the outcome of interest. This classic “causes of effects” approach stands in sharp contrast to an experimental design—and it has its own distinctive intellectual lineage, one that is not much discussed in Gerring’s analysis.

The implication of this discussion is that still more work needs to be done on the case study method. Gerring has done a remarkably good job of discussing the inner logic and methodology of case studies that seek to generalize about causal patterns. But case studies are often only secondarily interested in producing generalizations. A crucial next step is to start where Gerring’s epilogue leaves us: with case studies that primarily seek to explain particular outcomes in specific cases.

On Common Ground: Case Studies in Comparative Methodological Perspective

Rogers M. Smith
University of Pennsylvania
rogerss@sas.upenn.edu

It is perhaps worth noting at the outset that I am fascinated by discussions of methodology in the context of particular, actual research projects. But I find most books and articles on methodology to be rather tedious. Some are heavy-handed in selling the author’s pet approach, and in the case of even the best advice, I often find myself thinking of all the intellectual and logistical reasons it would have to be modified for real projects of which I’m aware. I’ve nonetheless occasionally participated in more general and abstract discussions of methodology, for two reasons: I know they are a necessary part of promoting shared understandings in our field of what constitutes good or at least legitimate political science, and I know that students entering the field do need some general guidance before they can come to grips with the specific challenges presented by their own research interests.

So we do need good methodological texts, and I’m happy to report that in terms of content, accessibility, and even “antitedium” factor, John Gerring has written an outstanding text. It is not heavy-handed but it does have interesting, important, distinctive themes, and what’s more, I have no major quarrel with them. I take at least some of these main themes to be:

1. especially when they are used to explore causal mechanisms and claims, case studies must be understood as, at least implicitly, instances of broader political phenomena, so that the difference between case studies and cross-case studies is one of degree;

2. because case studies necessarily involve analyses of many pertinent observations internal to the case, they can involve both quantitative and qualitative techniques;

3. the logic of good causal case studies is at bottom the same as that of experiments, in the sense that we are trying to determine whether a particular variable or set of variables produces an outcome that would not have occurred, all other things being equal (ceteris paribus);

4. the most basic methodological challenge really comes in the satisfaction of ceteris paribus assumptions (Gerring 2007, 171). We are looking for ways that we can plausibly say that the outcomes in which we are interested are traceable to the causes in which we are interested, but we can only really be confident of that if we have good reason to believe that, if everything else were just the same but those causes had been absent, the outcome would not have occurred. I want to stress this because, as I’ll note shortly, making ceteris paribus claims in social science research is very difficult, whether or not we use particular quantitative or qualitative techniques.
But my main response to Gerring’s book is to urge us to give greater weight to a few points that he gets right but does not stress as much as I think we should do and as I am about to do.

The points I’d emphasize are first, that quantitative techniques are always, inescapably embedded in qualitative techniques, a point often lost when we consider whether particular points should be addressed in quantitative or qualitative fashion; and second, that when we are assessing the larger significance of research projects, quantitative or qualitative, our judgments always involve some reliance on what appear here as “process tracing” and “counterfactual” techniques, as well as always corrigible ceteris paribus assumptions. I stress these points because I think we in the discipline need to remind ourselves, over and over again, that achieving greater rigor in our work can never simply be a matter of finding better statistical techniques for dealing with uncertainties or even of designing randomized field experiments, useful as those things are. Political science research inevitably rests on debatable premises about how we should define and categorize what we observe and also on debatable estimates about what particular research findings tell us about how larger political processes operate. Work can be and, I fear, often is rigorous in regard to its statistical analyses of observations but not very thoughtful about the categories used to create those observations. Research may also produce reliable causal arguments about particular patterns of political behavior, but these seem exciting only because scholars incorporate them into much broader claims about politics, explicit or implicit. Much in those broader claims is not rigorously examined or defended, or sometimes even acknowledged.

Let me elaborate those points by reference to a couple places in Gerring’s text where he makes legitimate arguments in ways that I’d have him put differently. On p. 34, Gerring has a sentence that is potentially misleading: he says “only in...case studies does qualitative analysis comprise a significant portion of the research.” He means that, once constructed, a large-n cross-case analysis doesn’t rely on qualitative methods to explore its particular causal claims, and I agree with that. On p. 41, Gerring also has a potentially misleading claim when he says “evidence gathered from a cross-case research design can be interpreted in a limited number of ways. It is therefore more reliable.” He means that once we understand the definitions and causal hypotheses deployed in a cross-case research project, it is hard to, for example, interpret findings of no statistical significance as proving statistical significance, and I also agree with that.

The reasons those sentences can nonetheless be misleading are made evident by points Gerring makes on pp. 52-53 and 69-71. There he notes that when we construct a set of observations such as “cases of democracy” or “rich” and “poor” nations and then judge particular cases as sufficiently similar to be counted together or sufficiently different to be counted apart, we are relying on theories of what political phenomena are, theories on how the world works, and judgments of what really matters that all represent what he terms “an ontological element of research design.” Subsequent quantitative research can tell us whether studies defined with certain assumptions produce statistically meaningful results, but whatever those conclusions may be, they cannot do all the work of justifying our initial categorizations and assumptions of what count as decisive similarities or differences. Our quantitative data sets are embedded in broader “visions of the world as it really is,” as Gerring puts it, and that means they are imbedded in accounts that have inevitably been arrived at through processes that have been far more “qualitative” than purely “quantitative,” if we must use that dichotomy. To be sure, particular assumptions can be subjected to quantitative empirical tests and it is often valuable to do so. But that doesn’t alter the embeddedness of qualitative research in qualitative judgments, because it is never possible to test all the assumptions that go into constructing our quantitative tests. We run into a kind of infinite regress—in principle we’d have to test all the assumptions that go into our research designs for testing the assumptions of our main research design, and then we’d have to test all the assumptions that go into our research designs for testing the assumptions of the research designs for testing the assumptions of our main research design, etc. There’s no way out from relying on some ontological judgments, as Gerring rightly stresses.

This is admittedly the old point that all research designs are “theory laden,” but I think it worth making, because it means that in cross-case studies, qualitative methods do form a significant portion of the overall research endeavor. We make partly qualitative or at least non-empirical judgments when we deem something a “democracy” and when we draw a boundary line for what constitutes a “rich” or “poor” nation, and since those are fundamental to the whole statistical inquiry, how can they not be a significant portion of the research?

This “theory-laden” feature of all political science research also means that, though within the assumptions used to construct a particular cross-case study, there may be a limited number of ways of interpreting the results, there are likely to be lots of ways of challenging the assumptions, and therefore dismissing or reinterpreting the results. If we say, for example, that the U.S. was not truly a democracy until the 1965 Voting Rights Act, lots of results based on counting it previously as a democracy have to be reinterpreted. If we say a nation is not truly “rich” whatever its GNP unless levels of economic inequality in its population are relatively low—otherwise, we might say, some elites are rich, but the nation understood as the people is not really rich—then, again, particular statistical results might be reinterpreted in a much wider range of ways.

I have similar concerns about Gerring’s discussions of “counterfactuals” and “process tracing” later in the book. He presents these as methods that, despite real limitations, deserve to be “in the toolkit of social science.” Again, though I agree with his discussions of these methods and their limitations, I think his formulation may overstate the role that counterfactuals and process tracing inevitably and always play, not in constructing particular research projects, but in our broader disciplinary assessments of the significance of those research projects. Just as we can’t, even in principle, rigorously quantify and test all the assumptions that go into our research
designs, we can’t ever test ALL the assumptions that we employ when we decide what to make of a particular, solid research finding. Take, for example, the influential findings of Green and Gerber (2004) in randomized field experiments, that voters are more likely to respond when contacted personally door-to-door than when they receive a recorded phone message or a postcard. That is a rigorously arrived-at, reliable finding; but what are to make of its significance?

For lots of campaigns that were investing heavily in recorded phone messages, the lesson has seemed clear: spend more money on door-to-door. But note what they’re also assuming. They’re assuming that campaigns that contact voters in some fashion are more likely to win elections than those that do not—a plausible assumption, but not one that is self-evident if most voters actually respond to certain sorts of dramatic events or media presentations or informal “word of mouth” discussions far more than any actual campaign contact. It might still be wiser to arrange the crisis (perhaps by tempting your opponent in an airport men’s room) or to find the magic media message or seek to spark a “word of mouth” campaign. Another, perhaps more telling point: those involved in campaigns also assume that winning elections is important, presumably because doing so will put into power people and policies they wish to have in power. But perhaps power and policies can be more effectively won outside electoral processes, through lobbying, gifts, coercion, blackmail, or simply by becoming so essential to the economy that everybody has to pay attention to you however they got elected. If that’s true (and I am not asserting that it is), it doesn’t undercut the Green and Gerber finding, as far as it goes; but it greatly alters our sense of the importance of that finding for understanding politics.

And I believe that point can be generalized: any time we attribute broader significance to a research finding, we are implicitly rejecting a whole set of untested counterfactuals that might undercut not the rigor or reliability, but the broader import we are attaching to the research; and we are also implicitly adopting an understanding of how the political world works that has features like those Gerring rightly attributes to process tracing—reliance on heterogeneous sorts of evidence to support various chains in the causal claims and on many untested “contextual assumptions and assumptions about how the world works” (Gerring 2007, 185). Though we can and should seek to find evidence for some of those assumptions, we are never going to be able to test them all—some will be left at the level of “background knowledge,” which, as Gerring states, “informs all causal analysis.” When we judge the broader significance of particular research projects, we inevitably place in the context of a broader account of political causes, some of which rests on unexamined “background knowledge.”

It may be said that the points I’m making—there are qualitative, judgmental, interpretive elements that are ineradically part of all research designs and that are ineradically part of all judgments of the significance of research findings—are true of all types of scientific research, and I agree. But the ontological, interpretive, and causal assumptions that go into constructing and assessing political science research tend, not surprisingly, to be more politically charged and controversial than those for many other kinds of research—so I think we have a special obligation to pay attention to these features of our work. As an example, Gerring says on p. 70 that “the social sciences are defined by their focus on decisional behavior—actions by human beings and humanly created institutions that are not biologically programmed.” But lots of the social sciences do in fact focus on behavior that is claimed to be biologically programmed, and some social scientists in the past and present believe that all behavior is ultimately biologically programmed. The assumptions we make on just how and how far we are biologically programmed are enormously relevant to the research projects we think worthwhile constructing and on how we interpret their significance—and again, though we can test some of our assumptions about biological determinism in specific research projects, we can never test them all. So our work is inevitably premised on and interpreted to advance different visions of human agency in ways that may well be more fraught for the conduct of our political lives than other sorts of research are likely to be.

Let me close with a final point in the same spirit. In making the preceding remarks, I may seem to have moved away from Gerring’s main topic, the case study. But I take the broader significance of Gerring’s elucidation of case study methods to be an indication of what case studies can and cannot contribute to social science, written, as he stresses, in part because “the methodological status of the case study is still highly suspect” (7). His book does an excellent job of presenting what case studies might and might not be good for and why in regard to causal arguments (as he acknowledges, he presumes without discussion the value of case studies in suggesting fresh descriptions of political phenomena). I think that his arguments are strengthened by stressing, even more than he does, that all social science research projects, cross-case research as much as case studies, are constructed on what most in the field would view as “qualitative” conceptual foundations and interpreted through broader understandings of how the world works that are for the most part not subjected to rigorous testing, instead representing the conventional world views that constitute “everyday” or “background knowledge” or “common sense,” world views of a sort that historically have often come to be discredited over time, at least in part.

That point is only underlined by recognizing that all our research designs involve ways of making credible “ceteris paribus” claims—because these claims are perhaps especially difficult to make in the social sciences. Scholars like Jamie Druckman have been arguing, for example, that we should be very careful about moving from lab experiments to claims about how people process political information in the far more complex environment of actual everyday life, because the lab context is not likely to approximate all the factors that may influence such processing in more conventional settings. Scholars like Lynn Sanders (forthcoming) and Sarah Igo (2006) have stressed how survey research, too, is an unconventional context producing results that may well not map onto what happens in the settings of actual political decision-making (though
they may then influence behavior in those settings). Even randomized field experiments ultimately rely in part on a measure of faith—that all the unobserved, unstudied factors that might influence outcomes, of which there are inevitably a large number, are randomly distributed among our test groups. It may always turn out down the line that this wasn’t the case, that ceteris was not paribus, so we cannot regard the results of even these research designs with complete confidence.

Because the sorts of characteristics and limitations I’ve been detailing are features of all political science research, including case study research and cross-case research, they admittedly say nothing about the general question of whether case studies or cross-case studies are preferable. But they do indicate that case studies are not necessarily inferior because they involve qualitative elements, including counterfactuals and process tracing and imperfect means of making ceteris paribus claims. And keeping these features of our work in mind also can help us to consider, when we get down to actual research projects, whether or not a particular case study is more wisely constructed than a cross-case study that speaks to the same general topics. I think that if we do so, we will find that many case studies continue to make valuable contributions to the progress of political science, as Gerring says. And his book will help us all to do them better.

References


Sanders, Lynn M. Forthcoming. Interracial Opinion in a Divided Democracy.

On the Road to Consensus?
A Reply to Coppedge, Lieberman, Mahoney, and Smith

John Gerring
Boston University
jgerring@bu.edu

First off, let me thank the participants in this symposium for their insightful comments. The first nightmare of every author—no one will read my book—is clearly assuaged. Of course, this serves to raise the second nightmare of every author: Christ, they’ve actually read it!

I am relieved to discover that there are many points of agreement among the participants of this symposium with respect to the methodology of the case study. This is not so surprising, since I appropriated the wisdom of the ages—including work by members of this symposium—in writing the book. Even so, one does not always find consensus, even when one is seeking it. Perhaps this is a sign that we are beginning to realize a degree of cumulation in the field.

To be sure, whatever consensus may exist in the field does not yet extend to the definition of key concepts, which remains hotly contested, as this symposium attests. However, I would suggest that beneath the semantic debates there is increasing consensus over key methodological principles. In the present instance, while there is disagreement over what definition we ought to adopt for “case study,” I venture to guess that for any given definition the authors represented in this symposium would agree on a similar set of methodological precepts. That is, “If case study means X, Y follows. If, on the other hand, case study means Z, Q follows.”

If I am correct in this diagnosis, then much of the debate that we are now experiencing in the field is largely semantic in nature. I do not wish to imply merely semantic, for surely the definition of key terms is an important issue. But let us not overstate the Sturm und Drang.

I shall follow convention in focusing my comments on areas of apparent disagreement. This is usually more enlightening, or at any rate more entertaining. In doing so, I shall endeavor to combine issues addressed by several authors. These fall into six general categories: (1) the distinction between studies that are theory generating and those that are theory testing, (2) the distinction between case studies and single-outcome studies, (3) the problem of representativeness, (4) the experiment as a template for case study research design, (5) pragmatic considerations impinging upon the choice of cases, and (6) process tracing. In a final section I briefly discuss a grab-bag of additional issues about which I have less to say.

Theory Generating and Theory Testing

Michael Coppedge suggests that we maintain a clear distinction between work whose purpose is theory-generating and work whose purpose is theory-testing. He argues that case studies—understood here as a generalizing form of analysis in which the scope of the inference extends to a larger population—are useful for the former but not for the latter.

I agree that the case study strategy is usually more compelling when employed in an exploratory mode than in a confirmatory/disconfirmatory mode (as discussed in chapter 3). I would not go so far as to dismiss the use of case studies in the latter mode, however, even when the purpose is to validate (or invalidate) an inference that extends to a larger population of cases.

Consider, for starters, that where the population of an inference is small, each individual case matters to a degree that it does not (all other things being equal) where the population is large. If one is generating insights into country-level relationships operative within Western Europe, for example, an intensive investigation of a single case, or several cases, may be sufficient to refute a theory. This is a simple matter of numbers. One case within this restricted population constitutes roughly one-twelfth of the population (depending upon how one defines “Western Europe”), two cases one-sixth, and so forth. The case study mode is virtually axiomatic. (I
presume here that the small scope of the theory can be justified by the author, and is not an arbitrary domain restriction.)

A second circumstance is found in a population where variation on key dimensions (the outcome and/or the explanatory variable of interest) is limited, i.e., a theory about rare events. Again, it seems virtually axiomatic that those cases where variation occurs (or at least some of them) should be studied intensively, i.e., in a case study fashion, and that the results of that study might lead one to alter one’s priors on the theory in question.

A third circumstance involves populations that are large, with ample variation on key parameters, but where evidence (e.g., quantitative data, archival sources, access to key informants) is available only for one or several sources, effectively prohibiting a cross-case approach to the problem at hand. Here, too, the case study format may be employed in a theory-testing manner—not because it is so methodologically compelling but rather because the alternative mode of investigation is simply not feasible.

A different sort of opportunity for case study research in the theory-testing mode is provided where the theory in question elaborates specific and testable hypotheses about causal mechanisms. It is often the case that mechanisms linking the exogenous cause with the outcome of concern can only be effectively tested in a small number of cases, perhaps because of data insufficiencies. This is sometimes referred to as “pattern matching” (following Donald Campbell). (Coppedge explores the logic of this style of analysis, but does not view it as viable when testing a general theory.)

So, Coppedge and I agree that a cross-case format for theory-testing is superior. But I think Coppedge fails to acknowledge that it is not always possible, by virtue of limitations encountered in the research domain. In these circumstances, a case study mode of theory-testing is eminently justifiable; it is, indeed, the best available method. Whether it produces a high level of confidence in results—a decisive acceptance or rejection of the theory—may be doubted. But the same skepticism also usually greets theory-testing in a cross-case mode. Indeed, even where the sample is large and the treatment can be randomized across groups, results from a single test are rarely sufficient to prove or falsify a general proposition (because of doubts about replicability or external validity). Confirmation/disconfirmation is perhaps better articulated as a relative matter—a narrowing of confidence intervals or decreasing variance between prior and posterior distributions.

Case Studies and Single-Outcome Studies

Mahoney argues that a primary use of the case study method is to explain single cases rather than a population of cases. If so, my definition of the case study effectively defines out of the subject its largest and most important area of employment. On one level, the issue is semantic—how shall we choose to define the term? In French, analyse de cas implies the analysis of a single outcome within a single bounded unit. In English, however, when one says one is conducting a case study one is usually implying that a broader population

is at issue. A case study is a study of a subject broader than the case itself. If, let us say, the object of one’s investigation is to explain the outbreak of World War I, an author will say that she is writing a study. If, on the other hand, her objective is to shed light on wars in general (or some subset of wars), then the same study of World War I will be articulated as a case study. Thus, I think that my usage is consistent with ordinary (English) language, and a fortiori, with language within the (Anglo-American) social sciences.

Whatever one’s choice of terms, it is absolutely vital to distinguish among studies that seek to elucidate the causal features of a single outcome and those that seek to elucidate the features of a broader population of cases (using one or several cases as examples). In order to distinguish these different methodological objectives, I have coined a new term—single-outcome study—for the former. It matters not whether others adopt this term; it does matter (contra Lieberman), that we distinguish these two forms of analysis. Despite some commonalities, they are distinctive, as the epilogue makes clear.

But let us address the larger, and surely more important, point. Perhaps single-outcome studies are the best defense, and best employment, of the case study method. The point is well taken. If one’s interest is to explain World War I, one really ought to conduct an intensive study of World War I. This goes more or less without saying—or so I thought. Perhaps it should have been granted greater emphasis. Even so, one should not neglect the ways in which a broader population of cases—perhaps interrogated through large-N cross-case analysis—might also help to shed light on that particular outcome, a matter that both Coppedge and Lieberman have written about. Within-case and cross-case techniques for evaluating single-outcomes are thus explored in the epilogue.

Before moving on, I should call attention to a pervasive ambiguity. Coppedge points out that case studies are often employed to answer the following genre of question (which he refers to as intensive testing): Does theory X explain case Y? Or, more open-endedly, which extant theory best explains case Y? The ambiguity is that the resulting study may be regarded either as a test of a general theory (a case study, in the sense in which I employ the term) or as an attempt to explain a single case (a single-outcome study, using my vocabulary). Often, both moments are present in the same study. Thus, Graham Allison’s famed study of the Cuban Missile Crisis may be read as a case study or a single-outcome study.

One’s view of this debate is necessarily colored by a larger question that I have not sought to resolve—namely, whether studies of individual cases are or ought to be focused on broader populations or on outcomes specific to each case. I assume that the latter style of “idiographic” analysis is typical of the humanities and of traditional historical studies. I believe that the former is more resonant with the goals of the social sciences (at least in the Anglo-American world), where the objective is usually to construct general models of human behavior, rather than (or in addition to) the accumulation of knowledge about specific cases. We are a species that is drawn—some might say prone—to theory. If so, then this book quali-
fies as a methodology appropriate for case studies in the social sciences, but not (or to a lesser extent) for case studies as they are approached in history and the humanities, or on the interpretive end of our discipline.

The Problem of Representativeness

Coppedge points out that there is always reason to doubt whether a sample of one (or two or three) is representative of a much larger population. This is central to his skepticism about the viability of case studies when employed to test theories (as opposed to generating hypotheses). I concur. However, it is important to clarify that the problem of representativeness (i.e., bias) is a product of three factors: the size of the sample, its method of selection (random selection, relative to the dimensions that might affect the outcome of interest, is superior), and the presumed homogeneity of the underlying population. Briefly, a larger sample is better, though increases in size bring diminishing returns. Randomization is superior to purposive selection if the sample is large, though purposive elements may be incorporated into the randomization procedure (via stratification of the sample) and purposive case selection can overcome some doubts about representativeness, as discussed in chapter 5. Finally, claims of representativeness are more questionable wherever the population of interest is highly heterogeneous; likewise, issues of sample size and sampling techniques are less important where the underlying population is highly homogeneous. Thus, Wittgenstein felt justified in choosing any native speaker of a language (e.g., his maid) in order to discover its underlying properties; there is, in other words, no value in taking a random sample of units that are effectively uniform with respect to the phenomenon under study.

I raise these familiar issues only to underline that problems of representativeness are not unique to case study forms of investigation. Consider the problem faced by crossnational studies where the population is (a) limited (to 190 or so nation-states), (b) impossible to sample from (without running into degrees-of-freedom problems and problems of sample bias), and (c) highly heterogeneous. Indeed, the whole notion of sampling (and with it the statistical theory that it undergirds) is questionable in this context.

Consider the proposition that federalism contributes to economic growth. What is the true population of this common inference? Is it (a) societies within the developing world? (If so, one is compelled to divide the world into developed and undeveloped societies, a difficult matter.) Is it (b) democracies only (in which case the division between regime types must be policed)? Is it (c) all nation-states from 1900 to 2007? Is it (d) all nation-states since the founding of nation-states (another ambiguous dividing line)? Or perhaps (e) all possible nation-states (including those that exist as well as those that might have existed)?

These questions are vexing because none of these putative populations can be tested in a rigorous manner. They are matters of presumption—“ontologies,” to use Peter Hall’s much-abused term. (The same points are made eloquently by Smith in his commentary.) All are plausible, but some seem more plausible than others—depending upon one’s theory of federalism and growth. Of course, one can test various hypotheses. My point is that such tests, by themselves, are not always definitive. One cannot conclude, for example, that because there is a strong empirical relationship between federalism and growth in the developing world—and not in the developed world (or vice versa)—that the theory is valid in the former but not the latter. One must be wary of defining the population of an inference as “that sample which yields the best results.” Evidently, the author must provide a coherent theoretical argument for why the inference is valid in context A but not in context B. And this takes us beyond the realm of empirics, tout court.

The Experimental Template

Mahoney suggests that the experimental template, introduced in chapter 6, is inappropriate for case study analysis because case studies—unlike experimental studies—are strong on internal validity, and weaker on external validity. Yet, surely, this is a similarity rather than a difference. Indeed, experimental work is often criticized for its lack of generalizability. By contrast, large-N cross-case samples drawn from “natural” (non-experimental) data usually have a stronger claim on external validity (because samples are drawn randomly from natural settings), but weaker claims on internal validity. My usual response to claims based on large-N representative samples drawn from surveys of the general public is that while it might not be true for the sample, if it is true for the sample it is also likely to be true for the population.

(Note that because the treatment is manipulated by the researcher in an experimental research design it is usually possible—and always desirable—to have large treatment and control groups. This moves the style of analysis into a large-N cross-case format, and explains why the traditional equation of case study research with observational data is generally true—though not true by definition.)

Pragmatic Considerations

Lieberman raises the point that cases are often chosen for pragmatic reasons, e.g., because a writer has special funding, access to, or knowledge about a particular research site. The same pragmatic factors, of course, also often come into play in large-N cross-case research. One hears—informally—that research on a topic was spurred by an author’s discovery of a new dataset, learning a new technique of analysis, or the development of new software (making easy what had heretofore been impossible or computationally expensive). We recognize these pragmatic imperatives, which may be justified on grounds of efficiency. We look for keys under lamp-posts because that is where the light shines.

However, they do not constitute methodological defenses of a particular research design. Because I happen to speak a certain language, or be conversant with a particular quantitative technique, is no reason for employing that linguistic or statistical technique in a particular instance. The latter choice must rest on methodological grounds, if it is to be justified at all. Thus, to say that I am studying Syria be-
cause I speak Arabic, or come from Syria, is akin to saying that I am employing a hierarchical linear analysis because I know how to do it (and I have a nifty software program that allows me to do it, just by punching a single key). Neither should be taken seriously, except as a matter of expediency. Both are varieties of the well-known methodological syndrome: "I have a hammer and now I am looking around for a nail to hit with it."

Problems—not techniques or special skills—should lead the way in our investigations, as Smith emphasizes in his comments for this symposium (and in other published work).

It is important, therefore, to distinguish carefully between justifications that are prudential and those that are properly methodological. This is why pragmatic concerns are given short shrift in the book.

Having said this, I realize—and the book recognizes—that case-selection strategies are sometimes justified methodologically only after the fact. This is particularly so when the point of the investigation shifts from a context of discovery to a context of verification (from theory-proposing to theory-testing). And it also the case in situations where cases are chosen by reason of expediency. I see no problem with this (though strict Popperians will demur), so long as there happens to be a strong ex post facto justification for choosing this case rather than another. Again, the question of interest rests on methodological issues, not expediency. So, I agree with Lieberman that the problem of (ex ante) case-selection sometimes resolves into the problem of (ex post) case-justification. The same might be said for many large-N cross-case analyses, which also exhibit a continual back-and-forth (soaking and poking) between methods and results.

Process Tracing

Lieberman cites insufficiencies in the chapter on process tracing—the last chapter in the book, the shortest, and perhaps the least satisfying. There is a mystical quality to this technique of analysis, which has an empirical component but rests largely upon background assumptions about the case and/or about human behavior in general. Because it is so heavily context-dependent, it is not clear whether any general methodological precepts apply. That is, one cannot tell a good process-tracing study from a poor one unless one knows a great deal about the case under investigation. It is these particular judgments about plausible counterfactuals—what A would have done, or could have done, at a particular point in time—that render this form of analysis convincing or unconvincing. By contrast, other aspects of the case study fit more neatly into the conventional methodological frameworks that we apply, for example, to large-N cross-case research (either experimental or observational). So, while I recognize the defects in my own treatment of this issue—and both Craig Thomas (who collaborated on this chapter) and I wish we could have gone further—our defense is that there may not be too much more to say from a general methodological perspective.

Few rules or principles, beyond those that we set forth in chapter 7, seem to apply. Or we have not yet stumbled across them. (The apparent obviousness of the subject matter may prevent us from accurately perceiving what is going on.)

Lieberman also suggests that process tracing, rather than being an element of case study analysis, may be its defining characteristic. As with the previously vetted debate (over whether the term is properly arrogated for studies that generalize rather than particularize), there is plenty of scope for rival definitions. However, I find this particular definitional choice might not be very salutary. In effect, it would constrain the sort of evidence and analysis that one would be able to consider (by definitional fiat) whenever one undertakes the intensive study of a single case. Insofar as process tracing is a tool of case study analysis—but not the tool—a broader definition of the term “case study” is justified.

Additional Points, Briefly Considered

I agree with Smith’s admonition (echoing Giovanni Sartori) that a great deal of conceptualizing and theorizing goes on prior to (or coincident with) a scholar’s empirical engagement with a subject. This sort of “qualitative” exercise (in the sense that it takes place with concepts and a priori logic rather than with datasets) underlies all research, whether case-based or based in large-N cross-case samples. As such, it lies outside the scope of the book, which does not purport to cover the subject of methodology generally, but only one particular method. But it is an important, and neglected, point nonetheless.

I agree with Lieberman and Mahoney that the focus of this book on causal analysis omits an important category of case-based work whose purpose is largely descriptive (i.e., interpretive). In this, I followed the lead of the broader field of methodology, where we have yet to come to grips with the challenge posed by descriptive inference (a subject that I am currently working on).

Coppendge wonders whether a cross-sectional time-series analysis would qualify as a dynamic comparison? Of course it would, but it would not qualify as a form of case study analysis, and as such is not envisioned as part of the typology presented in chapter 6.

Lieberman wonders whether the pathway case is a truly distinct method of case selection. I have argued (in chapter 5) yes. This method builds on the logic of the crucial case, first elaborated by Harry Eckstein. However, since Eckstein envisioned the crucial case—and its softer versions, the least- and most-likely cases—as studies that would test the existence of causal relationships, it appears that we require a different term for situations in which the covariational relationship (the causal effect) has been proven through cross-case analysis, but we still don’t know much (or are uncertain) about the causal mechanisms that might be at work. In these situations, one promising technique focuses on one or a few cases with the following characteristics: (a) the causal factor of interest has a large effect on this case (as judged by the changing residuals for the case when this factor is entered and removed from the cross-case model) and (b) the case lies close to the predicted outcome in the full model (its residual is small when the causal factor of interest is included, along with other relevant controls). This case-selection technique is quite different from others sketched in chapter five. Whether it will
prove useful to scholars is another matter, and remains to be seen (see Nome 2007).

Finally, a mea culpa. In the book, I appropriate the term "nested analysis" (from Lieberman) for a style of single-outcome analysis that employs large-N cross-case evidence. Clearly, Lieberman has in mind both generalizing and particularizing styles of case study when he uses the term. (I should have chosen Coppedge’s term—nested induction—which is consistent with what I had in mind.) One of the gratifications of post-publication symposia is the opportunity to set the record straight. Sorry, Evan!

Reference

Representation Revisited: Concepts, Typologies, and Case Selection

Jason P. Casellas
University of Texas, Austin
casellas@mail.utexas.edu

Empirical scholars in political science have generally allowed normative theorists to conceptualize key concepts such as democracy, accountability, and representation (Collier and Adcock 1999; Pitkin 1967). For some empiricists, taking the time to revisit the very concepts that they are purportedly measuring and testing seems at best too philosophical, and hence out of their domain (although see Goertz 2006 for a comprehensive treatment of social science concepts). Consequently, the empirical literature on representation has focused too heavily on statistical roll-call analyses, which to a certain degree can help us ascertain the extent to which legislators represent their constituents in legislatures and Congress. Substantive representation, however, involves much more than how legislators vote. In order for political scientists to understand why, we must think carefully about what representation involves. This essay will examine the concept of representation by briefly considering what normative theorists such as Jane Mansbridge and Hanna Pitkin have said about the subject, and then analyzing the concept of representation through the lens of what more recent empirical researchers have said about concepts, typologies, and case selection.

Normative Conceptions of Representation

Starting with Pitkin (1967), political scientists have regarded representation as either descriptive, substantive, or symbolic. Descriptive representation refers to citizens being represented by legislators who share particular demographic characteristics (race, gender, or ethnicity), while substantive representation involves legislators representing citizens’ interests or particular preferences. Symbolic representation refers to descriptive representation without the substantive component. Scholars of black representation have debated the merits of which type of representation is most effective, with Swain (1993) arguing that substantive representation is what really counts, while Mansbridge (2003) places more value on descriptive representation. Mansbridge, however, is a normative theorist who has argued that descriptive representation is essential for advancement of minorities and women in the American political system. Pitkin’s analysis did not really deal with minority representation, but the concept she presented has been extended to such studies. To date, no work on racial representation has challenged Pitkin’s conceptual framework or analyzed the concept of representation using more recent empirical research on methods.1

The Concept(s)

What does it mean to be represented? What does it mean to represent others? We are always asking others to represent us either before a lawmaking body, a court, or other institution. When one is represented by an attorney before a court, the attorney acts in the material interests of her client. It does not matter whether the attorney looks like her client physically. As long as the attorney defends her client well, then the client will be satisfied. In terms of political representation, however, surely more is at stake than just material interests. As Aristotle observed long ago, political issues deal with how we ought to order our lives together in the larger community. Additionally, politics deals with how individuals will be treated, including policies such as affirmative action, immigration, and English Only laws, just to name a few. This distinction is crucial because when Pitkin talks about descriptive and substantive representation, she is referring to political representation. As Gerring explains, concepts are not static and “progress in the cultural sciences occurs, if it occurs at all, through changing terms and definitions” (Gerring 2001). This may seem like a minor point because a normative political theorist is obviously referring to political representation. When we are trying to think clearly about concepts, though, it is important to consider the ladder of abstraction in order to avoid conceptual stretching and, at the other extreme, narrowing our concepts to infinitesimal degrees. Additionally, issues of race and representation have dramatically changed since Pitkin’s exegesis, which requires conceptual revisions.

As Gerring points out, it matters how we define our terms when we conduct empirical research. When researcher A speaks about democracy, she may mean something completely different than researcher B. While King, Keohane, and Verba (1994) insist that researchers choose observable and thus testable concepts, many of the most intriguing questions in the study of politics involve seemingly unmeasurable concepts. One cannot deny that the study of terrorism is appropriate for political scientists, yet understanding what motivates terrorists is hardly a simple question since it defies many of the primary motives for human behavior, such as utility maximization, and self-preservation (KKV 1994). Instead of giving up on conceptualizing difficult concepts, it would be better to tackle some of the most difficult substantive political problems and develop a