QUALITATIVE METHODS IN INTERNATIONAL RELATIONS

Jack S. Levy

Three decades ago, Sartori complained about the number of "unconscious thinkers" in the field of comparative politics, and the same could be said of the study of international relations.[1] Most qualitative analyses were idiographic rather than nomothetic,[2] historically specific rather than theoretically driven, and too little concerned with the logic of inference and questions of generalizability. Scholars gave little attention to the problem of how to control for extraneous variables in situations in which the number of variables typically exceeded the number of cases, or to the question of whether there are alternative methods for validating causal inferences in a single case. Both critics and advocates of case study methods agreed that qualitative research in the 1950s and 1960s precluded the cumulation of knowledge across historical cases.[3] This lack of scientific rigor created an image of qualitative analysis as highly subjective, pliable in fitting facts to theoretical arguments, nonreplicable, and essentially nonfalsifiable.

Much has changed in the last two or three decades, both in terms of the growing body of literature on the methodology of qualitative analysis and the increasing social science orientation of most qualitative research in international relations. My aim in this essay is to survey the expanding literature on qualitative methods and to highlight some of its major themes. Qualitative analysis includes everything from interpretive ethnographic studies to macrohistories spanning millennia to microanalyses of particular events, and individual cases have been analyzed by quantitative as well as qualitative methods. Thus we should not equate qualitative methods with the case study method. The core of the literature on the methodology of qualitative research in the international relations field focuses on comparative and case study methods from a positivistic perspective, however, and that is the focus of this essay. I give particular attention to the meaning of the case concept, the different kinds and purposes of case studies, the different types of case study designs for testing theories, and the distinctive contributions and limitations of comparative and case study methods.

A well-defined body of literature on comparative and case study methods was beginning to emerge just as Sartori lamented unconscious thinking in the field.[4] These efforts to develop a methodology of comparative analysis were informed by comparable work in sociology,[5] and were stimulated by the success of historical sociology.[6] By the early 1990s, the methodology of comparative and case study analysis had a well-defined place in the fields of comparative and international politics.[7]

The publication of King, Keohane, and Verba's *Designing Social Inquiry* triggered another surge of interest in the methodology of qualitative research.[8] This was in part a response to the "KKV" argument that there is a single logic of scientific inference, that qualitative analysis is not fundamentally different from large-N statistical analysis, and that the former needs to conform more closely to the latter. These arguments led to a major debate in the discipline. Symposia in the *American Political Science Review* (1995) and in the *APSA-CP Newsletter* (1998) and subsequent articles and books dealt with the "single logic" question, the advantages and limitations of case study approaches, the different types of case study approaches, and specific problems of case selection and other aspects of research design.[9] One indicator of the expanding interest in qualitative methodology is the increasing number of graduate courses offered in this area. Another is the organization of the Inter-University Consortium on Qualitative Research Methods (CQRM), which includes instructional workshops, collection of syllabi, and on-line discussion groups.[10]

Advances in the methodology of qualitative research have been paralleled by an increasing methodological self-consciousness and sophistication in applications of qualitative methods to empirical research. The current generation of qualitative international relations researchers are much more nomothetic in their general orientations, both in terms of their interest in using explicitly stated theories to guide historical interpretations and in their greater interest in developing theoretical generalizations valid across time and space. They are more attentive to the question of proper case selection and other aspects of research design, more sensitive to potential threats to valid inference, and generally more social scientific in orientation.[11]

Definitions

Although some define the case study method as distinct from the comparative method,[12] nearly all case studies involve comparisons, whether they be explicit or implicit, across cases or within cases. In addition, the literature has evolved in such a way that the comparative method has come to be associated with the analysis of a small number of cases,[13] further eroding any distinction between the case study and comparative methods. Some use the term *case study* to refer to a single case study and *comparative method* to refer to comparisons among a small number of cases, but the similar logic

underlying both single and multiple case studies leads most scholars to use *case study method* to refer to both.[14]

Despite the pervasiveness of case study analysis in international and comparative politics, scholars are far from agreement on how to define either case or case study. Ragin and Becker have edited a book entitled *What Is a Case?* but provide no clear definition or answer.[15] Precisely how one defines "case" depends in part on the purposes of inquiry. Those who are primarily interested in understanding and interpreting a particular historical episode, with little interest in constructing or testing more general theoretical propositions, tend to define case in terms of a set of events bounded in time and space. For such scholars, World War I or the cold war may each be a "case" to be explained.

Given the shift toward a more theoretical orientation in the international relations field and an increasing interest in constructing and validating theoretical propositions, the vast majority of scholars conceive of a case as an *instance* of something else, of a theoretically defined class of events.[16] The question is always, What is this a case of? In this sense World War I itself is not a case, but certain theoretically defined aspects of World War I may be cases of some broader phenomenon, such as deterrence, balance of power, power transition, diversionary action, or war termination. This is explicit in George's conception of the method of "structured, focused comparison," which focuses on a particular analytically defined aspect of a set of events and uses a well-defined set of theoretical questions to structure empirical inquiry. This analytical orientation recognizes that observation is theory laden, and that researchers' theoretical preconceptions will lead them to decide which of the myriad of events constituting a historical episode are to be selected out, studied, and recast in terms of "variables."[17]

From this perspective, "cases come wrapped in theories" and do not exist independently of the analytic framework a scholar brings to a particular subject.[18] Cases are analytical constructions. They are made, not found; invented, not discovered.[19] This does not imply that there are multiple, equally valid answers to our questions about the world, only multiple questions that we might ask.

It is important to emphasize that a case generally includes many observations on the same variable. The July 1914 crisis, for example, includes a substantial number of observations of the use of coercive threats, domestic influences, and misperceptions, to name but a few variables. Indeed, one of the main strategies of case study analysis, just like large-N analysis, is to generate as many testable implications of one's hypotheses as possible in a given case.[20] This standard view of multiple observations per variable per case is more useful than Eckstein's definition of a case as "a phenomenon for which we report and interpret only a single measure on any pertinent variable,"[21] and most methodologists have moved away from Eckstein's definition.

Types of Case Studies

There are many types of case studies and numerous ways of classifying them. Most classifications focus on the purpose or function of case studies

and build on Lijphart's typology of atheoretical, interpretive, hypothesis-generating, theory-confirming, theory-infirming, and deviant case studies.[22] Eckstein suggests a similar typology: configurative-idiographic, disciplined-configurative, heuristic, and crucial case studies based on most likely and least likely designs.[23] Eckstein suggests the additional category of plausibility probe. These are ideal types, and in practice many case studies combine several of these aims. The following classification represents a combination of the Lijphart and Eckstein categories.

Atheoretical or *configurative-idiographic* case studies are traditional single-case analyses often associated with area studies. They are highly descriptive and aim to understand and interpret a single case as an end in itself rather than to develop broader theoretical generalizations. Idiographic case studies are inductive; they involve a minimum of a priori theoretical preconceptions, and the interpretation emerges from the case itself. The analyst generally attempts to create a gestalt or holistic picture of a historical episode, explaining all aspects of the episode and all of their interconnections. This is "total history," which "cannot decide to leave out *any* aspect of human history a priori."[24]

Interpretive or *disciplined-configurative* case studies also aim to explain/interpret a single case, but that interpretation is explicitly structured by a theory or well-developed theoretical framework that focuses attention on some theoretically specified aspects of reality and neglects others. This is analytic history rather than total history, but it is still idiographic in that it aims to explain a particular historical episode rather than develop or test theoretical generalizations.

Although such "case-explaining" case studies are common in the international relations field, they are less highly valued than work that generates or tests theories,[25] and many international relations theorists believe that historically specific research is best left to historians. This is understandable but unfortunate, because the explicit and structured use of theory to explain discrete cases often provides better explanations and understandings of those cases—or at least some 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 validate or invalidate empirically. Theoretically guided, interpretive case studies can significantly enhance our descriptive understanding of the world, and international relations scholars have much to contribute to this task.[26]

Whereas atheoretical and interpretive case studies are basically idiographic in their explanatory objectives, *hypothesis-generating* or *heuristic* case studies are more nomothetic in their aims. They examine a particular case or perhaps several cases for the purpose of developing more general theoretical propositions, which can then be tested through other methods, including large-N methods and perhaps alternative case study methods. Case selection is driven by theoretical considerations, not by the intrinsic interest or historical importance of the case itself, and only particular aspects of the case are investigated.

It would be more correct to say that case studies can contribute to the *process* of theory construction than to theory itself, for the latter—defined as a logically interconnected set of propositions about empirical phenomena—requires a more deductive orientation than case studies provide. Thus Achen and Snidal argue, "the logic of comparative case studies inherently provides too little logical constraint to generate dependable theory," and they complain that findings of case studies have "too often . . . been interpreted as bodies of theory."[27] The same can be said of statistical or experimental methods. The application of each of these methods can stimulate the imagination and suggest new hypotheses, but isolated hypotheses fall short of theory, and neither case study methods nor statistical methods can by themselves generate theory.

Case studies permit an intensive examination of particular historical sequences, and in doing so they can contribute to the process of theory development by helping to clarify the meaning of key variables and the validity of empirical indicators used to measure them, and by suggesting additional causal mechanisms, causal variables, and interaction effects. They can also help to identify the contextual variables that affect hypothesized causal processes and to identify the scope conditions under which particular theories are valid. These are all important steps in the theory-building process. Thus Achen and Snidal argue that "because they are simultaneously sensitive to data and theory, case studies are more useful for these purposes [of developing analytic theory] than any other methodological tool."[28]

The role of case studies in generating hypotheses, or at least in refining and sharpening them, is enhanced by the close interaction of theory and data in case study analysis. The analyst begins with a theory, uses it to interpret a case, and simultaneously utilizes the case to suggest important refinements in the theory, which can then be tested on other cases or perhaps even on other aspects of the same case.[29] George and Smoke's analysis of deterrence in American foreign policy, for example, is organized sequentially in terms of theory specification, application of the theory to historical cases, and reformulation of the theory based on the cases.[30] The interplay between theory and evidence is also explicit in "analytic narratives" in which formal rational choice theory guides analytic histories, the anomalies of which are then used to refine the theory.[31]

Although hypothesis-generating case studies can sometimes contribute to the development of entirely new hypotheses, through unexpected discoveries in the process of investigating other phenomena, case studies are often more useful in helping the researcher to refine existing hypotheses, as in George's method of structured, focused comparison.[32] The greater the theoretical structure guiding inquiry, and the better defined the researcher's hypotheses, the more efficient the hypothesis-generating process. But sometimes a full-fledged test of a hypothesis is premature. *Plausibility probes*, like pilot studies in experimental or survey research, are intermediate steps between hypothesis construction and hypothesis testing. They enable the researcher to refine the hypothesis or theory before engaging in a costly and

time-consuming research effort, whether through the massive collection of quantitative data or through extensive fieldwork.

Deviant case studies, which focus on empirical anomalies in established theoretical generalizations in order to explain them and refine existing hypotheses, are a particularly powerful strategy for theory refinement, much like examining the residuals in a statistical analysis. Deviant case studies can help to validate the measurement of key variables; identify omitted variables, interaction effects, or alternative causal paths; or identify the scope conditions under which a particular theory is valid. Revised hypotheses can then be tested over a broader set of cases or unexplored aspects of the same case. The analysis of "borderline" cases in the hypothesized absence of war between democracies is a form of deviant case analysis.[33]

Another way case studies can contribute to theory development is through what Lazarsfeld described as the "analysis of the dependent variable."[34] The analyst examines cases that on the surface appear to be similar on the dependent variable, in order to identify its different subtypes. This facilitates the development of a more differentiated conception of the dependent variable and a more nuanced set of hypotheses. George and subsequently George and Bennett develop this idea in their discussion of "typological theory."[35] They emphasize that there may be a number of alternative causal paths to a given outcome (equifinality) and attempt to identify these alternative causal paths and, if possible, to specify the conditions under which each is most likely to occur. An example is George and Smoke's exploration of the various ways in which deterrence can fail and their elaboration of the different causal paths associated with each. Similar logic underlies Collier and Levitsky's effort to articulate various subtypes of democracy.[36]

Many scholars emphasize this role of comparative and case study methods in contributing to the process of theory building, in stimulating the imagination and generating hypotheses that can then be more rigorously tested. Stinchcombe suggests that "lots of facts" can be "good hard stones for honing ideas."[37] Lijphart argues that a comparative perspective (as distinct from the comparative method per se) can be a helpful element in discovery.[38] He quotes Stretton, who argues that "the function of comparison is less to stimulate experiment than to stimulate imagination. . . . Comparison is strongest as a choosing and provoking, not a proving, devise; a system for questioning, not for answering."[39]

In addition to playing an essential role in the explanation of individual historical episodes and a contributory role in the generation of hypotheses, case studies can also be used to test hypotheses and theories. This is Lijphart's *theory-confirming* and *theory-infirming* roles of case studies, which are best collapsed into a single theory-testing category.[40] As Lijphart recognizes, however, the use of case studies for testing hypotheses faces a very serious problem: the existence of many variables in conjunction with a relatively small number of cases. As a consequence, outcomes are overdetermined, and it is difficult if not impossible to be certain that changes in

the dependent variable are due to changes in the hypothesized independent variables and not to the effects of extraneous variables. Much of the literature on the comparative method and on case study analysis deals with the problem of how to make causal inferences in small-N research, when the number of variables generally exceeds the number of cases.[41]

Lijphart recognized the many variables/small-N problem and offered a number of possible solutions.[42] Three try to directly influence the ratio of cases to variables: increase the number of cases by expanding the domain of the analysis both spatially and temporally; reduce the number of variables, either by collapsing conceptual categories or by data reduction techniques such as factor analysis; or focus on a relatively small number of key variables by constructing more parsimonious theories. Lijphart's fourth solution focused on "comparable cases"—cases similar in terms of the control variables but different in terms of hypothesized explanatory variables that one wants to investigate.

In emphasizing an increase in the ratio of cases to variables, Lijphart basically accepted the utility of large-N analysis and conceded that the comparative method was inferior to the experimental or statistical methods for the purposes of causal inference. He concluded that comparative analysis was a "first stage" devoted to the careful formulation of hypotheses, which are then tested in a "second stage" statistical analysis.[43] He also acknowledged a certain tension between the goals of increasing the number of cases and focusing on similar cases, because the latter narrows the range of possible cases.

In a subsequent article on the comparative method, Lijphart conceded that the maximizing-N strategy—which is based on controlling for extraneous variables through partial correlations—and the comparable-cases strategy—which is based on control through a carefully selected set of matched cases—involved different logics and were more fundamentally opposed than he had initially acknowledged.[44] He retracted his earlier argument that the comparative method works best with a larger number of cases, and defined the comparative method as equivalent to the small-N analysis of comparable cases. In this way Lijphart moved closer to Eckstein who stressed the advantages of small-N analysis, including single-case analysis.[45]

Following Lijphart, scholars now conceive of the comparative method and case study methods as strategies for dealing with a relatively small number of cases. The comparative method is often defined as a strategy for conducting research on naturally occurring phenomena in a way that aims to control for potential confounding variables through careful case selection and matching rather than through experimental manipulation or partial correlations.[46]

Varieties of Case Selection Strategies

The comparable-cases strategy is closely related to John Stuart Mill's *method of difference*, which focuses on cases that have different values on the dependent variable and similar values on all but one of the independent variables.[47] In terms of the logic of inference, this facilitates the identifica-

tion of causal factors that vary with the dependent variable by eliminating all variables that are constant over the similar cases. Mill's *method of agreement* focuses on cases that are similar on the dependent variable and different on all but one of the independent variables, in order to eliminate all factors that vary across cases on the independent variable and that therefore cannot account for similar outcomes across cases on the dependent variable.

Mill's methods of agreement and difference are comparable to "most-different" and "most-similar" systems designs, respectively.[48] A *most-different systems* design identifies cases that are different on a wide range of explanatory variables but not on the dependent variable, while a *most-similar systems* design identifies cases that are similar on a wide range of explanatory variables but different on the value of the dependent variable.[49] The former eliminates extraneous causal variables that vary across cases, while the latter eliminates causal variables that do not vary across cases.

Mill argued that the method of difference was more powerful than the method of agreement in establishing causation. Similarly, Lijphart and Smelser each preferred most-similar systems designs.[50] Przeworski and Teune, on the other hand, preferred most-different systems designs, which maximize the number of extraneous variables that can be eliminated because they vary while the dependent variable does not.[51] The basic 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.

It is rare that either strategy alone can fully eliminate extraneous variables, and the best strategy generally involves the combination of most-similar and most-different systems designs. Mill recognized this and argued for a method of "concomitant variation" involving a combination of the methods of agreement and difference.

One problem in the application of Mill's methods and of most-similar and most-different systems designs is the difficulty of identifying cases that are truly comparable—identical or different in all respects but one. In addition, because of the possibility that several different sets of conditions may lead to the same outcome—which Mill identified as the "plurality of causes" and which modern systems theorists refer to as "equifinality"— 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. Mill acknowledged this and concluded that for this reason the application of the methods of agreement and difference to political science was "completely out of the question."[52]

Proponents of comparative case methods argue that Mill was far too cautious, and that insisting on precise and absolute comparability imposes "a too exacting scientific standard."[53] Comparative researchers emphasize that experimental and statistical methods themselves are imperfect, and they focus on the question of how best to overcome the acknowledged limitations of the comparative method. They give particular emphasis to strategies of case selection and to process tracing to supplement basic controlled comparisons. I return to process tracing in the next section and focus here on case selection strategies.

Within-case comparisons of hypothesized relationships at different points in time within the same case are particularly powerful. Such longitudinal comparisons generally take the form of most-similar systems designs because they are able to hold so many variables constant: political history, culture, institutions, geography, and other variables that change only slowly (if at all) over time. This facilitates the identification of the small number of variables that vary with the dependent variable of interest. George, and George and Bennett, use the label of the "congruence method" for this kind of within-case comparison within the methodology of structured, focused comparison.[54] Rosen's analysis of Indian strategic doctrine over time and how it changed as a function of changing strategic culture within a relatively static geopolitical context is a good example of a longitudinal most-similar systems design.[55]

Another useful case study design, which facilitates the control over additional variables, involves a combination of across-case and within-case comparisons. A good example is Snyder's study of imperial overextension, which combines comparisons of the behaviors of different states, different individuals within the same state, and the same individuals within a given state over time.[56]

While comparative researchers argue about the relative merits of alternative case selection strategies, one thing they agree on is that the strategy of random selection of cases, so useful in large-N statistical analysis, will often generate serious biases in small-N research.[57] Scholars generally argue that the analysis of a limited number of cases is better served by a careful selection of nonrandom cases.

One of the most serious dangers in the deliberate selection of nonrandom cases involves overrepresenting 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, because it results in a reduction in the slope estimates generated by regression analyses (assuming linear relationships) and thus underestimates the strength of causal effects.[58] Selecting cases with very high (or very low) values of the dependent variable are common as well as consequential. Precisely because of their historical significance and high salience, major wars and major revolutions are the kinds of cases that most attract scholarly attention.

The selection of cases with extreme values of the independent variable does not have a comparable effect, and this asymmetry is the basis for warnings of the dangers of "selecting on the dependent variable."[59] These problems apply to small-N as well as large-N research.[60] Research based on the selection of cases with no variation at all on the dependent variable ("no-variance" designs) are particularly problematic. In the study of the causes of war, for example, if the analyst were to examine only wars and observed a particular factor present in every case, she could not infer that this factor systematically contributes to the outbreak of war because there might be countless other cases not observed in which the same factor were present but in which war did not occur.

Although warnings to avoid selecting on the dependent variable have been useful in reminding students of wars, revolutions, and other phenomena to include cases in which wars or revolutions did not occur—to think about the "dogs that didn't bark"—the mechanical application of this basic rule obscures some important situations in which selection of observations on the dependent variable might be a useful strategy for research. One involves the strategy of studying deviant cases for the purpose of analyzing why they deviate from theoretical predictions. Another is situations in which the hypothesis posits necessary conditions for the occurrence of a particular outcome.

For testing hypotheses that posit necessary conditions, the basic logic of inference requires the selection of cases on a particular value of the dependent variable because the only observations that can falsify the hypothesis in question are those in which a particular outcome of the dependent variable occurs in the absence of a condition that is hypothesized to be necessary for that outcome.[61] If the hypothesis is deterministic in its assertion of necessary conditions, and if one is confident that no measurement error is present, the observation of a single case in which the posited condition is absent is in principle sufficient to falsify the hypothesis. If we allow for some measurement error, even a small number of anomalous cases in which the hypothesized causal factor is not present can significantly undermine our confidence in a hypothesis based on necessary conditions.[62]

Similarly, causal propositions positing sufficient conditions for a given outcome can usefully be tested through a case study research design, though here it is essential to select cases on a particular range of values of the independent variable posited to be sufficient for a given outcome.[63] If the prediction is strong and if measurement error is negligible, hypotheses positing sufficient conditions can be seriously undermined by identifying a very small number of cases in which the hypothesized sufficient condition is not followed by the predicted outcome.

The analysis gets more complicated if there is more than one necessary or sufficient condition, or if there are multiple causal paths that can lead to the same outcome. A particular condition might be necessary for one sequence to operate, and that sequence may be sufficient for a particular outcome to occur, but there may be other sequences that also lead to the same outcome but that do not involve the key condition in the first sequence. The impact of some variables may be contingent on the values of other variables, so that simple additive models will not work, and the analyst must examine the combinations or interaction effects of different sets of factors. Ragin refers to this general problem as "multiple conjunctural causation."[64] He argues that standard statistical methods cannot easily deal with this phenomenon,[65] and develops "qualitative comparative analysis" based on Boolean algebra to identify and test combinatorial hypotheses.[66]

Another strategy for case selection for the purposes of testing theories involves what Eckstein called *crucial case studies*, which are related to the concepts of *most-likely* or *least-likely* case research designs.[67] A most-likely

case is one that almost certainly must be true if the theory is true, in the sense that all the assumptions of a theory are satisfied and all the conditions hypothesized to contribute to a particular outcome are present, so the theory makes very strong predictions regarding outcomes in that case. If a detailed analysis of a most-likely case demonstrates that the theory's predictions are not satisfied, then our confidence in the theory is seriously undermined. The logic of inference is Bayesian, in the sense that the marginal impact of the data on our confidence in the validity of a hypothesis depends on our a priori judgments of the validity of the hypothesis.[68] The greater the a priori likelihood of the hypothesis, the lower the impact of confirmatory data and the greater the impact of disconfirmatory data.

Similar logic applies to a least-likely case design, which selects "hard" cases in which the predictions of a theory are quite unlikely to be satisfied because few of its facilitating conditions are satisfied. If those predictions are nevertheless found to be valid, our confidence in the theory is increased, and we have good reasons to believe that the theory will hold in other situations that are even more favorable for the theory. Least-likely case research designs follow what I call the "Sinatra inference"—if I can make it there I can make it anywhere. Most-likely case designs follow the inverse Sinatra inference—if I cannot make it there I cannot make it anywhere.

Most-likely and least-likely case designs are often based on a strategy of selecting cases with extreme values on the independent variables, which should produce extreme outcomes on the dependent variable, at least for hypotheses positing monotonically increasing or decreasing functional relationships. Alternatively, a most-likely case design can involve selecting cases where the scope conditions for a theory are fully satisfied, while a least-likely case design identifies cases in which the theory's scope conditions are satisfied weakly if at all.

A good example of a most-likely case design is Lijphart's study of political cleavages and stability in the Netherlands.[69] Pluralist theory, which was widely accepted in the discipline, argued that cleavages that cut across various social groups promoted social peace and political stability, while cleavages that were mutually reinforcing across various social groups contributed to social conflict and political instability.[70] Because there were very few cross-cutting cleavages in the Netherlands, pluralist theory predicted high levels of social conflict and low levels of political stability. By demonstrating that the opposite was true, Lijphart's analysis contradicted the unconditional statement of the theory and went a long way toward refuting it. This is a good example of the power of a well-selected individual case study to seriously undercut a widely accepted theory.[71]

The power of most-likely and least-likely case analysis is further strengthened by defining most likely and least likely not only in terms of the predictions of a particular theory but also in terms of the predictions of leading alternative theories. This builds on the idea that a theory is falsified not by the data alone but by a "three-cornered test" involving the theory, the data, and a rival theory.[72] The strongest support for a theory comes when a case is least likely for a particular theory and most likely for the rival

theory, and when observations are consistent with the predictions of the theory but not those of its competitor.

A good example here is Allison's application of three models of foreign policy decision making to the Cuban missile crisis.[73] Allison argued that the missile crisis was a least-likely case for the organizational and bureaucratic models of decision making and a most-likely case for the rational-unitary model. We might expect organizational routines and bureaucratic politics to affect decision making on budgetary issues and on issues of low politics, but not in cases involving the most severe threats to national security, where rational calculations to maximize the national interest should dominate and where politics should stop "at the water's edge." If Allison could show that bureaucratic and organizational factors had a significant impact on key decisions in the Cuban missile crisis, we would have good reasons to expect that these factors would be important in a wide range of other situations.[74]

Process Tracing

The preceding discussion of case selection strategies suggests that certain kinds of individual case studies can contribute to hypothesis testing as well as to hypothesis construction. Most between-case and within-case comparisons are correlational in nature and examine whether a particular set of conditions is associated with hypothesized outcomes, while holding constant as many other factors as possible. Thus George, and George and Bennett, refer to within-case comparisons of hypothesized relationships at different points in time within the same case as the "congruence method" and include it within the methodology of structured, focused comparison.[75]

There is another approach to within-case analysis, one that is quite common in the practice of case study research but that is often neglected in attempts to formally describe case study methodology, and that is *process tracing*.[76] Process tracing follows a different logic and tries to uncover the intervening causal mechanisms between conditions and outcomes through an intensive analysis of the evolution of a sequence of events within a case. The logic of inference is much more similar to what philosophers of history call *genetic explanation*[77] than to explanations based on covering laws and deductive nomological logic.[78]

Process tracing provides several comparative advantages for testing many kinds of intervening causal mechanisms, particularly those involving propositions about what goes on inside the "black box" of decision making and about the perceptions of actors.[79] One of the implications of the democratic peace proposition, for example, is that democracies are perceived differently than autocracies and that these differences have a significant impact on behavior. These perceptions are often better explored through small-*N* case study methods than through large-*N* statistical methods.

Case study process-tracing methods can also be extremely useful in the empirical analysis of nonlinear propositions involving critical inflection points. The testing of such propositions is extremely sensitive to the accurate identification of these inflection points. In order to avoid circular infer-

ences, this must be done with indicators measured independently of the behavior predicted by the theory. It may be difficult to identify empirical indicators of these inflection points that are valid across a large number of cases for the purposes of a large-N analysis, and case study methods can be used to help identify these tipping points and why they occur.[80] Similarly, process tracing can be extremely useful in the exploration of path-dependent macrohistorical processes that are extremely sensitive to patterns of timing and sequence.[81]

Many proponents of case study analysis argue that process tracing has a comparative advantage over large-N statistical methods in validating intervening causal mechanisms, because statistical methods are limited to establishing correlations while case studies can trace the steps in a causal chain.[82] While there may be some truth to this argument—particularly for propositions that involve equifinality, complex and contingent interaction effects, or path dependencies—this argument goes too far and needs to be qualified. There is a tendency among some case study researchers both to underestimate the possible utility of statistical analysis for empirically differentiating among hypothesized causal mechanisms[83] and to exaggerate the utility of process tracing for this purpose.[84]

Those engaged in a close process tracing of a causal chain still face the problem identified by Hume, the impossibility of establishing causality from empirical observation. We cannot know for certain that a particular outcome y is the causal result of a set of factors x rather than another set of factors z that have been omitted from the formal analysis. Our confidence in such an inference is greatest, however, if each link in the causal chain is based on a well-established empirical regularity (probabilistic or otherwise) that has been confirmed by large-N studies or possibly comparative case studies in other comparable empirical domains.[85] It is true that we have few strong regularities in international relations, and that this limits our ability to infer causality based on covering laws. But this is a general problem of any approach that attempts to infer causality from a sequence of empirical observations, not just a problem with the covering law model.

It is more useful to think of causality as an analytical construct, a component of our theories rather than something that can be inferred directly from empirical observation.[86] All theories about the empirical world have testable implications.[87] Many of these implications concern the relative frequencies or magnitudes of readily observable events and can be best validated by large-N statistical studies. Other implications deal with hypotheses that posit necessary or sufficient conditions, that fall within the rather opaque black box of decision making, or that are for various reasons difficult to measure with validity and accuracy over large numbers of cases in different historical and cultural contexts. These implications can often be effectively analyzed through process tracing. The greater the empirical validation of the testable implications of a theory, by whatever method, the more confidence we can have in a theory, and hence in the causal mechanisms posited by the theory.

Limitations of Case Study Methods

I have argued that case studies are essential for the description, explanation, and understanding of particular historical episodes, and that they can also be useful in the development and refinement of more general theoretical propositions. Case studies can demonstrate that certain constellations of variables generate predictions with nonempty cells, even if they cannot establish the relative frequency with which the predicted event occurs. Case studies can also play a role in theory testing, particularly if the theory is very strong and makes point predictions or posits necessary or sufficient conditions. Few of our theories of international relations satisfy these criteria, however, and for the purposes of testing most theories case study methods have a number of rather serious limitations.[88]

One is the large number of variables relative to the small number of cases, which is probably the central issue in the literature on the comparative method and case studies over the past three decades. In attempting to demonstrate that her hypothesized causal variables, and not other variables, explain various outcomes, the case study researcher can achieve some degree of control through careful case selection based on most-similar or most-different systems designs, or preferably a combination of both. The researcher can gain additional leverage over her theory through most-likely and least-likely designs, defined both in terms of the theory and the leading rival theories. If a theory's scope conditions are fully satisfied, if its testable implications are precise, if measurement is valid and accurate, and if cases are carefully selected, the case study researcher can often make a plausible argument that the theory is either supported or disconfirmed for the cases under investigation.[89] It is much less likely, however, that she will be able to convincingly demonstrate that her findings are valid for comparable instances of the same phenomenon beyond her immediate study. The case study researcher gains leverage on internal validity, but only at the expense of external validity.[90]

A related problem is that case study methods cannot easily get at "probabilistic" theories, whether those theories involve probabilistic causal mechanisms or whether the operationalization of more deterministic theories involves substantial measurement error. Both kinds of theories lead to probability distributions of predicted outcomes rather than to point predictions and can be falsified with confidence only with a fairly large number of cases. This is a major strength of statistical analysis and a serious limitation of small-N research.[91]

Many case study researchers acknowledge the limitations of their method for the analysis of probabilistic relationships. Unlike large-N researchers, case study researchers want to explain *all* variation and leave to chance no variation or anomalous results, whether because of omitted variables or measurement error. Ragin and Zaret, for example, argue that comparative methods "are logical and not statistical in nature because they are used to identify *invariant* relationships, not statistical or probabilistic relationships

... to identify patterns of constant association, not to explain variation."[92] Ragin argues that "the comparative method does not work with samples or populations but with all relevant instances of the phenomenon of interest. . . . [Explanations] are not conceived in probabilistic terms because every instance of a phenomenon is examined and accounted for if possible. . . . The comparative method is relatively insensitive to the relative frequency of different types of cases."[93] Similarly, Becker argues that "narrative analysts . . . are not happy unless they have a completely deterministic result. Every negative case becomes an opportunity to refine the result, to rework the explanation so that it includes the seemingly anomalous case."[94]

This emphasis on refinement and reworking is the key. Rather than stopping with a probabilistic relationship that explains a certain amount of the variation in outcomes, with omitted variables and measurement error captured by an error term, case study researchers continue to probe in an attempt to reduce further both sources of error and explain additional variation. In doing so they can generate more complete explanations, but of a smaller number of cases and with a loss of parsimony and generalizability.

It may be possible to explain nearly all variation in a handful of variables of interest in a modest number of cases, and some phenomena we want to explain might involve a relatively small number of cases (hegemonic decline, for example), but most phenomena of interest to international relations theorists occur too frequently to conduct detailed case studies of all of them. We need some means of generalizing beyond our sample of cases. Statistical methods do this through a combination of control through partial correlations and randomization of other extraneous influences, but the latter works only if N is large. Although case study methods based on least-likely or most-likely system designs permit generalization, these generalizations are based more on deductive logic (the Sinatra inferences) than empirical demonstration and must be tested on other cases.

Case study researchers face another problem: they have difficulty in assessing the relative causal weights of the various factors influencing a particular outcome, unless those factors are either necessary or sufficient for certain outcomes to result. Case study methods can be useful in determining the presence or the absence of a particular variable and its impact on the presence or absence of outcomes. They might also be able to establish empirically the *direction* of a variable's impact, and perhaps provide a very rough approximation of its impact in terms of categories of high and low. Case studies are much more limited in their ability to determine empirically the relative magnitude of various causal influences when those factors are neither necessary nor sufficient for a given outcome.[95] Because necessary or sufficient conditions are rare, this is another important limitation of case study methods.

The ability to estimate different causal effects empirically is a major strength of statistical analyses of larger numbers of cases (assuming the functional relationship is correctly specified and the key variables are measurable across cases, which can be quite problematic). Regression analyses can estimate the amount of variance explained by each variable, the additional

amount of variance explained when another variable is added to the model, and the proportional effects on the dependent variable of comparable changes in each of the independent variables.

Combining Case Study and Statistical Methods

Statistical methods have their own limitations, of course, particularly concerning the validity of concepts and the operational indicators used to measure them across a large number of cases (the "unit homogeneity" assumption). Proponents of comparative case studies also worry about this problem. They talk at length about balancing the need for "conceptual traveling" (using concepts that are valid across time and space, which facilitates generalization) with the dangers of "conceptual stretching" (applying concepts in historical and cultural contexts in which they have a different meaning or are otherwise not appropriate).[96] This is not the place for a detailed comparative evaluation of the advantages and disadvantages of statistical and comparative methods, but many analysts from each methodological perspective have increasingly come to the conclusion that by combining both statistical and case study methods, researchers can use the advantages of each to partially offset the limitations of the other.

These methods can be combined in a single study or sequentially as part of a "research cycle" in a larger research program.[97] Work that integrates statistical and case study methods in a single study includes Huth's analysis of the success and failure of extended deterrence, Ray's study of the democratic peace, Martin's analysis of the role of international institutions in multilateral economic sanctions, and Simmons's study of the politics of adjustment to international economic pressures of the 1920s and 1930s.[98] Examples of the sequential integration of statistical and case study methods in a larger research program include the International Crisis Behavior Project,[99] Doyle's research program on the democratic peace,[100] and the Mansfield and Snyder project on democratization and war.[101] Although scholars agree on the utility of mixed-method approaches, and although we have begun to see more efforts of this kind, scholars have made few efforts to elaborate on exactly how different methods can be combined or the proper sequence for combining them.

While Russett, Lijphart, and others suggest a sequence involving comparative methods for refining hypotheses followed by statistical methods to test them, reversing this sequence can also be useful.[102] In the democratic peace research program, for example, statistical methods were first used to establish the extraordinarily strong empirical relationship between democratic dyads and the absence of war. Case study methods were then used to validate whether states were properly classified as democracies or nondemocracies; to explore the intervening causal mechanisms linking peaceful outcomes to the characteristics of democracies or possibly to alternative causal mechanisms; and to explore additional testable implications of the democratic peace hypothesis, including differences in leaders' perceptions of democratic and nondemocratic adversaries.[103]

Case study methods can also be usefully combined with formal rational choice theories, in part because of the difficulty of systematically measuring some of the key concepts in rational choice theory (preferences, utilities, probabilities, and informational environments, for example) across a large number of cases. In his study of the July 1914 crisis, for example, Levy empirically examined how political leaders perceived the set of feasible outcomes of the crisis and rank-ordered their preferences over those outcomes, and used this framework to anchor an analytic case study of the outbreak of World War I.[104] Examples of a more deductively structured use of rational choice theory to guide case study analysis include the "analytic narratives" research program[105] and Bueno de Mesquita's analysis of church-state relations in medieval Europe.[106]

Conclusions

In a volume on quantitative and qualitative methods it is natural to focus on questions of method. The utility of particular methods cannot be separated from questions of theory, however, and in many respects the greatest potential for advances in our knowledge about international relations remains theoretical. Developing better theories is particularly important for those who wish to test theories with qualitative methods, because the number of observations needed to test a theory is inversely related to the precision of a theory's predictions. The stronger the theory, the more specific its predictions, and the greater the divergence in predictions from those of a rival theory, the fewer the number of observations that are necessary to provide a satisfactory test of the theory, and thus the more valuable case study methods. This is why case study methods are so useful for testing theories that posit necessary or sufficient conditions. If theories are weaker, and if divergence between the predictions of competing theories is smaller, a greater number of observations are required for a meaningful empirical test, and case study methods are at a disadvantage.

Building better theories is not the only solution here. Another is to think more carefully about the testable implications of existing theories, in terms of quality as well as quantity. Competing theories generate both overlapping and divergent testable implications. The former are irrelevant for testing competing theories, but the latter are critical. As a field we have probably done a better job building theories and developing methods for testing them than thinking creatively about identifying those testable implications of competing theories that are most divergent and consequently most conducive to providing definitive tests between rival theories. The application of all methods would benefit from more imaginative and clever thinking at this critical stage of research, at the juncture of theory and research design, but qualitative case study researchers have particularly strong incentive to do this.

Notes

1. Giovanni Sartori, "Concept Misinformation in Comparative Politics," *American Political Science Review* 64, no. 4 (1970): 1033–53.

2. Idiographic inquiry aims to describe, understand, and interpret individual events or a temporally and spatially bounded series of events, whereas nomothetic inquiry aims to generalize about the relationships between variables and, to the extent possible, construct lawlike propositions about social behavior. See Jack S. Levy, "Explaining Events and Testing Theories: History, Political Science, and the Analysis of International Relations," in *Bridges and Boundaries*, ed. Colin Elman and Miriam Fendius Elman (Cambridge: MIT Press, 2001), 39–83.
3. For example, Brecher, Steinberg, and Stein argued that most empirically oriented foreign policy analyses were single country in orientation and "usually devoid of theoretical value." Michael Brecher, Blema Steinberg, and Janice Stein, "A Framework for Research on Foreign Policy Behavior," *Journal of Conflict Resolution* 13, no. 1 (1969): 75.
4. Sidney Verba, "Some Dilemmas in Comparative Research," *World Politics* 20, no. 1 (1967): 111–27; Adam Przeworski and Henry Teune, *The Logic of Comparative Social Inquiry* (New York: Wiley, 1970); Arend Lijphart, "Comparative Politics and the Comparative Method," *American Political Science Review* 65, no. 3 (1971): 682–93; Arend Lijphart, "The Comparable Cases Strategy in Comparative Research," *Comparative Political Studies* 8, no. 2 (1975): 133–77; Harry Eckstein, "Case Study and Theory in Political Science," in *Handbook of Political Science*, vol. 7, ed. Fred I. Greenstein and Nelson W. Polsby (Reading, Mass.: Addison-Wesley, 1975), 79–138; Alexander L. George, "Case Studies and Theory Development," in *Diplomacy: New Approaches in Theory, History, and Policy*, ed. Paul Lauren (New York: Free Press, 1979), 43–68.
5. Arthur Stinchcombe, *Constructing Social Theories* (New York: Harcourt, Brace and World, 1968); Neil Smelser, "The Methodology of Comparative Analysis," in *Comparative Research Methods*, ed. Donald P. Warwick and Samuel Osherson (Englewood Cliffs, N.J.: Prentice Hall, 1973), 42–86; Theda Skocpol and Margaret Somers, "The Uses of Comparative History in Macrosocial Inquiry," *Comparative Studies in Society and History* 22, no. 2 (1980): 156–73.
6. Barrington Moore Jr., *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World* (Boston: Beacon Press, 1966); Charles Tilly, ed., *The Formation of National States in Western Europe* (Princeton, N.J.: Princeton University Press, 1975); Theda M. Skocpol, *States and Social Revolutions* (New York: Cambridge University Press, 1979).
7. Charles Tilly, *Big Structures, Large Processes, Huge Comparisons* (New York: Russell Sage Foundation, 1984); Alexander George and Timothy McKeown, "Case Studies and Theories of Organizational Decision Making," in *Advances in Information Processing in Organizations*, ed. Robert Coulam and Richard Smith (Greenwich, Conn.: JAI Press, 1985), 43–68; Charles C. Ragin, *The Comparative Method* (Berkeley: University of California Press, 1987); Charles C. Ragin, "Introduction to Qualitative Comparative Analysis," in *The Comparative Economy of the Welfare State*, ed. Thomas Janoski and Alexander M. Hicks (New York: Cambridge University Press, 1994), 299–319; Barbara Geddes, "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics," in *Political Analysis*, vol. 2, ed. James A. Stimson (Ann Arbor: University of Michigan Press, 1990), 131–50; Charles C. Ragin and Howard Becker, eds., *What Is a Case?* (New York: Cambridge University Press, 1992); David Collier, "The Comparative Method," in *Political Science: The State of the Discipline II*, ed. Ada Finifter (Washington, D.C.: American Political Science Association, 1993), 105–19; David Collier and James Mahoney, "Conceptual Stretching Revisited: Adapting Categories in Comparative Analysis," *American Political Science Review* 87, no. 4 (1993): 845–55.
8. Gary King, Robert Keohane, and Sidney Verba, *Designing Social Inquiry* (Princeton, N.J.: Princeton University Press, 1994).
9. David Collier and James Mahoney, "Insights and Pitfalls: Selection Bias in Qualitative Research," *World Politics* 49, no. 1 (1996): 56–91; Douglas Dion, "Evidence and Inference in the Comparative Case Study," *Comparative Poli-*

tics 30, no. 2 (1998): 127–46; Timothy J. McKeown, "Case Studies and the Statistical World View," *International Organization* 53, no. 1 (1999): 161–90; Charles C. Ragin, *Fuzzy-Set Social Science* (Chicago: University of Chicago Press, 2000); Alexander L. George and Andrew Bennett, *Case Studies and Theory Development* (Cambridge: MIT Press, forthcoming).

10. For further information about the consortium, and for syllabi on qualitative methods, see http://www.asu.edu/clas/polisci/cqrm.

11. There has been an interesting reversal in the "balance" between the literature on qualitative methodology and applications of qualitative methods over the last three decades. In the early 1970s, the methodological literature on comparative and case study methods failed to adequately reflect the theoretical and methodological sophistication of some of the best applied research in the field, including work by George and Brecher. Today, applied qualitative research probably lags behind developments in the literature on the methodology of qualitative analysis. Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy* (New York: Columbia University Press, 1974); Michael Brecher, *The Foreign Policy System of Israel: Setting, Images, Process* (New Haven, Conn.: Yale University Press, 1972).

12. Lijphart, "Comparative Politics and the Comparative Method," 682.

13. Collier, "The Comparative Method," 105.

14. Andrew Bennett, "Case Study: Methods and Analysis," in *International Encyclopedia of the Social and Behavioral Sciences*, ed. Neil J. Smelser and Paul B. Baltes (New York: Pergamon, forthcoming).

15. Ragin and Becker, *What Is a Case?*

16. See George, "Case Studies and Theory Development."

17. If we think of cases as instances of broader theoretical categories, and if we acknowledge that historians are more idiographic in orientation than are political scientists, it should not be surprising that historians rarely use the term *case*. Levy, "Explaining Events and Testing Theories."

18. John Walton, "Making the Theoretical Case," in Ragin and Becker, *What Is a Case?* 122.

19. For further discussion of whether cases are made or found see Ragin and Becker, *What Is a Case?*

20. See King, Keohane, and Verba, *Designing Social Inquiry.*

21. Eckstein, "Case Study and Theory in Political Science," 85.

22. Lijphart, "Comparative Politics and the Comparative Method," 691.

23. Eckstein, "Case Study and Theory in Political Science," 96–123. The first two in the list come from Verba, "Some Dilemmas in Comparative Research."

24. Eric Hobsbawm, *On History* (New York: New Press, 1997), 109. Because of their "total" orientation toward the subject matter, atheoretical case studies do not fit the definition of a "case" as an *instance* of a broader class of phenomena.

25. Stephen Van Evera, *Guide to Methods for Students of Political Science* (Ithaca, N.Y.: Cornell University Press, 1997); Levy, "Explaining Events and Testing Theories."

26. One example is Allison's use of three alternative theoretical frameworks to interpret the Cuban missile crisis (though this study had theoretical aims as well). Graham Allison, *Essence of Decision: Explaining the Cuban Missile Crisis* (New York: Little, Brown, 1971).

27. Christopher Achen and Duncan Snidal, "Rational Deterrence Theory and Comparative Case Studies," *World Politics* 41, no. 2 (1989): 145.

28. Ibid., 169.

29. The widely accepted injunction against testing a theory or hypothesis against the same *data* from which it was generated does not preclude us from testing hypotheses generated in a case study against different data from the same case. If a case study of the origins of the cold war leads a researcher to formulate hypotheses about enduring rivalries, and if those hypotheses have implications for the termination of enduring or militarized rivalries, then those same hypotheses can be tested against evidence from the end of the cold war. Indeed,

one of the most useful ways of validating a historical interpretation is to derive additional implications of that interpretation for other aspects of that particular historical episode.

30. George and Smoke, *Deterrence in American Foreign Policy.*

31. Robert Bates, Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry Weingast, *Analytic Narratives* (Princeton, N.J.: Princeton University Press, 1998).

32. George, "Case Studies and Theory Development."

33. See James Lee Ray, *Democracies and International Conflict* (Columbia: University of South Carolina Press, 1995); John Owen IV, *Liberal Peace, Liberal War: American Politics and International Security* (Ithaca, N.Y.: Cornell University Press, 1997).

34. Noted in Howard S. Becker, "Cases, Causes, Conjunctures, Stories, and Imagery," in Ragin and Becker, *What Is a Case?* 209.

35. George, "Case Studies and Theory Development"; George and Bennett, *Case Studies and Theory Development.*

36. David Collier and Steven Levitsky, "Democracy with Adjectives: Conceptual Innovation in Comparative Research," *World Politics* 49, no. 3 (1997): 430–51. One caveat is worth noting. If the proliferation of subtypes and of the number of causal paths leading to them is taken to an extreme, the theory might explain all the anomalies but only at the cost of a loss of analytic power and parsimony. This problem is minimized if the variety of causal paths is linked to the conditions under which each is most likely to occur.

37. Arthur L. Stinchcombe, *Theoretical Methods in Social History* (New York: Academic Press, 1978), 5.

38. Lijphart, "The Comparable Cases Strategy in Comparative Research," 159.

39. Hugh Stretton, *The Political Sciences* (London: Routledge and Kegan Paul, 1969), 245–47.

40. Lijphart, "Comparative Politics and the Comparative Method," 692.

41. The many variables/small-N problem affects the use of case studies for theory testing but not for describing and explaining individual historical episodes or for generating or refining hypotheses to be tested by other means. If the universe of cases is relatively small (hegemonic wars, for example), and if the aim of case study analysis is to generate hypotheses, there may be advantages of using a small number of cases. This leaves a maximum number of cases for hypothesis testing, given the need to test hypotheses on data that are independent of the data from which the hypotheses are generated.

42. See also King, Keohane, and Verba, *Designing Social Inquiry,* 217–18.

43. Lijphart, "Comparative Politics and the Comparative Method," 685.

44. Lijphart, "The Comparable Cases Strategy in Comparative Research," 163.

45. Campbell also emphasized the utility of a single case, but only after retracting his earlier argument to the contrary. Donald Campbell, "Degrees of Freedom and the Case Study," *Comparative Political Studies* 8, no. 2 (1975): 178–93. See also McKeown, "Case Studies and the Statistical World View."

46. John Frendreis, "Explanation of Variation and Detection of Covariation: The Purpose and Logic of Comparative Analysis," *Comparative Political Studies* 16, no. 2 (1983): 255.

47. John Stuart Mill, *A System of Logic* (1875; reprint, London: Longman, 1970).

48. Przeworski and Teune, *The Logic of Comparative Social Inquiry;* Theodore Meckstroth, "'Most Different Systems' and 'Most Similar Systems': A Study in the Logic of Comparative Inquiry," *Comparative Political Studies* 8, no. 2 (1975): 133–77.

49. This terminology has generated some confusion. Mill defines agreement or difference in terms of the *dependent* variable, whereas most similar and most different are defined in terms of *explanatory* variables. Thus Mill's method of agreement is equivalent to a most-different systems design, and Mill's method of difference is equivalent to a most-similar systems design.

50. Lijphart, "The Comparable Cases Strategy in Comparative Research"; Smelser, "The Methodology of Comparative Analysis."

51. Przeworski and Teune, *The Logic of Comparative Social Inquiry.*
52. Mill, *A System of Logic,* bk. 6, ch. 7.
53. Lijphart, "Comparative Politics and the Comparative Method," 688.
54. George, "Case Studies and Theory Development"; George and Bennett, *Case Studies and Theory Development.*
55. Stephen Peter Rosen, *Societies and Military Power: India and Its Armies* (Ithaca, N.Y.: Cornell University Press, 1996).
56. Jack Snyder, *Myths of Empire: Domestic Politics and International Ambition* (Ithaca, N.Y.: Cornell University Press, 1991).
57. See King, Keohane, and Verba, *Designing Social Inquiry,* 126; Collier and Mahoney, "Insights and Pitfalls."
58. King, Keohane, and Verba, *Designing Social Inquiry,* ch. 4.
59. Ibid., ch. 4.
60. Collier and Mahoney, "Insights and Pitfalls."
61. An example involving large-N designs is Bueno de Mesquita's empirical analysis of his core hypothesis that a positive expected utility for war is a necessary condition for a state to initiate war. Bruce Bueno de Mesquita, *The War Trap* (New Haven, Conn.: Yale University Press, 1981).
62. Dion, "Evidence and Inference in the Comparative Case Study."
63. The proposition that democracies rarely if ever go to war with each other posits that joint democracy is a sufficient condition for peace, and for the analysis of this proposition scholars focus exclusively on cases of democratic dyads.
64. Ragin, *The Comparative Method.*
65. Ragin argues that because the number of interaction effects necessary to capture combinatorial effects increases rapidly with the number of variables, serious degrees-of-freedom problems result. Ibid. For a new statistical approach to modeling multiple causal paths see Bear F. Braumoeller, "Modeling Multiple Causal Paths: Logic, Derivation, and Implementation" (manuscript, Harvard University, 2000).
66. For an application of Boolean models to the study of deterrence see Frank P. Harvey, "Practicing Coercion: Revisiting Successes and Failures Using Boolean Logic and Comparative Methods," *Journal of Conflict Resolution* 43, no. 6 (1999): 840–71.
67. Eckstein, "Case Study and Theory in Political Science," 113–23.
68. This assumes that measurement error is low.
69. Arend Lijphart, *The Politics of Accommodation* (Berkeley: University of California Press, 1968).
70. See David Bicknell Truman, *The Governmental Process* (New York: Knopf, 1951).
71. Ronald Rogowski, "The Role of Theory and Anomaly in Social-Scientific Inference," *American Political Science Review* 89, no. 2 (1995): 467–70. Lijphart also contributed to theory construction by suggesting hypotheses on why the theory of cross-cutting cleavages broke down in the Netherlands. See Lijphart, *The Politics of Accommodation.*
72. Imre Lakatos, "Falsification and the Methodology of Scientific Research Programmes," in *Criticism and the Growth of Knowledge,* ed. Imre Lakatos and Alan Musgrave (Cambridge: Cambridge University Press, 1970), 91–196.
73. Allison, *Essence of Decision.*
74. Similarly, if the rational-unitary model cannot explain state behavior in an international crisis as acute as the one in 1962, we would have little confidence that it could explain behavior in situations of noncrisis decision making.
75. George, "Case Studies and Theory Development"; George and Bennett, *Case Studies and Theory Development.*
76. Ibid.
77. Ernest Nagel, *The Structure of Science* (Indianapolis: Hackett, 1979), 564–68; W. B. Gallie, "The Historical Understanding," *History and Theory* 3, no. 2 (1963): 149–202.
78. Carl G. Hempel, "The Function of General Laws in History," *Journal of Philosophy* 39 (1942): 35–48.

79. Experimental methods may be superior for testing many of these hypotheses because of their ability to control for extraneous variables. It is often difficult, however, to generalize from laboratory settings that do not incorporate the relevant political variables and that cannot fully replicate the stakes and emotions inherent in the contexts of foreign policy decision making.
80. Sidney Tarrow, "Bridging the Quantitative-Qualitative Divide in Political Science," *American Political Science Review* 89, no. 2 (1995): 474.
81. Paul Pierson, "Increasing Returns, Path Dependence, and the Study of Politics," *American Political Science Review* 94, no. 2 (2000): 251–67; Ira Katznelson, "Periodization and Preferences: Contributions of Comparative Historical Social Science" (manuscript, Columbia University, 2000).
82. Andrew Bennett and Alexander L. George, "The Alliance of Statistical and Case Study Methods: Research on the Interdemocratic Peace," *APSA-CP Newsletter* 9, no. 1 (1998): 6–9.
83. A good example of a large-N study that empirically distinguishes between alternative causal mechanisms within the black box of decision making is Schultz's statistical test of competing institutional constraint and informational signaling models of the democratic peace proposition. Kenneth A. Schultz, "Do Democratic Institutions Constrain or Inform? Contrasting Two Perspectives on Democracy and War," *International Organization* 53, no. 2 (1999): 233–66.
84. Many of those committed to large-N analysis exaggerate in the opposite directions.
85. Clayton Roberts, *The Logic of Historical Explanation* (University Park: Pennsylvania State University Press, 1996).
86. Thomas D. Cook and Donald T. Campbell, *Quasi-Experimentation* (Chicago: Rand McNally, 1979), ch. 1.
87. If theories have no observable consequences, they cannot explain the variation in outcomes that is the primary task of social science, and hence those theories have no explanatory power. This does not imply that all theories are easy to test.
88. For an argument that hypotheses involving necessary conditions are more common, see Gary Goertz and Harvey Starr, eds., *Necessary Conditions: Theory, Methodology, and Applications* (Lanham, Md.: Rowman and Littlefield, forthcoming).
89. These are very demanding criteria. They lead Lieberson to conclude that Mill's methods cannot be applied if the number of cases is small because they do "not allow for probabilistic theories, interaction effects, measurement errors, or even the presence of more than one cause." Stanley Lieberson, "Small N's and Big Conclusions," in Ragin and Becker, *What Is a Case?* 105–18; Stanley Lieberson, "More on the Uneasy Case for Using Mill-Type Methods in Small-N Comparative Studies," *Social Forces* 72 (June 1994): 1225–37. Multiple causes and interaction effects can probably be dealt with, but only by increasing the number of cases in order to encompass all possible interaction effects. This creates a practical problem for intensive case study methods. Some case study researchers acknowledge this problem and supplement Mill's methods with process tracing (George and Bennett, *Case Studies and Theory Development*), while others have adopted a Boolean-based "qualitative comparative analysis" (Ragin, *The Comparative Method*).
90. The large-N researcher makes the opposite trade-off.
91. Whether social science theories can have causal mechanisms that are truly probabilistic—as opposed to those that reflect measurement error, the misspecification of functional forms, or omitted variables that have yet to be understood and specified—is quite problematic. See Wesley C. Salmon, *Four Decades of Scientific Explanation* (Minneapolis: University of Minnesota Press, 1990), and George and Bennett, *Case Studies and Theory Development.* I prefer to speak of a theory's probabilistic testable implications rather than probabilistic causal mechanisms. My thinking (admittedly provisional) on this complex issue has benefited from conversations with Andrew Bennett.

92. Ragin and Zaret, 1983, 744.
93. Ragin, *The Comparative Method*, 15. Ragin's exploration of the relevance of the logic of fuzzy sets for social science analysis is an attempt to deal with the problem of measurement error and probabilistic testable implications. Ragin, *Fuzzy-Set Social Science*.
94. Becker, "Cases, Causes, Conjunctures, Stories, and Imagery," 212.
95. Dorothy Vaughan, "Theory Elaboration: The Heuristics of Case Analysis," in Ragin and Becker, *What Is a Case?* 184. Proponents of "offense/defense theory," for example, theorize about whether military technology favors the offense or defense but fail to establish the magnitude of this effect relative to that of the balance of material capabilities and related variables. Sean M. Lynn-Jones, "Offense-Defense Theory and Its Critics," *Security Studies* 4, no. 4 (1995): 660–91.
96. Sartori, "Concept Misinformation in Comparative Politics"; Collier and Mahoney, "Conceptual Stretching Revisited."
97. Skocpol and Somers, "The Uses of Comparative History in Macrosocial Inquiry."
98. Paul Huth, *Extended Deterrence and the Prevention of War* (New Haven, Conn.: Yale University Press, 1988); Ray, *Democracies and International Conflict*; Lisa L. Martin, *Coercive Cooperation* (Princeton, N.J.: Princeton University Press, 1992); Beth A. Simmons, *Who Adjusts?* (Princeton, N.J.: Princeton University Press, 1994).
99. Brecher, *The Foreign Policy System of Israel*; Michael Brecher, with Benjamin Geist, *Decisions in Crisis: Israel, 1967 and 1973* (Berkeley: University of California Press, 1980); Michael Brecher and Jonathan Wilkenfeld, *A Study of Crisis* (Ann Arbor: University of Michigan Press, 1997).
100. Michael Doyle, "Liberalism and World Politics," *American Political Science Review* 80, no. 4 (1986): 1151–70; Michael Doyle, *Ways of War and Peace* (New York: Norton, 1997).
101. Edward D. Mansfield and Jack Snyder, "Democratization and the Danger of War," *International Security* 20, no. 1 (1995): 5–38; Jack Snyder, *From Voting to Violence: Democratization and Nationalist Conflict* (New York: Norton, 2000).
102. Bruce M. Russett, "International Behavior Research: Case Studies and Cumulation," in *Approaches to the Study of Political Science*, ed. Michael Haas and Henry S. Kariel (Scranton, Penn.: Chandler, 1970), 425–43; Lijphart, "Comparative Politics and the Comparative Method."
103. Bennett and George, "The Alliance of Statistical and Case Study Methods."
104. Jack S. Levy, "Preferences, Constraints, and Choices in July 1914," *International Security* 15, no. 3 (1990–91): 151–86.
105. Bates et al., *Analytic Narratives*.
106. Bruce Bueno de Mesquita, "Popes, Kings, and Endogenous Institutions: The Concordat of Worms and the Origins of Sovereignty," *International Studies Review* 2, no. 2 (2000): 93–118.