

Conflict Management and Peace Science

<http://cmp.sagepub.com/>

Case Studies: Types, Designs, and Logics of Inference

Jack S. Levy

Conflict Management and Peace Science 2008 25: 1

DOI: 10.1080/07388940701860318

The online version of this article can be found at:

<http://cmp.sagepub.com/content/25/1/1>

Published by:



<http://www.sagepublications.com>

On behalf of:



Peace Science Society (International)

Additional services and information for *Conflict Management and Peace Science* can be found at:

Email Alerts: <http://cmp.sagepub.com/cgi/alerts>

Subscriptions: <http://cmp.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://cmp.sagepub.com/content/25/1/1.refs.html>

>> [Version of Record](#) - Feb 1, 2008

[What is This?](#)

PRESIDENTIAL ADDRESS

Case Studies: Types, Designs, and Logics of Inference

JACK S. LEVY

Department of Political Science
Rutgers University
New Brunswick
New Jersey, USA

I focus on the role of case studies in developing causal explanations. I distinguish between the theoretical purposes of case studies and the case selection strategies or research designs used to advance those objectives. I construct a typology of case studies based on their purposes: idiographic (inductive and theory-guided), hypothesis-generating, hypothesis-testing, and plausibility probe case studies. I then examine different case study research designs, including comparable cases, most and least likely cases, deviant cases, and process tracing, with attention to their different purposes and logics of inference. I address the issue of selection bias and the “single logic” debate, and I emphasize the utility of multi-method research.

Keywords case studies, comparable cases, multiple-method, process tracing, research design

Introduction

The study of peace and war cuts across many disciplines and across theoretically and methodologically defined research communities within political science. While scholars within different research communities have long worked in isolation from each other, many have increasingly come to believe that the cumulation of knowledge is furthered if scholars try to learn from and build upon work conducted in other research communities. This is true for method as well as for theory, and we see a growing trend toward multi-method research in the study of international conflict and in international relations more generally. This is evident in the continued integration of formal and statistical approaches and in the growing interest of each in incorporating case study analyses into multi-method research designs. An increasing number of doctoral dissertations involve multi-method research.

One obstacle to bridging existing methodological divides is the rapid growth of the methodological literature in various research communities throughout the social sciences. This is true for qualitative methods as well as for statistical and formal methods.¹ As a

Address correspondence to Jack S. Levy, Department of Political Science, Rutgers University, 89 George Street, New Brunswick, NJ 08901-1411. E-mail: jacklevy@rci.rutgers.edu

¹After a wave of influential work in the 1970s (Lijphart, 1971, 1975; Eckstein, 1975; Campbell, 1975; George, 1979), there has been an explosion of qualitative methodology in the last decade, particularly after the publication of King, Keohane, and Verba (1994). The growing interest in qualitative methods is reflected by convention panels, publications, graduate courses, and by the success of the Qualitative Methods section of APSA and of the Arizona State University Institute for Qualitative Research Methods), where student attendance increased from 45 in 2002 to 127 in 2007 (Elman, 2008). Reflecting the growing appeal of multi-method research, the above-mentioned section and Institute now refer to “Qualitative and Multi-Method Research.”

result, the best qualitative research is more theoretically driven and methodologically self-conscious than it was three or four decades ago. It has much more potential for contributing to the cumulation of knowledge, by itself and in conjunction with formal and quantitative methods. The common view that good case study research lacks a method is unwarranted.²

On the assumption that greater dialogue across research communities is facilitated by greater familiarity, I summarize some important developments in qualitative methodology. Given the expansive literature and limited space, I focus on comparative and case study methods,³ and more specifically on those that aim to produce causal explanations based on a logically coherent theoretical argument that generates testable implications. This includes the vast majority of contemporary case study research relating to peace and war, but it excludes postmodern narratives and other analyses that reject the possibility of making causal statements or of bringing empirical evidence to bear on the question of their validity.

There are many good general reviews of case study methodology (George & Bennett, 2005; Bennett & Elman, 2006, 2007; Mahoney & Goertz, 2006), and there is no need for another one at this time. Instead, after a brief discussion of definitions, I suggest a new typology of case studies based on their research purposes. I then analyze different research designs that advance these various objectives.

What Is a Case Study?

Despite the widespread use of case study methods throughout the social sciences, **no consensus has emerged as to the proper definition**, either of a case or a case study (Ragin & Becker, 1992; Gerring, 2007: chapter 2). Most of us probably think of a case study as an attempt to understand and interpret a spatially and temporally bounded set of events. With the shift of political science toward a more theoretical orientation in the last three decades, qualitative methodologists began to think of a **case as an instance of something else, of a theoretically defined class of events**. They were willing to leave the explanation of individual historical episodes to historians, and to focus instead on how case studies might contribute to the **construction and validation of theoretical propositions**. To this end, George (1979) argued that case study researchers should adapt the method of the historian but convert descriptive explanations of particular outcomes to analytic explanations based on variables. George and Bennett (2005: 5, 17) build on that conceptualization and define a case as **“an instance of a class of events,”** and a case study as **“the detailed examination of an aspect of a historical episode to develop or test historical explanations that may be generalizable to other events.”** Thus a central question to ask of any case study is “what is this a case of?”⁴

From this perspective, now the dominant one among qualitative methodologists, a historical episode like the Cuban missile crisis is not itself a case, but different aspects of the Cuban missile crisis are cases of broader, theoretically defined classes of events, such as coercive diplomacy, crisis management, the operational codes of political leaders, etc. This conception of case studies is explicit in the method of “structured, focused comparison,” which George (1979) defined as the use of a well-defined set of theoretical questions or propositions to structure an empirical inquiry on a particular analytically defined aspect of a set of events.⁵

²Maoz (2002, 164–165) may be correct that many case studies are still “free-form research where everything goes.” That, however, is not a reflection of the method but of the individual research scholar. The utility of a method should be evaluated in terms of best practices.

³Other qualitative methods include ethnography, elite interviews, macrohistorical analysis, and “qualitative comparative analysis” based on Boolean and fuzzy set methods (Ragin, 1987, 2000).

⁴Gerring (2007: 13, 19–20) defines “case study” as “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).”

⁵One problem with this conception of a case as an instance of a broader class of events is that it excludes studies that aim to explain or interpret a single case but not to generalize beyond the

It is important to note that “case” is not equivalent to “observations.” In an early critique of case study approaches, Campbell and Stanley (1966) argued that case studies are inherently limited in their ability to establish causation because of the “degrees of freedom” problem, with many potential causal (and control) variables but only a single case or small number of cases. Lijphart (1971) made the same argument, and suggested various strategies for increasing the “N/V” ratio in order to better emulate the inferential logic of experimental and large-N statistical methods (see also King et al., 1994: 217–28).

Campbell (1975) later retracted his argument, as did Lijphart, 1975. In his discussion of cross-cultural research designs, Campbell (1975: 181–82) argued that a theory designed to explain key differences between cases “also generates predictions or expectations on dozens of other aspects of the culture. . . . In some sense, he has tested the theory with degrees of freedom coming from the multiple implications of any one theory.” Nearly all qualitative methodologists now conceive of a case to include many observations on the same variable,⁶ and emphasize that one of the main tasks of case study analysis is to generate as many testable implications of one’s hypotheses as possible in a given case (King et al., 1994).⁷

Many conventional treatments equate case studies with a narrative approach, but that is too restrictive. We can think of detailed studies of individual “cases” that incorporate substantial statistical analysis, often with the aim of generalizing to other cases. The studies of perception and misperception in the 1914 crisis by North (1967) and by his colleagues (Holsti, 1965) might be examples. Thus Gerring (2007: 33–36) suggests that the association of case study analysis with a qualitative approach is a “methodological affinity, not a definitional entailment.”

Typology of Case Studies

Most typologies of case studies reflect some variation of Lijphart’s (1971: 691) categories of atheoretical, interpretive, hypothesis-generating, theory-confirming, theory-informing, and deviant case studies and Eckstein’s (1975: 96–123) categories of configurative-idiographic, disciplined-configurative, heuristic, plausibility probe, and crucial case studies. Such typologies combine research objectives and case selection techniques, and consequently they result in nonparallel categories. A deviant case study is a research design or case selection technique for the purpose of refining or replacing an existing theory or hypothesis, and thus serves the objective of hypothesis generation. Eckstein’s (1975: 104–08) heuristic case studies, designed to “stimulate the imagination,” also serve a hypothesis-generating function. Crucial case studies are most/least likely case designs for the purpose of hypothesis testing.

We can construct a simpler and more useful typology by focusing on the theoretical (or descriptive) purposes or research objectives of a case study and distinguishing those from various research designs or case selection techniques used to advance those objectives. The basic typology consists of *idiographic* case studies, which aim to describe, explain, or interpret a particular “case” and which can be either inductive or theory-guided; *hypothesis generating case studies*; *hypothesis testing cases*, which combine Lijphart’s theory-confirming and theory-informing categories; and *plausibility probes*, which are an intermediary step between hypothesis generation and hypothesis testing and which include “illustrative” case studies. These are ideal types, and in practice case studies often combine several of these aims, often (and preferably) in sequence as a part of a multi-stage research program, one that may involve other methods.

case. Gerring (2007: 187–210) tries to get around this problem by distinguishing case studies from “single-outcome studies” involving a purely idiographic analyses of a single historical episode.

⁶The Cuban missile crisis, for example, includes many observations of coercion, crisis management, signaling, etc.

⁷Thus case study researchers reject Eckstein’s (1975: 85) definition of a case as “a phenomenon for which we report and interpret only a single measure on any pertinent variable.”

Idiographic Case Studies

The aim is to describe, explain, interpret, and/or understand a single case as an end in itself rather than as a vehicle for developing broader theoretical generalizations. The work of most (but not all) historians falls into this category. We can identify two subtypes, depending on the degree to which the analysis of the case is guided by an explicit theoretical framework.

Inductive Case Studies

Inductive case studies, which Verba (1967) and Eckstein (1975: 96–99) label configurative-idiographic and which Lijphart (1971) labels atheoretical, are highly descriptive and lacking an explicit theoretical framework to guide the empirical analysis. Inductive case studies often take the form of “total history,” which assumes that everything is connected to everything else and which consequently aims to explain all aspects of a case and their interconnections.⁸

I prefer “inductive” or perhaps “descriptive” to Lijphart’s (1971) “atheoretical” label, since a purely atheoretical analysis is inconceivable. In the absence of an explicit conceptual framework the analyst’s unstated theoretical preconceptions and biases structure the interrogation of the case. Thus, few contemporary scholars would embrace the epistemology underlying what I call the “Dragnet” conception of history: “Just the facts, ma’am, just the facts” (Levy, 2001: 52). Still, we have all read enough history to be able to distinguish between descriptive histories about the sequence of events (“who said what to whom”) and analytic histories that are explicitly structured and focused by theoretical concepts and hypotheses.

Theory-Guided Case Studies

Theory-guided case studies are also idiographic, in that they aim to explain and/or interpret a single historical episode rather than to generalize beyond the data.⁹ Unlike inductive case studies, they are explicitly structured by a well-developed conceptual framework that focuses attention on some theoretically specified aspects of reality and neglects others. Many efforts by political scientists to explain the origins of World War I or the end of the Cold War fit this category, as do some explicitly realist, Marxist, and feminist historical analyses.¹⁰

Although many have argued that social scientists ought to leave the explanation of individual cases to historians and focus exclusively on the task of constructing and testing generalizable theories, I think this argument reflects an excessively narrow view of the potential contributions of social science. While historians’ training in archival research gives them a comparative advantage in the conduct of inductive case studies, that advantage does not

⁸ As Hobsbawm (1997: 109) argues, “basically all history aspires to . . . ‘total history,’” in which the analyst “. . . cannot decide to leave out *any* aspect of human history *a priori*.” In contrast, theory-driven social science adopts a partial equilibrium perspective and assumes that the benefits of simplification by focusing on a restricted set of variables and relationships will outweigh the costs (Lake and Powell, 1999: 17).

⁹ These are also called interpretive (Lijphart, 1971: 691), disciplined-configurative (Eckstein, 199–104), and case-explaining (Van Evera, 1997: 74–75) case studies.

¹⁰ The idiographic/nomothetic distinction, often mistakenly equated with the distinction between work that is theoretical and work that is not, is best defined in terms of what one is trying to explain rather than how one explains it. Both inductively driven interpretations and theory-driven explanations of individual cases are idiographic, while attempts to generalize beyond the immediate data are nomothetic. Most historiography is idiographic and most social science is nomothetic, but what really separates the two is the fact that social scientists are much more sensitive than are historians to the question of how to construct research designs that maximize the ability to make inferences beyond the data (Levy, 2001).

extend to theory-guided case studies, where social scientists' explicit and structured use of theory to explain discrete cases often provides better explanations and understandings of the key aspects of those cases than do less structured historical analyses. The more case interpretations are guided by theory, the more explicit their underlying analytic assumptions, normative biases, and causal propositions; the fewer their logical contradictions; and the easier they are to empirically validate or invalidate. While political scientists should not make explaining individual cases their primary goal, neither should they abandon that task to historians. As I have argued elsewhere, "history is too important to leave to the historians" (Levy, 2001).

Hypothesis-Generating Case Studies

Unlike idiographic case studies, which aim to describe, interpret, or explain an individual historical episode, hypothesis-generating case studies aim to generalize beyond the data. They examine one or more cases for the purpose of developing more general theoretical propositions, which can then be tested through other methods, including large-N methods. Given their close proximity to and familiarity with the data, case study analysts are well positioned to suggest additional explanatory and contextual variables, causal mechanisms, interaction effects, and scope conditions (Collier, 1999).¹¹

It is important to note, however, that hypothesis-generating case studies contribute to the *process* of theory construction rather than to theory itself. Theory, defined as a logically interconnected set of propositions, requires a more deductive orientation than case studies provide. Thus Achen and Snidal (1989: 145) argue that "... the logic of comparative case studies inherently provides too little logical constraint to generate dependable theory," and they complain that case studies have "too often ... been interpreted as bodies of theory."¹²

Case studies can be particularly useful in explaining cases that do not fit an existing theory, in order to explain why the case violates theoretical predictions and to refine or replace an existing hypothesis or perhaps specify its scope conditions. Beyond their role in their analysis of "deviant cases," which I discuss later, case studies can help refine and sharpen existing hypotheses in any research strategy involving an ongoing dialogue between theory and evidence. A theory guides an empirical analysis of a case, which is then used to suggest refinements in the theory, which can then be tested on other cases (through statistical as well as case study methods). This is quite explicit in George's method of structured, focused comparison, and it is illustrated in George and Smoke's (1974) analysis of the multiple paths to deterrence failure. The "analytic narratives" research program (Bates et al., 1998) is driven by a more formally specified theory but follows a similar research strategy. This pattern of a continuous interaction between theory and evidence in an alternating sequence of conjectures and refutations (Levy, 2007b) characterizes the evolution of the democratic peace research program. This suggests that case studies can play different functions at different stages of a research program.

Another widely emphasized contribution of case studies to the process of hypothesis generation involves the specification of causal mechanisms, which is a leading research

¹¹Qualitative methodologists commonly argue that small-N qualitative researchers are far more inclined than are large-N statistical researchers to establish the scope conditions of their theories, whereas large-N researchers aim to establish more universal propositions based on statistical tests that incorporate the largest possible populations and hence the maximum statistical power (Mahoney and Goertz, 2006). There is some truth here, but those making this argument push it too far. Recent arguments that quantitative researchers should abandon large-scale regression analyses that incorporate many dummy and control variables into the analysis of large populations, and instead focus on particular subsets of cases and limit the use of control variables (Achen, 2005; Ray, 2005), for example, reflect a sensitivity to scope conditions (Levy, 2007a).

¹²A similar argument can be applied to statistical or experimental findings.

objective of many case study analysts.¹³ Qualitative researchers have long argued that the methodology of *process tracing* (George, 1979), which involves an intensive analysis of the development of a sequence of events over time, is particularly well-suited to the task of uncovering intervening causal mechanisms and exploring reciprocal causation and endogeneity effects. By focusing on what Brady et al. (2004: 12) call “causal-process observations,” case study researchers get inside the “black box” of decision making and explore the perceptions and expectations of actors, both to explain individual historical episodes and to suggest more generalizable causal hypotheses. Many propositions about bureaucratic politics, for example, originate in Allison’s (1971) intensive study of the Cuban missile crisis. Process tracing can also contribute to the testing of certain theoretical propositions, which I discuss in a later section.

The comparative method can also contribute to hypothesis generation, but most qualitative methodologists emphasize its hypothesis-testing function, and I discuss it further in that context. Thus Liphart (1975: 159), while emphasizing that “the primary function of the comparative method is to test empirical hypotheses,” argues that “a comparative *perspective*—not to be confused with the comparative method—can be a helpful element in discovery.” Similarly, Stretton (1969, 246–247) argues that “The function of comparison is less to simulate experiment than to stimulate imagination. . . . Comparison is strongest as a choosing and provoking, not a proving, device: a system for questioning, not for answering.”

Hypothesis Testing

While many scholars question the utility of case studies for hypothesis testing, qualitative methodologists emphasize that well-designed case studies can play a role in testing certain types of hypotheses. Eckstein (1975) emphasized the hypothesis-testing contributions of crucial case studies based on most/least likely case designs, and Lijphart (1975: 164) actually *defined* the comparative method as a “method of testing hypothesized empirical relationships. . . .” I discuss the comparative method, process tracing, and other research strategies more fully in a subsequent section on research designs for hypothesis testing.

Plausibility Probes

A plausibility probe is comparable to a pilot study in experimental or survey research. It allows the researcher to sharpen a hypothesis or theory, to refine the operationalization or measurement of key variables, or to explore the suitability of a particular case as a vehicle for testing a theory before engaging in a costly and time-consuming research effort, whether that effort involves a major quantitative data collection project, extensive fieldwork, a large survey, or detailed archival work. As Eckstein (1975: 110) suggested, plausibility probes can be “cheap means of hedging against expensive wild-goose chases, when the costs of testing are likely to be very great.” Thus plausibility probes are generally nomothetic in orientation, since the analyst probes the details of a particular case in order to shed light on a broader theoretical argument.¹⁴

“Illustrative” case studies, which are quite common in the international relations field and in the social sciences more generally, also fit this category. Such case studies are often quite brief, and fall short of the degree of detail needed either to explain a case fully or to test a theoretical proposition. Rather, the aim is to give the reader a “feel” for a theoretical argument by providing a concrete example of its application, or to demonstrate the empirical

¹³Causal mechanisms are central to scientific realist epistemology, which is often invoked by case study researchers to validate their approach (George & Bennett, 2005: Chapter 7).

¹⁴A similar logic could apply to the role of plausibility probes in theory-driven idiographic studies.

relevance of a theoretical proposition by identifying at least one relevant case (Eckstein, 1975: 109).¹⁵ Plausibility probes are particularly useful in combination with formal models or statistical analyses,¹⁶ but they can also be used to set up more intensive case studies.

If applied in a methodologically self-conscious way, plausibility probes can serve an important function in theory development, particularly in a multi-stage research strategy. In practice, however, the plausibility probe concept is often used rather loosely. Scholars who recognize the growing expectation in the discipline for theoretically and methodologically self-conscious empirical work but who have yet to think through the theoretical purposes of their case studies often invoke the “plausibility probe” category as a legitimizing device. This skeptical view is captured by the unsolicited comment of a colleague in response to my reference to the frequent classification of case studies as plausibility probes: “That is everybody’s favorite kind because it is a cop-out.”

The fact that the plausibility probe category is often used loosely as a residual category should not detract from the potentially important role that properly conceived plausibility probes can play in the process of theory development. Indeed, one might argue, as Eckstein (1975: 108) did over three decades ago, that social science would be well served by the use of a greater number of plausibility probes as an intermediary stage before moving directly from hypotheses construction to time-consuming empirical tests.

Varieties of Case Study Research Designs

Having identified different types of case studies, defined in terms of their research objectives, I now turn to a variety of case study designs or selection strategies. Qualitative methodologists increasingly insist that scholars doing case studies must justify their selection of cases in terms of theoretical criteria. Considerations of “intrinsic interest” or “historical importance” are no longer regarded acceptable criteria for case selection, unless the aim is the purely idiographic one of explaining a particular case as an end in itself.

Even if the aim is to explain a single case, however, there might be some advantages in including additional cases. A compelling explanation of an individual case requires both demonstrating that the hypothesized explanation fits the evidence in the case *and* that it fits the evidence better than do leading alternative explanations (George & Bennett, 2005; Maoz, 2002), which is logically equivalent to controlling for extraneous variables. Demonstrating that the outcome in question is the causal effect of the hypothesized explanatory variables and not of other factors is often difficult to establish with enough confidence by operating within the confines of the case itself (Ray, 1995: 134). Additional leverage can often be brought to bear by going beyond the case, since nearly all interpretations of a case have testable implications for related cases. We could learn a lot about the causes of World War I, for example, by analyzing why the numerous crises in the decade before 1914 did not lead to a general European war.

Turning to more nomothetic research goals, there are fewer rules of case selection for the purposes of hypothesis generation than for hypothesis testing, corresponding to the notion that there is a logic of scientific confirmation but not of scientific discovery (Popper, 1965). One common strategy for exploratory case studies of this kind is to focus on cases with extreme values on independent or dependent variables (defined in term of deviation from a mean or mode), based on the logic that causality ought to be clearest in cases where variables take on their extreme values.¹⁷ Equally common is the strategy of looking at

¹⁵This is the empirical equivalent of an “existence proof” in mathematics.

¹⁶Examples include Bueno de Mesquita (2000), Reiter and Stam (2000), and Fortna (2004).

¹⁷If one wants to get ideas about the causes of revolution, France would be a good place to start. The logic assumes linear relationships.

several cases that maximize variation across the independent and/or dependent variables, based on the logic that comparison stimulates the imagination. Such strategies might be useful for exploratory studies in the earliest stages of research. More useful, particularly after the researcher has constructed preliminary hypotheses and perhaps conducted a large-N statistical analysis, is a strategy of examining deviant cases (discussed below). The aim is to explain *why* observed outcomes do not fit the theory, and subsequently to modify the theory for subsequent testing or perhaps to establish its scope conditions.¹⁸

Some issues of case selection that are important in hypothesis testing are of less concern at the hypothesis-generating stage. One is warnings about “selecting on the dependent variable” and other forms of selection bias, which is less critical at the hypothesis-generation stage (leaving questions of efficiency aside). Another is the appropriate number of cases for investigation. Although more cases are generally better than fewer cases (up to a point) for the purposes of hypothesis testing, since they enhance control over extraneous causal influences, the same is not always true for hypothesis generation. If the population of cases about which we want to theorize is relatively small (hegemonic wars or revolutions in advanced industrial states, for example), it may be preferable to have fewer cases. The more cases used to construct a theory, the fewer that remain for testing it, since tests can only be conducted on cases (or aspects of cases) that were not used to construct the theory. Thus if the population of cases is small, inferential leverage at the hypothesis-testing stage is inversely related to the number of cases used at the hypothesis generation stage. The logic is similar to that underlying “cross validation” in statistical studies (Picard & Cook, 1984). The data is partitioned into subsets; the model is initially estimated on the first or “training” set; and the remaining sets are used to validate the model and test its predictive capacity.

Let us now turn to hypothesis-testing case studies, for which the careful selection of cases is the most critical. I begin with a general discussion of selection bias.

Selection Bias

While random sampling is central to most statistical analysis, there is a consensus that random selection will often generate serious biases in small-N research, and that the analysis of a small number of cases requires the careful, theory-guided selection of nonrandom cases (King, Keohane, & Verba, 1994: 124–128; Collier et al., 2004; Gerring, 2007: 87–88). This raises dangers of selection bias. Besides the obvious problem of picking a case that was used to generate the hypothesis, or picking a case because it fits the researcher’s hypothesis, one potentially serious problem is over-representing cases from either end of the distribution of a key variable. This is particularly serious when it involves cases with extreme values on the dependent variable (especially “no-variance” designs, in which all cases have the same outcome), because it underestimates the strength of causal effects (King et al., 1994: 128–139).

This problem of “selecting on the dependent variable” applies to comparisons among a small number of cases as well as to statistical studies, and qualitative researchers now routinely incorporate “negative cases” into their case research designs. They have also begun to analyze the utility of different selection criteria for negative cases (Mahoney & Goertz, 2004; Gerring, 2007: chapter 5). They argue, however, that selecting on the dependent variable is not a problem for process-tracing within case studies, which does not involve comparisons and which follows an arguably different inferential logic (George & Bennett, 2005: 22–25; Collier, Mahoney, & Seawright, 2004: 96; Bennett & Elman, 2006: 460–463).

¹⁸Gerring (2007: Chapter 5) describes these as *extreme*, *diverse*, and *deviant* strategies of case selection for hypothesis generation, and also mentions *typical* cases. Deviant cases are defined with respect to causal propositions, and the others only with respect to the values of variables.

We need to qualify the injunction against selecting on the dependent variable. If the hypothesis in question posits necessary conditions, the only observations that can falsify the hypothesis in question are those in which a particular outcome of the dependent variable occurs despite the absence of hypothesized necessary condition. A no-variance design of selecting on the dependent variable would be appropriate.¹⁹ Causal propositions positing sufficient conditions can be falsified by cases in which the condition is present but where the hypothesized outcome is absent, and thus an optimal test involves a no-variance design on the independent variable.²⁰ If measurement error is low, even a single case can falsify a hypothesis that posits necessary or sufficient conditions.²¹ Similarly, a small number of cases can falsify probabilistic hypotheses of the form “*x* is nearly always followed by *y*” or “*y* is nearly always preceded by *x*” (Dion, 1998).

It is true, of course, that social science offers relatively few non-trivial bivariate relationships positing necessity and sufficiency, at least for large populations of cases. There are some, however, as illustrated by the proposition that joint democracy is a sufficient condition for peace (Ray, 1995) or the proposition that social revolutions will not occur in the absence of either peasant revolts or state breakdown (Skocpol, 1979). Claims of necessity and sufficiency are much more common in interpretations of individual historical outcomes (Goertz & Levy, 2007). They are also common in theories involving multiple conjunctural causation (Ragin, 1987) or equifinality (George & Bennett, 2005: 25–27), in which there are multiple paths to an outcome and the presence of one condition might be a necessary condition for the impact of another variable along that particular causal path. Such theories often posit “INUS” causes, defined as factors that are an insufficient but necessary part of a condition that is itself unnecessary but sufficient for the outcome to be present (Mackie, 1974).

Another form of selection bias is unique to scholars who rely on secondary historical accounts in their case studies. Historians do not produce theoretically neutral analyses. If case study researchers rely on historians who share their own set of analytic biases, then the data upon which they rely (unconsciously or otherwise) may predispose them toward certain theoretical interpretations (Lustick, 1996). Although the potential for bias cannot be completely eliminated, it can be minimized if scholars make a serious effort to test their explanations against alternative interpretations. This is facilitated if the researcher, before conducting empirical research, specifies leading alternative interpretations, the observable implications of each, and the evidence that would lead them to accept or reject of each explanation, including his own (Maoz, 2002: 469–470; Gochal & Levy, 2004). Ideally, the researcher should select cases where the predictions of alternative interpretations contradict those of his/her own explanation.²²

¹⁹Bueno de Mesquita (1981) adopted this strategy in his large-N analysis of the proposition that a positive expected utility is a necessary condition for war.

²⁰Scholars have invoked Bayesian logic in a debate about the utility of no-variance designs for testing hypotheses involving necessary or sufficient conditions (Seawright, 2002; Clarke, 2002; Braumoeller & Goertz, 2002; Goertz, 2006).

²¹I define falsification in the Bayesian sense of significantly reducing our confidence in the validity of a hypothesis.

²²Larson (2001: 337–343) argues that researchers can minimize these “secondary” biases by conducting their own archival research (though the researcher’s own biases would still be present). This raises the question of the tradeoff between the intensive examination of a small number of cases (through archival research) and the more extensive but less detailed examination of a larger number of cases. This is the tradeoff between internal and external validity, which is also applied to comparisons between case studies and statistical approaches. As Skocpol (1984: 382) argues, “a dogmatic insistence on redoing primary research for every investigation would be disastrous. It would rule out most comparative historical research.”

Comparable-Case Research Designs

The first criticism quantitative researchers raise regarding the use of case studies for theory testing is that the number of variables (including necessary control variables) often exceed the number of cases, creating a degrees of freedom problem that leaves outcomes causally underdetermined. Lijphart (1971: 685) initially favored strategies that increased the N/V ratio, but later emphasized that the comparative method followed a different logic than did statistical or experimental methods, that it imposed controls not by partial correlations but by selecting comparable cases, and that it worked most effectively with a small number of comparable cases (Lijphart, 1975: 163).

Most qualitative methodologists now accept this conception of the comparative method as a strategy for conducting research of naturally occurring phenomena in a way that controls for potential confounding variables through careful case selection and matching rather than through experimental manipulation or partial correlations (Freundreis, 1983: 255). This suggests that the logic of inference is quite similar in statistical and comparable case methods (though not necessarily in process tracing, which I discuss later), though their specific research designs are different.²³

Different case selection designs involve different strategies of matching. In his *System of Logic* (1970 [1875]), John Stuart Mill suggested two closely related methods for the empirical testing of theoretical propositions: the method of difference and the method of agreement. The *method of difference* selects cases with different values on the dependent variable and similar values on all but one of the possible causal variables, while the *method of agreement* focuses on cases that are similar on the dependent variable and different on all but one of the independent variables.²⁴ Similar logic underlies Przeworski and Tuene's (1970) concepts of "most different" and "most similar" systems designs. A *most different systems* design, which corresponds with Mill's method of agreement, identifies cases that are different on a wide range of explanatory variables but similar on the dependent variable, while a *most similar systems* design, which corresponds to Mill's method of difference, identifies cases that are similar on a wide range of explanatory variables but different on the value of the dependent variable. The basic inferential logic of the two designs is the same—to identify patterns of covariation and to eliminate independent variables that do not covary with the dependent variable. Most different systems designs eliminate extraneous variables that vary across cases, while most similar designs eliminate extraneous variables that do not vary across cases.

A major problem confronting any comparable case research design is the difficulty of identifying cases that are truly comparable—identical or different in all respects but one. This goal is often easier to approximate in longitudinal designs involving a single state over time—where political culture, political structure, history, rivalries, historical lessons, etc. change very slowly if at all—than in most cross-case designs. George and Bennett (2005) use the label of the "congruence method" (a subset of structured, focused comparison) for this kind of within-case comparison. Additional inferential leverage can be gained from a combination of longitudinal and cross-sectional designs.²⁵

Comparable case designs also face the problem of causal complexity. While Mill's methods work fine for bivariate hypotheses involving a single explanatory variable,

²³Thus Collier, Mahoney, and Seawright (2004: 94) refer to cross-case comparisons as "intuitive regression." Note the parallels between the strategy of matched cases in case study methodology and the recent attention to matching in statistical methodology (Ho et al., 2007).

²⁴Mill's *method of concomitant variation* combines the methods of agreement and difference.

²⁵Snyder's (1991) study of imperial overextension combines comparisons of the behaviors of different states, of different individuals within the same state but in different bureaucratic roles, and of the same individuals over time.

particularly if measurement error is low, they are more problematic in situations involving complex causation involving interaction effects, and particularly if there are several different sets of conditions that may lead to the same outcome (Ragin, 1987; Lieberman, 1992). Under such conditions Mill's methods can lead to spurious inferences if they are used mechanically or not supplemented with the use of within-case methods like process tracing to rule out spurious inferences.

Ragin (1987) describes such causal complexity as "multiple conjunctural causation." He argues that standard statistical methods cannot easily deal with this phenomenon because the number of interaction terms necessary to capture combinatorial effects increases rapidly with the number of variables and quickly overwhelms the degrees of freedom.²⁶ Ragin develops "qualitative comparative analysis" based on Boolean algebra to identify and test combinatorial hypotheses. This framework is particularly useful in dealing with hypotheses positing necessary or sufficient conditions, but is more problematic if a theory posits probabilistic causation. This led Ragin (2000) to apply fuzzy set methods to capture uncertainty inherent in causal complexity involving necessary conditions. Other qualitative methodologists develop explanatory typologies or *typological theory*, which focuses on configurative causation in terms of different combinations of variables rather than on the average causal effects of variables across cases (Ragin, 2000: 67–82; George & Bennett, 2005; chapter 11; Elman, 2005).

Process Tracing

Comparable case strategies involve inter-case comparisons and/or intra-case comparisons (including longitudinal comparisons) and are fundamentally correlational. Boolean and fuzzy set methods, which involve the classification of a case into categories or fuzzy sets, follow a similar comparative logic. All such methods face the problem of demonstrating that observed patterns of covariation reflect a causal relationship. While incorporating the proper controls (whether statistically or through matching) can help eliminate some causal inferences, case study researchers generally emphasize the role of process tracing or "causal process observations" (Brady et al., 2004: 12) in providing additional evidence about cause and effect. Process tracing can "make up for the limitations of . . . controlled comparison . . . (and of) Mill's methods of agreement and difference," and it is "particularly useful as a supplement in large-N statistical analyses" and for "obtaining an explanation for deviant cases" (George and Bennett, 2005: 214–215).

Process tracing has a comparative advantage in the empirical analysis of decision making at the individual, small group, and organizational levels, including the analysis of leaders' perceptions, judgments, preferences, internal decision-making environment, and choices. It can also be useful in exploring other kinds of theoretical propositions, which often generate observable implications for which process tracing often has a comparative advantage in investigating. One of the implications of the democratic peace proposition, for example, is that political leaders differ in their perceptions of democracies and autocracies, and that these differences have a significant impact on behavior. Process tracing in case studies is well-suited for such questions.

Process tracing can be effectively combined with other methods (experimental, statistical, comparable cases, and the most/least likely and deviant case strategies described below), in order to examine empirically the alternative causal mechanisms associated with observed patterns of covariation. Many attempts to combine large-N statistical studies with case studies involve process tracing (Walter, 2002; Sambanis, 2004). Similarly, many attempts to combine formal modeling approaches with case studies also utilize process tracing,

²⁶For new statistical approaches to modeling interaction effects see Braumoeller (2004).

in part to validate the preferences and decision-making calculus attributed to political leaders and other actors (Bates et al., 1998; Brams, 1994; Bueno de Mesquita, 2000).

Process tracing can also be useful in the empirical analysis of various forms of complex causation, which have been attracting growing attention by qualitative methodologists.²⁷ The analysis of critical junctures and path dependence, for example, are extremely sensitive to the accurate identification of the precise timing of these key turning points (Pierson, 2000). Process-tracing case studies can make a critical contribution in providing a more precise measurement of these critical junctures and tipping points in individual cases (Tarrow, 1995: 474).

Crucial Case Designs

Crucial case studies, based on *most-likely* or *least-likely* designs, can be useful for the purposes of testing certain types of theoretical arguments, as long as the theory provides relatively precise predictions and measurement error is low (Eckstein, 1975: 113–123). Most/least likely designs are based on the assumption that some cases are more important than others for the purposes of testing a theory. They are implicitly based on a Bayesian perspective in which the weight of the evidence is evaluated relative to prior theoretical expectations (McKeown, 1999; Bennett & George, 2005). If one's theoretical priors suggest that a particular case is unlikely to be consistent with a theory's predictions—either because the theory's assumptions and scope conditions are not fully satisfied or because the values of many of the theory's key variables point in the other direction—and if the data supports the theory, then the evidence from the case provides a great deal of leverage for increasing our confidence in the validity of the theory. Similarly, if one's priors suggest that a case is likely to fit a theory, and if the data confound our expectations, that result can be quite damaging to the theory. The inferential logic of least likely case design is based on the “Sinatra inference”—if I can make it there I can make it anywhere. The logic of most likely case design is based on the inverse Sinatra inference—if I cannot make it there, I cannot make it anywhere (Levy, 2002: 442).

Inferential leverage from a least likely case is enhanced if our theoretical priors for the leading alternative explanation make it a most likely case for that theory, while inferential leverage from a most likely case is maximized if our priors make the case least likely for the alternative theory. The logic of inference in most/least likely case analysis is asymmetric. Evidentiary support for a theory from a least likely case or lack of support from a most likely case provides substantial theoretical leverage, and induces a significant shift in our confidence in the theory. Evidentiary support for a theory from a most likely case or lack of support for a least likely case, however, leads to only a modest shift in one's confidence in the validity of a theory.

A good example of a crucial case design is Allison's (1971) application of his three models of foreign policy decision-making to the Cuban missile crisis, which he framed as a most likely case for the rational unitary actor model of foreign policy decision making and simultaneously a least likely case for alternative organizational process and governmental politics models. The fact that the evidence appeared to contradict many predictions of the rational unitary actor model but to fit predictions of the organizational process and governmental process models increased scholars' confidence in the generalizability of the organizational and governmental politics models. If Allison had picked a case of noncrisis decision making or budgeting, evidence consistent with models II and III would not

²⁷ See the special issue of *Political Analysis* (2006). Also Goertz and Mahoney (2005).

have been surprising and consequently would not have significantly altered scholars' prior assessments of the broader validity of those models.²⁸

This discussion of most and least likely case study research designs, in conjunction with our earlier discussion of the role of case studies in testing hypotheses positing necessary or sufficient conditions, suggests that a small number of case studies, and possibly even a single case, can be quite valuable for the purposes of testing certain types of theoretical propositions. As Gerring (2007: 115) argues, they provide "the strongest sort of evidence possible in a nonexperimental, single-case study." The argument holds, however, only if the theory yields fairly precise predictions, if the researcher specifies in advance the kinds of evidence that would lead him to accept or to reject the theory, and if cases are selected in a way that maximizes leverage on the theory.

Deviant Case Designs

Deviant case study research designs focus on observed empirical anomalies in existing theoretical propositions, with the aim of explaining *why* the case deviates from theoretical expectations and in the process refining the existing theory and generating additional hypotheses. Thus deviant case designs serve the primary purpose of refining existing hypotheses. The logic of inquiry is similar to the examination of the residuals in a statistical analysis. In fact, a deviant case approach can be usefully combined with statistical methods, since the most significant deviations from the regression line in a statistical analysis are ideal cases for selection for more thorough examination by case studies. The examination of deviant cases is not the end of inquiry, as the theory refined on the basis of deviant case analysis must be subject to subsequent testing against new evidence, whether in either large-N or small-N analysis, by applying the revised hypotheses to other cases or to unexamined aspects of the same case (Lakatos, 1970).

It is conceivable that a detailed examination of deviant case will lead a researcher to conclude that the case does not violate the theory's core predictions—because of measurement error, inappropriate operationalization of key concepts, failure to incorporate important contextual variables, recognition that the case falls outside of the theory's scope conditions, or for other reasons. The result, though not necessarily the intention, of such an inquiry is essentially to "save" a theory from damaging evidence.²⁹ Because deviant case research designs can result in rescuing a theory from potentially damaging evidence, they can contribute to hypothesis testing as well as to hypothesis refinement and generation.³⁰

The verification that the operational indicators of a theory's key concepts have been properly specified and measured correctly, and also that the case satisfies the scope

²⁸Although some argue that most/least likely research designs are appropriate only for case studies and not for large-N studies (Gerring, 2007: 121), the logic is more generalizable. If we pick a subset of the population where a theory is most (least) likely to be valid, conduct a statistical analysis, and find that the evidence disconfirms (supports) the theory, then we can use most/least likely logic to leverage our inferences from the data. An example is Levy and Thompson's (2005) quantitative study of balancing in Europe, which they argue is a most likely case for balance of power theory.

²⁹Gerring (2007: 105–115) labels this an "influential case" selection strategy and distinguishes it from a deviant case strategy. I prefer to collapse the two categories, since the design is the same, and whether an anomalous case is actually consistent with a theory is determined only as a result of the empirical investigation. Paradoxically, the researcher either demonstrates that a case is not really deviant, or, by refining the theory to eliminate the anomaly, eliminates its status as deviant. Thus the "deviance" of a case is a function of the stage of a research program. The purpose of deviant case analysis is to eliminate the set of deviant cases.

³⁰The researcher should be alert to the danger of subconsciously using a deviant case analysis to dismiss evidence that disconfirms his/her own preferred theory.

conditions of the theory, plays a central role in deviant case analysis. The measurement validation function of case studies is also important in the analysis of “borderline” cases. The aim is to check for the possibility of measurement error in key variables that might affect the classification of cases or the validity of the unit-homogeneity assumption (King et al., 1994: 91–94).³¹

An illustration is efforts by democratic peace researchers to ascertain whether certain cases fit the categories of joint democracies or of wars. Most of these investigations (Ray, 1995) conclude that borderline cases either involve a nondemocracy or fail to qualify as a war. Different assessments of the measurement of these variables in these and other contested cases in democratic peace research would have been particularly damaging to the dyadic democratic peace hypothesis because the deterministic form of the hypothesis (joint democracy is sufficient for peace) means that a single empirical disconfirmation would be very damaging to the theory. The last point can be generalized. The potential bearing of a deviant case analysis on a theory is significantly enhanced if the theory posits necessary or sufficient conditions or generates precise point predictions. A deviant case study design can also be combined with a most or least likely case design, which gives the case additional leverage over the theory.³²

Conclusions

I have argued that the rapidly expanding literature on case study methodology reflects an increasing theoretical orientation and methodological self-consciousness among case study researchers. They now generally see cases primarily as vehicles for constructing and supporting broader theoretical generalizations, and even most idiographic studies are guided by a well-developed theoretical framework. The role of theory is particularly evident in the criteria for case selection and logics of interpretation in most/least likely designs, deviant case strategies, and comparable-case designs.

Qualitative methodologists are increasingly catholic in their orientations. They generally emphasize that methodological debates are separable from theoretical debates and that case study methods are compatible with any theoretical orientation (George & Bennett, 2005: 4–9). They also argue that case study, formal, and quantitative methods are complementary, and exactly how these methods might be combined is now a leading area for research.³³

In my discussion of selection biases I noted that qualitative methodologists argue that process tracing, unlike large-N and cross-case comparative work, is not susceptible to the problem of selecting cases on the dependent variable, because process tracing follows a different logic of inference (Collier et al., 2004: 96; George & Bennett, 2005: 22–25; Bennett and Elman, 2006: 460–463). It resembles what philosophers of history call *genetic explanation* (Nagel, 1979: 564–568; Gallie, 1963),³⁴ which philosophers of history developed

³¹See Sambanis' (2004) discussion of the role of case studies in uncovering substantial unit heterogeneity in cases of civil wars.

³²An example is Ripsman and Levy's (2007) study of the absence of a preventive war under the seemingly optimal theoretical conditions of the 1930s.

³³See Laitin's (2002) elaboration of the “tripartite method,” Lieberman's (2005) discussion of “nested analysis,” and the symposiums in the *Qualitative Methods* Newsletter in 2006 (4,1) and 2007 (5,1).

³⁴Gerring (2007: chapter 7) offers a different view of process tracing.

as an alternative to Hempel's (1942) covering law (or nomological) model of historical explanation.³⁵

This raises the larger question regarding the existence of a "single logic of inference" in quantitative and qualitative analysis. This was a central theme in King et al. (1994: 3–7) but is contested by most qualitative methodologists, even those who define themselves as positivists. I think George and Bennett (2005: 11) are right that the single logic argument may apply to the level of epistemology but not to the level of methodology.³⁶ What quantitative and qualitative researchers shared is the logic of "deriving testable implications from alternative theories, testing these implications against quantitative or case study data, and modifying theories or our confidence in them in accordance with the results," which George and Bennett (2005: 11) describe as central to the "still-evolving positivist tradition." What quantitative and qualitative researchers do not share are specific methodological rules about case selection, the role of process tracing, and the relative emphasis on the role of causal mechanisms as the basis for explanation.

I think that this shared epistemological ground among quantitative, formal, and case study researchers is far greater than some of the methodological differences that divide them. I also think that some of the differences that do exist (see Mahoney & Goertz, 2006, for a superb discussion) have been exaggerated by some qualitative methodologists (Levy, 2007a). This suggests that the impediments to incorporating multiple methods into research programs are few, while the benefits are many.

Acknowledgments

I thank Andy Bennett, Colin Elman, and Gary Goertz for helpful comments on an earlier draft of this paper.

References

- Achen, C. H. 2005. Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science* 22: 327–339.
- Achen, C. H., and D. Snidal. 1989. Rational deterrence theory and comparative case studies. *World Politics* 41(2): 143–182.
- Allison, G. T. 1971. *Essence of decision*. New York: Little Brown.
- Bates, R., A. Greif, M. Levi, J.-L. Rosenthal, and B. Weingast. 1998. *Analytic narratives*. Princeton: Princeton University Press.
- Bennett, A., and C. Elman. 2006. Qualitative research: Recent developments in case study methods. *Annual Review of Political Science* 9: 455–476.
- Bennett, A., and C. Elman. 2007. Qualitative methods: The view from the subfields. *Comparative Political Studies* 40(2): 111–121.
- Brady, H. E., D. Collier, and J. Seawright. 2004. Refocusing the discussion of methodology. In Brady, H., and D. Collier, eds. *Rethinking social inquiry: Diverse tools, shared standards*, Lanham, MD: Rowman & Littlefield, pp. 3–21.
- Brady, H. E., and D. Collier, eds. 2004. *Rethinking social inquiry: Diverse tools, shared standards*. Lanham, MD: Rowman & Littlefield.

³⁵While the logic of inference in process tracing and genetic explanation differs in important respects from that in statistical analysis and a comparable cases strategy, our confidence in inferences from process tracing is greatest if each link in the causal chain is based on a well-established empirical regularity (probabilistic or otherwise) that has been confirmed by statistical or comparative case study analysis (Roberts, 1996).

³⁶Brady and Collier (2004) emphasize a similar theme, as reflected in the subtitle of their book: "Diverse Tools, Shared Standards."

- Brams, S. J. 1994. *Theory of moves*. New York: Cambridge University Press.
- Braumoeller, B. F. 2004. Hypothesis testing and multiplicative interaction terms. *International Organization* 58(4): 807–820.
- Braumoeller, B. F., and G. Goertz. 2002. Watching your posterior: Bayes, sampling assumptions, falsification, and necessary conditions. *Political Analysis* 10(2): 198–203.
- Bueno de Mesquita, B. 1981. *The war trap*. New Haven: Yale University Press.
- Bueno de Mesquita, B. 2000. Popes, kings, and endogenous institutions: The Concordat of Worms and the origins of sovereignty. *International Studies Review* 2(2): 93–118.
- Campbell, D. T. 1975. Degrees of freedom and the case study. *Comparative Political Studies* 8(2): 178–193.
- Campbell, D. T., and J. C. Stanley. 1966. *Experimental and quasi-experimental design for research*. Chicago: Rand McNally.
- Causal complexity and qualitative methods. 2006. Special issue. *Political Analysis* 14(3): 223–368.
- Clarke, K. 2002. The reverend and the ravens. *Political Analysis* 10(2): 194–197.
- Collier, D. 1999. Data, field work and extracting new ideas at close range. *APSA–CP Newsletter* Winter, 1–6.
- Collier, D., J. Mahoney, and J. Seawright. 2004. Claiming too much: Warnings about selection bias.” In Brady, H. E. and D. Collier, eds. *Rethinking social inquiry: Diverse tools, shared standards*. Lanham, MD.: Rowman & Littlefield, 85–102.
- Dion, D. 1998. Evidence and inference in the comparative case study. *Comparative Politics* 30(2): 127–146.
- Eckstein, H. 1975. Case studies and theory in political science. In Greenstein, F., and N. Polsby, eds. *Handbook of political science*, vol. 7, Reading, MA: Addison-Wesley, 79–138.
- Elman, C. 2005. Explanatory typologies in qualitative studies of international politics. *International Organization* 59(2): 293–326.
- Elman, C. 2008. Institutions for qualitative methods. In Box-Steffensmeier, J., H. Brady, and D. Collier, eds. *Oxford Handbook of Political Methods*. New York: Oxford University Press.
- Ethnography meets rational choice. 2006. *Qualitative Methods* (newsletter) 4(1): 1–33.
- Fortna, V. P. 2004. *Peace time: Cease-fire agreements and the durability of peace*. Princeton: Princeton University Press.
- Freundreich, J. 1983. Explanation of variation and detection of covariation: The purpose and logic of comparative analysis. *Comparative Political Studies* 16(2): 255–272.
- Gallie, W.B. 1963. The historical understanding. *History and Theory* 3(2): 149–202.
- George, A. L. 1979. Case studies and theory development. In Lauren, P., ed. *Diplomacy: New approaches in theory, history, and policy*, New York: Free Press, 43–68.
- George, A. L., and A. Bennett. 2005. *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press.
- George, A. L., and R. Smoke. 1974. *Deterrence in American foreign policy*. New York: Columbia University Press.
- Gerring, J. 2007. *Case study research*. New York: Cambridge University Press.
- Gochal, J. R., and J. S. Levy. 2004. Crisis mismanagement or conflict of interests? A case study of the Crimean war. In Maoz, Z., ed. *Multiple paths to knowledge in international relations*, Lexington, MA: Lexington Books, pp. 309–342.
- Goertz, G., and J. S. Levy. 2007. *Explaining war and peace: Case studies and necessary condition counterfactuals*. New York: Routledge.
- Goertz, G., and J. Mahoney. 2005. Two-level theories and fuzzy-set analysis. *Sociological Methods & Research* 33(4): 497–538.
- Ho, D. E., K. Imai, G. King, and E. A. Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15(3): 199–236.
- Hempel, C. G. 1942. The function of general laws in history. *Journal of Philosophy* 39: 35–48.
- Hobsbawm, E. 1997. *On history*. New York: The New Press.
- Holsti, O. R. 1965. The 1914 case. *American Political Science Review* 59(2): 365–378.
- King, G., R. Keohane, and S. Verba. 1994. *Designing social inquiry*. Princeton University Press.
- Laitin, D. 2002. Comparative politics: The state of the sub-discipline. In Katznelson, I., and H. V. Milner, eds. *Political science: The state of the discipline*, New York: Norton, 630–659.

- Lakatos, I. 1970. Falsification and the methodology of scientific research programmes. In Lakatos, I., and A. Musgrave, eds. *Criticism and the growth of knowledge*, New York: Cambridge University Press, 91–196.
- Lake, D. A., and R. Powell. 1999. International relations: A strategic choice approach. In Lake, D. A. and R. Powell, eds. *Strategic choice and international relations*, Princeton: Princeton University Press, 3–38.
- Larson, D. W. (2001) Sources and methods in Cold War history: The need for a new theory-based archival approach. In Elman, C., and M. F. Elman, eds. *Bridges and boundaries*, Cambridge: MIT Press, 327–350.
- Levy, J. S. 2001. Explaining events and testing theories: History, political science, and the analysis of international relations. In Elman, C., and M. F. Elman, eds. *Bridges and boundaries*, Cambridge: MIT Press, 39–83.
- Levy, J. S. 2002. Qualitative methods in international relations. In Brecher, M., and F. P. Harvey, eds. *Millennial reflections on international studies*, Ann Arbor: University of Michigan Press, pp. 432–454.
- Levy, J. S. 2007a. Qualitative methods and cross-method dialogue in political science. *Comparative Political Studies* 40(2): 196–214.
- Levy, J. S. 2007b. Theory, evidence, and politics in the evolution of research programs. In Lebow, R. N., and M. Lichbach, eds. *Theory and evidence in comparative politics and international relations*. New York: Palgrave Macmillan, 177–197.
- Levy, J. S., and W. R. Thompson. 2005. Hegemonic threats and great power balancing in Europe, 1495–2000. *Security Studies* 14(1): 1–30.
- Lieberman, E. S. 2005. Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review* 99(3): 435–452.
- Lieberson, S. 1992. Small N's and big conclusions. In Ragin, C., and H. Becker, eds. *What Is a Case?* New York: Cambridge University Press, 105–118.
- Lijphart, A. 1971. Comparative politics and the comparative method. *American Political Science Review* 65(3): 682–693.
- Lijphart, A. 1975. The comparable cases strategy in comparative research. *Comparative Political Studies* 8(2): 133–177.
- Lustick, I. 1996. History, historiography, and political science: Multiple historical records and the problem of selection bias. *American Political Science Review* 90(3): 605–618.
- McKeown, T. 1999. Case studies and the statistical world view. *International Organization* 53(1): 161–190.
- Mackie, J. L. 1974. *The cement of the universe: A study of causation*. Oxford: Clarendon.
- Mahoney, J., and G. Goertz. 2004. The possibility principle: Choosing negative cases in comparative research. *American Political Science Review* 98(4): 653–670.
- Mahoney, J., and G. Goertz. 2006. A tale of two cultures: Contrasting quantitative and qualitative research. *Political Analysis* 14(3): 227–249.
- Maoz, Z. 2002. Case study methodology in international studies: From storytelling to hypothesis testing. In Brecher, M. and F. P. Harvey, eds. *Millennial reflections on international studies*, Ann Arbor, MI: University of Michigan Press, 455–475.
- Mill, J. S. 1970 [1885]. *A system of logic*. London: Longman.
- Multi-method work, dispatches from the front lines. 2007. *Qualitative Methods* 5(1): 9–28.
- Nagel, E. 1979. *The structure of science*. Indianapolis: Hackett.
- North, R. C. 1967. Perception and action in the 1914 crisis. *Journal of International Affairs* 21:103–122.
- Picard, R. R., and R. D. Cook. 1984. *Journal of the American Statistical Association* 79(387): 575–583.
- Pierson, P. 2000. Increasing returns, path dependence, and the study of politics. *American Political Science Review* 94(2): 251–267.
- Popper, K. 1965. *The logic of scientific discovery*. New York: Harper Torchbacks.
- Przeworski, A., and H. Teune. 1970. *The logic of comparative social inquiry*. New York: Wiley.
- Ragin, C. C. 1987. *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.

- Ragin, C. C. 2000. *Fuzzy-set social science*. Chicago: University of Chicago Press.
- Ragin, C. C., and H. Becker. 1992. *What is a Case?* New York: Cambridge University Press.
- Ray, J. L. 1995. *Democracy and international conflict: An evaluation of the democratic peace proposition*. Columbia: University of South Carolina Press.
- Ray, J. L. 2005. Constructing multivariate analyses (of dangerous dyads). *Conflict Management and Peace Science* 22: 277–292.
- Reiter, D., and A. C. Stam. 2002. *Democracies at war*. Princeton: Princeton University Press.
- Ripsman, N., and J. S. Levy. 2007. The preventive war that never happened: Britain, France, and the rise of Germany in the 1930s. *Security Studies* 16(1): 32–67.
- Roberts, C. 1996. *The logic of historical explanation*. University Park, PA: Pennsylvania State University Press.
- Sambanis, N. 2004. Using case studies to expand economic models of civil wars. *Perspectives on Politics* 2(2): 259–279.
- Skocpol, T. 1979. *States and social revolutions*. Cambridge: Cambridge University Press.
- Skocpol, T. 1984. Emerging agendas and recurrent strategies in historical sociology. In Skocpol, T., ed. *Vision and method in historical sociology*, New York: Cambridge University Press, 156–173.
- Seawright, J. 2002. Testing for necessary and/or sufficient causation: Which cases are relevant? *Political Analysis* 10(2): 178–193.
- Stretton, H. 1969. *The political sciences: General principles of selection in social science and history*. London: Routledge & Kegan Paul.
- Tarrow, S. 1995. Bridging the quantitative-qualitative divide in political science. *American Political Science Review* 89(2): 471–474.
- Van Evera, S. 1997. *Guide to methods for students of political science*. Ithaca, New York: Cornell University Press.
- Verba, S. 1967. Some dilemmas in comparative research. *World Politics* 20(1): 111–127.
- Walter, B. F. 2002. *Committing to peace: The successful settlement of civil wars*. Princeton: Princeton University Press.