

2. Case Study Methods: Design, Use, and Comparative Advantages

Andrew Bennett

There is a growing consensus among social scientists that research programs advance more effectively through the iterative or collaborative use of different research methods than through the use of any one method alone. Making the most of the synergies among research methods requires an understanding of the relative comparative advantages, trade-offs, and limitations of each method and an ability to translate between different methods. The comparative advantages of case study methods include identifying new or omitted variables and hypotheses, examining intervening variables in individual cases to make inferences on which causal mechanisms may have been at work, developing historical explanations of particular cases, attaining high levels of construct validity, and using contingent generalizations to model complex relationships such as path dependency and multiple interactions effects. Particularly important is the ability to identify new hypotheses, which case studies can do through a combination of deduction and induction.

Recurrent trade-offs in the use of case study methods include the problem of case selection and the danger of selection bias, which can have more severe consequences in case studies than in statistical studies, and the tension between parsimony and richness in selecting the number of variables and cases to be studied. In addition, case study findings are usually contingent and can be generalized beyond the type of case studied only under specified conditions, such as when a case study shows that a variable is not

a necessary condition or a sufficient condition for an outcome, or when a theory fails to fit a case that it appeared most likely to explain. Potential limitations of case studies, though not inherent in every one, include indeterminacy or inability to exclude all but one explanation, lack of independence of cases, and the impossibility of perfectly controlling case comparisons.

The inherent limitations of case study methods include their relative inability to render judgment on the frequency or representativeness of particular cases and their weak capability for estimating the average “causal weight” of variables. These are inferential processes for which case studies are not designed and cannot be used except in a rudimentary manner. Fortunately, these inherent limitations correspond almost exactly with the comparative advantages of statistical methods, which give various measures of frequency and can estimate the expected causal weight of a variable.

This chapter defines and explicates case study methods and details their comparative advantages and limitations. It then more briefly reviews the strengths and limits of formal models and statistical methods. This analysis substantiates the conclusion that the comparative advantages of case study methods are complementary to those of statistical methods and formal models. It concludes with suggestions for increasing multimethod collaboration among researchers to make the best possible use of this complementarity.

Overview of Case Study Methods

Defining *Case* and *Case Studies*

A *case* is often defined as a “phenomenon for which we report and interpret only a single measure on any pertinent variable” (Eckstein 1975). This wrongly implies, however, that each case has only one observation on the dependent variable but many independent variables. If this were true, it would present an inherent problem of indeterminacy, or an inability to choose among competing explanations for a case.¹ Yet each “case” in fact has a potentially large number of observations on intervening variables and may allow several qualitative measures of various dimensions of the independent and dependent variables, so case studies do not necessarily suffer from indeterminacy (King, Keohane, and Verba 1994, 225; Campbell 1975, 179, 181–82). I therefore follow the definition of a *case* as an instance

of a class of events of interest to the investigator (George 1979a), such as an instance of revolution, type of governmental regime, kind of economic system, or personality type. A *case study* is thus a well-defined aspect of a historical happening that the investigator selects for analysis, rather than a historical happening itself. The Soviet revolution, for example, is an instance of civil conflict, war termination (the Soviet pullout from World War I), the role of personality in politics, and so on. The investigator decides which class of events, which facets of the Soviet revolution, and which variables to focus upon.²

There is also potential for confusion among the terms *comparative methods*, *case study methods*, and *qualitative methods*. I use the term *case study methods* to refer to both within-case analysis of single cases and comparisons among a small number of cases, as most case studies involve both kinds of analysis due to the limits of either method used alone. Even single-case studies usually draw implicit comparisons to wider groups of cases. As for the term *qualitative methods*, this is sometimes used to encompass both case studies carried out with a neopositivist view of the philosophy of science and those implemented with a postmodern or interpretive view. In the present chapter I use the term *case study* to refer only to studies that aspire to causal explanations, setting aside those interpretivist and postmodernist analyses that eschew such explanations or view them as unattainable.

Types of Theory-Building Contributions of Case Studies

Within this general definition of case studies, there are many types of case studies. Some methodological texts focus on theory-testing cases at the expense of theory development. It is important to keep in mind, however, that there are several kinds of contributions to theory, including the generation of new hypotheses (the “logic of discovery”) as well as the testing of existing ones (the “logic of confirmation”). In addition, there are several kinds of research objectives, including not only the development of generalized theories but the *historical explanation* of particular cases, that is, explanation of a sequence of events that produce a particular historical outcome in which key steps in the sequence are in turn explained with reference to theories or causal mechanisms. Case studies can contribute to all of these kinds of theory building, as Arend Lijphart (1971) and Harry Eckstein (1975) indicated in their similar taxonomies of different kinds of case studies, outlined in table 1 (from George 1979a).

Apart from the first type, which is simply a kind of chronological narrative, case studies have an explanatory or theory-building purpose. “Interpretive” or “disciplined configurative” cases use theoretical variables to provide historical explanations of particular cases. In other words, they use theories to show that in the particular historical circumstances of the case, the outcome was to be expected. Heuristic case studies seek to generate new hypotheses inductively. “Deviant” cases, or cases whose outcomes are not predicted or explained well by existing theories, can be particularly useful in identifying new or left-out variables. Finally, researchers can use case studies to test whether extant theories accurately explain the processes as well as the outcomes of particular cases. Herein, I use Eckstein’s terminology, which is more common, with the addition of Lijphart’s term for the study of “deviant” cases.

Within-Case Methods of Analysis

Process Tracing

There are three methods of within-case analysis: process tracing, congruence testing, and counterfactual analysis.³ Process tracing focuses on whether the intervening variables between a hypothesized cause and observed effect move as predicted by the theories under investigation. Put another way, process tracing looks at the observable implications of putative causal mechanisms in operation in a case, much as a detective looks for suspects and for clues linking them to a crime. The goal is to establish which of several possible explanations is consistent with an uninterrupted chain of evidence from hypothesized cause to observed effect. The power of process tracing arises from the fact that it requires continuity and com-

TABLE 1. Equivalent Terms for Types of Case Studies

Arend Lijphart	Harry Eckstein
Atheoretical case study	Configurative-ideographic case study
Interpretative case study	Disciplined-configurative case study
Hypothesis-generating case study	Heuristic case study
Deviant case study	(No comparable term or concept)
Theory-confirming/infirming case study	Crucial, most likely, least likely test cases

pleteness in explaining a case (although there are pragmatic limits on the ability or need to examine the infinite “steps between steps” in a temporal process). If even a single significant step in a hypothesized process is not as predicted, the hypothesis must be modified, sometimes trivially and other times substantially, if it is to explain the case. If, for example, 98 of 100 dominoes standing in a straight line knock one another over but the 99th domino does not fall or strike the final domino, we need a separate explanation for why the 100th domino has fallen.

This contrasts sharply with statistical methods, which rely on probabilistic associations but do not require continuity or completeness in any given case. In this sense, process tracing is different from the notion of “pattern matching” outlined by Donald Campbell (1975). Campbell does not elaborate in any detail on what he means by “pattern matching,” but he indicates that it involves finding similar patterns or sequences in different cases, and he does not define it to include an analysis of the full sequence of events in either case. This is potentially an important form of inference that combines elements of cross-case comparison with some degree of within-case analysis, but it does not require full continuity or completeness and hence cannot constitute a historical explanation of either case. The distinction is analogous to the difference between finding common short sequences in a long strand of DNA that may offer clues to its operation (pattern matching) and attempting to explain how the full strand operates to express itself in the life form to which the DNA belongs (process tracing).

In any particular study, there can be a deductive element to process tracing, an inductive element, or both. Deductively, the researcher uses theories to predict the values of intervening variables in a case and then tests these predictions. This may require filling in the predictions that under-specified theories should make in a case, and it is important to trace the predicted processes of alternative hypotheses as well as those of the main hypothesis of interest. Inductively, the researcher should be open to unexpected clues or puzzles that indicate the presence of left-out variables. This can lead to the development of new hypotheses.

One common misconception here is that it is always illegitimate to derive a hypothesis from a case and then test it against the same case. In fact, it may be possible to develop a hypothesis from a case and then test it against *different evidence* in the same case. Detectives, of course, do this all the time: clues may lead to a new “theory of the case,” which prompts the

detective to look for “new” evidence in the case that had previously been ignored or considered irrelevant. If the new evidence fits the prediction of the new theory, this is considered an independent corroboration.⁴

Process tracing is not infallible. Measurement error and omitted variables can lead to incorrect inferences in process tracing just as they can in statistical methods. There are also practical limits on our ability to observe or trace processes in all of their nearly infinite detail and to establish fully continuous sequences. The requisite evidence may not be available at key steps in the process, and even where evidence is available, we may not have the time to go through all of it. Yet by insisting that we establish explanations that document the intervening variables and processes through which the hypothesized independent variables are purported to have brought about the observed outcome, process tracing differs from and complements statistical inferences. Although no case study is undertaken in the infinite level of detail that would be needed to establish a *fully* continuous process, case study explanations are open to challenge if they are inconsistent with the finest level of detail that is observable. For example, if a rational choice theory posits that an individual should have gone through a rational calculation that led to a certain behavior, but it can be shown in a case study that the individual’s thinking process was actually very different from that posited by the theory, then the theory cannot constitute a satisfactory explanation of the case even if its predicted outcome is consistent with the observed outcome.

Congruence Testing

In congruence testing, the researcher focuses on the values of the independent and dependent variables rather than the intervening variables. Here, the researcher tests whether the predicted value of the dependent variable, in view of the values of the case’s independent variables, is congruent with the actual outcome in the case. Congruence tests are usually less conclusive than process tracing because in the social sciences we usually lack precise models of the value that the individual variables, individually and collectively, should produce in the dependent variable. In this sense, congruence tests in a single case or a small number of cases are a less reliable version of statistical tests of covariation or estimates of partial correlations among a large number of cases. Still, congruence tests may be able to rule out pro-

posed necessary or sufficient conditions, and they may weaken the plausibility of particular historical explanations of cases.

Counterfactual Analysis

Counterfactual analysis inverts the standard mode of inference for empirical testing of assertions such as “ x in a specified case was necessary for y .” This assertion poses a logically equivalent counterfactual, namely, “if not- x had occurred in the case, then not- y would have occurred.” Analogously, a claim that a variable x is sufficient for an outcome y , whether made for all y or only for specific contexts, can be assessed by looking at the equivalent counterfactual, “not- y could occur only if not- x .” Interest in counterfactual analysis has increased in recent years (Fearon 1991; Tetlock and Belkin 1996). At the same time, there is an obvious danger of confirmation bias and spuriousness if counterfactual analysis is carried out in an undisciplined way.

Philip Tetlock and Aaron Belkin have devised a useful taxonomy of counterfactual analyses. These include “idiographic case-study counterfactuals,” which focus on points of supposed historical contingency in individual cases, and “nomothetic counterfactuals,” which apply well-defined theories to specific antecedent counterfactual conditions. These authors argue that an especially important type of counterfactual combines these two, bringing together in-depth knowledge of particular cases with strong theories about the consequences of particular values of a variable to produce convincing accounts of what should have been true if one variable in a case had assumed a particular value. They illustrate this with the example of dinosaur extinction (1996, 6–11): If an asteroid of a size sufficient to cause climatic change had struck the earth 65 million years ago, what testable implications should be observable in contemporary geologic evidence? This combines the known specifics of the dinosaur extinction case with theories on asteroid impacts to produce testable assertions, and it thus moves from the counterfactual to the factual. The difficulty of applying this to the social sciences, as they note, is that we generally lack “idiographic-nomothetic syntheses of comparable scope and sweep in world politics” (11).

Tetlock and Belkin also offer sensible advice on criteria for defining good counterfactual analyses, including clarity in defining the variables,

minimization of the necessary rewriting of history, and consistency with established theories and statistical findings. Most important, they suggest that good counterfactuals must have testable implications in the factual world (1996, 18). They also note that if we find a causal argument plausible but its equivalent counterfactual argument implausible, or vice versa, we must reconcile the asymmetry in our thinking. This can help identify double standards, inconsistent causal reasoning, and hindsight bias (13). Subject to these criteria, counterfactual analysis is a useful tool in the explanation of individual cases and can provide a check on confirmation bias rather than an open license to rewrite history.

Research Design Tasks

There are five research design tasks common to both single and comparative case studies, many of them common to statistical studies as well (George 1979a; George and McKeown 1985). First, the researcher must define the research objective, including the class of events to be explained, the alternative hypotheses under consideration, and the kind of theory building to be undertaken. Second, the researcher must specify the independent, dependent, and intervening variables and decide which of these are to be controlled for and which are to vary across cases or types of cases. Third, the researcher selects the cases to be studied, possibly assisted by the typological space that results from the specification of the variables and alternative hypotheses. Fourth, the researcher should consider how best to describe variance in the independent and dependent variables, considering not only individual variables but also types of cases, or combinations of variables, and the sequential pathways that characterize each type. Finally, the researcher specifies the structured questions to be asked of each case in order to establish the values of the independent, intervening, and dependent variables.

An example from my own work illustrates how these tasks were accomplished in one study.⁵ I chose to study Soviet and Russian military interventionism and to try to explain the puzzle of why such interventionism appeared to increase in the 1970s, decrease in the 1980s, and increase once again in the mid-1990s. I first had to define *interventionism*, the propensity for intervention, as distinct from actual military *interventions*. This required defining in a general way what constituted an inviting or uninviting

“opportunity” for military intervention, which I did by looking at a typology of situational factors, and by comparing opportunities in which the Soviet Union or Russia intervened in one period to analogous opportunities in which there was no intervention, or a withdrawal from an existing intervention, in another period. I also decided to focus on one subtype of intervention: the high end of the scale involving the direct use of Soviet or Russian troops or commanders.

For the alternative explanations of patterns in Soviet-Russian interventionism, I included standard theories from the IR literature based on the systemic, domestic, organizational, and individual levels, as well as theories based on the arguments of area experts and policymakers. The explanation that interested me most was that Soviet and Russian leaders learned lessons from their ongoing experiences that made them more willing to resort to military intervention in the 1970s, less so in the 1980s, and more so once again in the mid-1990s.⁶ I specified the variables for each of the alternative explanations and carried out both congruence and process-tracing tests on each explanation. Explanations based on changes in the balance of military forces, for example, were consistent with the rise of Soviet interventionism in the 1970s, but not with its decline in the 1980s when Soviet forces were still strong or with its resurgence in 1994 when Russian forces were weak. For the “learning” explanation, eight specific beliefs on the efficacy of using force were listed, such as beliefs on whether “balancing” or “bandwagoning” is the most likely response by others to the use of force. I also defined corresponding behaviors, such as the intensity of efforts to get other regional states or national liberation movements to bandwagon with Soviet-Russian efforts.

After considering for study more than a dozen cases of interventions, noninterventions, and withdrawals, I chose to examine the Soviet-Cuban intervention in Angola in 1975, the Soviet intervention in Afghanistan in 1979, the ongoing occupation of Afghanistan through the 1980s, the withdrawal from Afghanistan in 1989, and the Russian intervention in Chechnya in 1994. I was assisted in this process by a chart outlining my preliminary knowledge on how each of the possible cases for study fit with respect to their values on the seven independent variables identified by the hypotheses. This helped ensure that the cases I chose included wide variation in both the independent and dependent variables. It also made clear which other cases might have been included, thereby “leaving up the scaf-

folding” for future researchers to build upon or future critics to question (King, Keohane, and Verba 1994). This is an important aspect of research design that is seldom adequately carried out.

I also used my preliminary knowledge of the cases to select from among the cases that history provided the ones that provided the most analytical leverage on my research objective and that best fit a strong “most similar cases” research design. Since the learning hypothesis focused on a path-dependent historical process, I chose cases that covered the entire historical period from 1973 (the lead-up to Soviet intervention in Angola) to 1996 (the peak of modern Russia’s first intervention in Chechnya). I also included cases of different types of intervention, such as direct intervention (Afghanistan, Chechnya) and proxy intervention (Angola). Also, the before-and-after cases of intervention in and withdrawal from Afghanistan provided a most-similar-case comparison that controlled for many variables, such as intrinsic geographic importance. Controlling for other variables that changed over time, such as U.S. policy, was done through a combination of case comparisons and process tracing. Thus, as is common, the research design included both within-case analysis of every case and cross-case comparisons.

The questions asked of each case included those that established the values of the independent and intervening values for each hypothesis and the outcome of the case. For the learning hypothesis, the case studies tested whether stated Soviet and Russian beliefs changed in response to experience and were congruent with Soviet behavior. They also tested whether the patterns and timing of changes in stated beliefs fit the dynamics predicted by theories of individual, organizational, and governmental learning. A particularly important test was whether individuals’ stated beliefs fit better with their apparent material interests, as many explanations argued they should, or with the experiences and information to which individuals were exposed, as learning theory predicted. Finally, the study designated fifty-five key Soviet and Russian officials whose stated views were traced through public statements, archival documents, interviews, and memoirs.⁷

The actual case studies found substantial changes in stated beliefs over time that correlated closely with actual Soviet-Russian behavior. It was also able to trace these changes of beliefs to ongoing Soviet-Russian experiences in the use of force and to show that beliefs were often correlated more closely with individuals’ experiences than with their bureaucratic or

material interests. Many military officers who fought in Afghanistan, for example, strongly protested the use of Russian troops in Chechnya, even to the point of losing their jobs and ending their careers. These conclusions constituted strong evidence for the general applicability of learning theory, as in many respects the closed Soviet system was a least likely case for learning. The cases studied did not include any crucial cases, however, as U.S. policy responses and Soviet-Russian domestic politics were also broadly consistent with changes in Soviet behavior.

Single-Case Research Designs

Within the context of general research design tasks, there are specific considerations that apply to single and comparative case studies. Some methodologists have downplayed the theory-building contributions that can be made by single-case research designs (King, Keohane, and Verba 1994, 209–11). In contrast, most case study researchers have argued that single-case studies can provide tests that might strongly support or impugn theories. Many influential research findings in political science have come from single-case studies that presented anomalies for accepted theories.⁸

An important single-case research design is the study of crucial, most likely, and least likely cases that pose severe tests of theories. Harry Eckstein developed the idea of a “crucial case,” or a case that “*must closely fit* a theory if one is to have confidence in the theory’s validity, or, conversely, *must not fit* equally well any rule contrary to that proposed” (1975, emphasis in original). Because true crucial cases were rare in Eckstein’s view, he pointed to the alternative of “most likely” and “least likely” cases. A most likely case is one that is almost certain to fit a theory if the theory is true for any cases at all. The failure of a theory to explain a most likely case greatly undermines our confidence in the theory. A least likely case, conversely, is a tough test for a theory because it is a case in which the theory is least likely to hold true. Eckstein’s conception is a useful starting point on theory testing in case studies, but it is at best incomplete because he does not address whether the cases in question are most or least likely for competing theories, or whether these theories predict the same outcome as the theory of interest or a different outcome altogether. Thus, a more complete version of Eckstein’s insight would be that a theory is most strongly supported when it makes a clear prediction on the outcome or process of a

case, all other theories make clear predictions that we should not find this outcome or process, and the first theory is corroborated in the case. Conversely, if both our theory of interest and the alternative theories make the same prediction on the outcome or process of a case, but this prediction proves wrong, then the theory of interest is strongly impugned because its failure cannot be explained away by the operation of other theories or mechanisms.⁹ Single-case studies that fit either of these situations can greatly increase or decrease our confidence in a theory or require that we alter its scope conditions, although we can never entirely rule out the possibility that the outcome or process of the case was caused by probabilistic processes analogous to those of quantum mechanics.

Another important single-case research design is the study of a deviant or outlier case. Research on deviant cases can help inductively identify variables and hypotheses that have been left out of existing theories. Deviant cases may also uncover measurement errors that may exist in less extreme forms in other cases.

Single-case study designs can fulfill the other theory-building purposes identified by Lijphart and Eckstein as well. Idiographic studies, while often disdained, may provide data for later more theoretically oriented case studies. Also, a study of a newly defined puzzle or phenomenon might begin with a fairly open-ended effort—sometimes called “soaking and poking” in the data—to generate hypotheses that can then be tested more systematically.¹⁰

Comparative Methods

Mill's Methods and Most-Similar and Least-Similar Case Comparisons

Comparisons between cases are a powerful source of causal inferences but also a potential source of inferential errors. One mode of case comparisons is Mill's method of agreement, in which the investigator looks for the potentially causal antecedent conditions that are the same between two cases that have the same outcome. Ideally, these would turn out to be necessary conditions. Thus, if we compared the following two cases using Mill's method of agreement, we might infer that the variable A is causally related to the outcome Y, as it is the only independent variable common to the two cases.

Case Study Methods

Mill's Method of Agreement

	Independent Variables	Dependent Variable
Case 1	A B C D E	Y
Case 2	A F G H I	Y

This method of agreement corresponds, somewhat confusingly, with what has been called the “least similar cases” research design. If, for example, we find that teenagers are “difficult” in both tribal societies and industrialized societies, we might be tempted to infer that it is the nature of teenagers rather than the nature of society that accounts for the difficulty of teenagers (Przeworski and Teune 1970).

In Mill's method of difference, the investigator would look for antecedent conditions that differ between two cases that have different outcomes, and they would judge that those antecedent conditions that were the same despite differing outcomes could not be sufficient to cause either outcome. In the following example (where $\sim A$ represents “not A”) the researcher would draw the inference that the variable A was causally related to the outcome because it is the only one that varies when the outcome varies.

Mill's Method of Difference

	Independent Variables	Dependent Variable
Case 1	A B C D E	Y
Case 2	$\sim A$ B C D E	$\sim Y$

This corresponds with the “most similar case” research design (Przeworski and Teune 1970). It has also been called the method of “controlled comparison,” because if two cases in fact are the same in all but one independent variable, then we have the functional equivalent of a controlled experiment. The practical limitation here, of course, is that two cases are almost never identical in all but one independent variable (George 1979a).¹¹

In actual practice, case study researchers almost never draw conclusions on the basis of Mill's methods alone because these methods require demanding and unrealistic assumptions in order to provide nonspurious inferences. One key limitation of Mill's methods, which Mill himself identified, is that they cannot work well in the presence of equifinality (George 1982). A condition of equifinality, or what Mill called a “plurality

of causes,” holds when the same outcome can arise through different pathways or combinations of variables. Thus, when equifinality is present, there might be no single necessary or sufficient variable for a phenomenon: it might be that either ABC or DEF causes Y, and that none of the variables A through F is by itself sufficient to cause Y. In such circumstances, pairwise comparisons of cases might lead us wrongly to reject variables that can cause an outcome in conjunction with some contexts but not others, and it might also lead us to accept a confounding variable as causal rather than recognizing that its relationship to the outcome is spurious.

Thus Mill’s methods can work well at identifying causal relations only under three conditions that are impossible to realize fully in practice. First, the causal relations being investigated must be deterministic regularities involving conditions that by themselves are either necessary or sufficient for a specified outcome. This implies that there can be no causally relevant interaction effects. Second, all variables that contributed causally to the outcome would have to be identified and included in the analysis. Third, cases that represent the full range of all logically and socially possible causal paths must be available for study (Little 1998; George and McKeown 1985).

Because these requirements are unrealistic, case study researchers use Mill’s methods in only a very general and preliminary way to identify potentially relevant variables, but they then rely heavily on process tracing to compensate for the evident weakness of Mill’s methods (Mahoney 1999).¹² For example, when it is not possible to find cases similar in all but one independent variable and the dependent variable, process tracing can test whether each of the potentially causal variables that differ between the imperfectly matched cases can be ruled out as having causal significance.¹³

Structured, Focused Comparison of Cases and the Development of Typological Theories

In response to the limitations of Mill’s methods and controlled comparison, Alexander George (1979a, 1979b) systematized case study procedures and developed the method of “structured focused case comparisons.” In this method, the researcher systematically (1) specifies the research problem and the class of events to be studied; (2) defines the independent, dependent, and intervening variables of the relevant theories; (3) selects

the cases to be studied and compared; (4) decides how best to characterize variance in the independent and dependent variables; and (5) formulates a detailed set of standard questions to be applied to each case.

In addition, consistent with his emphasis on equifinality, George argued that case studies could be especially useful in developing what he called “typological theories,” or contingent generalizations on “*the variety of different causal patterns that can occur for the phenomena in question [and] the conditions under which each distinctive type of causal patterns occurs*” (1979a, emphasis in original). He advocated a kind of “building block” approach to the development of theories. In this approach, each case, while rendered in terms of theoretical variables, might prove to be a distinctive pathway to the outcome of interest. Typological theories treat cases as *configurations* of variables that may involve complex interactions among all of the variable values in the case. While statistical methods can model interactions effects as well, this puts added pressure on the sample size necessary to be confident in one’s results, and statistical studies rarely model interactions among *all* the variables acting together, as a typological theory may do.

Typological theories make less restrictive assumptions about case comparisons than Mill’s methods. Specifically, typological theory assumes that if cases within the same type, or with the same mix of independent variables, have different outcomes on the dependent variable, the difference in the outcome is due to measurement error or left-out variables, not to the type of probabilistic relations theorized in quantum physics. This addresses a common misinterpretation of case study methods, namely, that they assume or require restrictive forms of determinism (Lieberson 1992). It is certainly true that all forms of case comparison are much stronger sources of inference when a variable is a necessary or sufficient condition for a particular outcome. But it is also true that some forms of case comparison require more deterministic assumptions than others, and most case study researchers appear to assume that equifinality is a common condition in social life (Ragin 1987). The minimal assumptions of typological theory are in fact similar to those of the statistical researchers who interpret the “error term” in their equations as including measurement error or left-out variables.¹⁴ This assumption sets aside a third possibility, which can never be definitively ruled out, namely, that the error term can also represent a fundamentally stochastic element analogous to the irreducible probabilism of quantum mechanics.

Comparative Advantages of Case Study Methods

Case study methods have considerable comparative advantages relative to statistical methods or formal models (Collier 1993). These include the operationalization and measurement of qualitative variables (construct validity), the heuristic identification of new variables or hypotheses, the examination of potential causal mechanisms within particular cases or contexts, the historical explanation of cases, and the incorporation of complex relations like equifinality and path dependency into typological theories.

Construct Validity

One of the greatest strengths of case studies is the opportunity to achieve high levels of construct validity, or the ability to measure in a case the indicators that best represent the theoretical concept we intend to measure. Many of the variables of interest to researchers, such as democracy, power, and political culture, are notoriously difficult to operationalize and measure. What constitutes a “democratic” procedure in one cultural context might be profoundly undemocratic in another. Thus, it is important to carry out “contextualized comparison,” that is, comparison that “self-consciously seeks to address the issue of equivalence by searching for *analytically equivalent* phenomena—even if expressed in substantively different terms—across different contexts” (Lock and Thelen 1998, 11). This requires detailed consideration of contextual variables, which is extremely difficult to carry out in statistical studies but common in case studies. Whereas statistical studies run the risk of “conceptual stretching” if they lump together dissimilar cases to get a higher sample size (Sartori 1970), case studies move in the opposite direction, refining concepts with a higher level of validity but doing so at the cost of producing generalizations applicable only over a smaller number of cases. Put in other terms, there is a trade-off between achieving a high level of construct validity, which is easier to do in case studies, and establishing a high level of external validity, or the ability to apply findings across a wide population of cases, which statistical studies are better suited to doing.¹⁵

Because case studies can achieve high construct validity, statistical research is not only usefully preceded by case study research to identify relevant variables, it is often followed by case study work that focuses on deviant cases and further refines concepts (Collier 1998). For example, after

a range of statistical studies suggested that democracies do not fight other democracies, case study researchers started to explore which aspects of democracy—democratic values, democratic institutions, the transparency of decision making in democracies, and so on—might be responsible for this apparent “inter-democratic peace” (George and Bennett forthcoming). Should these case studies indicate, to take a hypothetical example, that a free press and transparency are more important factors than competitive elections in producing an interdemocratic peace, then statistical databases that weighted competitive elections heavily in the definition of democracy will have to be redone, and new statistical tests performed.

Generating New Theories

Case studies can also heuristically identify new variables and hypotheses. This can take place through the study of deviant cases, as noted earlier, but it also happens in the ordinary course of fieldwork, such as archival research and interviews with participants, area experts, and historians. The popular refrain that observations are theory-laden does not mean that they are theory-determined. When a case study researcher asks a participant, “Were you thinking x when you did y ,” and they get the answer, “No, I was thinking z ,” they may have a new variable demanding to be heard. Statistical methods lack any counterpart for this process; some methods of “data mining” or “exploratory data analysis” can be used to identify potentially relevant variables, but even these methods can use only data that is already coded into data sets, or data that someone has already identified as sufficiently useful to be worth coding. Statistical studies that do not involve archival work or interviews to measure or code variables have no inductive means of identifying new variables, although deductive theorizing, whether by a researcher using statistical methods or a formal modeler, can also identify new variables.¹⁶

Making Inferences Regarding Causal Mechanisms

Case studies can use process tracing to examine in detail the observable implications of hypothesized causal mechanisms in individual cases. Causal mechanisms can be defined as the ultimately unobservable entities or structures that operate in specific contexts to generate the phenomena that we observe in the physical or social world.¹⁷

Thus, as the philosopher David Hume famously argued, we cannot directly observe theories or causal mechanisms. As noted earlier, process tracing, like all methods (even experimental ones), does not allow direct or infallible assessment of causal mechanisms as there is always the danger of measurement error, specification error, and omitted variables. Hume also noted, however, that we have several *sources of inference* on the operation of hypothesized causal entities, so that our inferences on underlying causal mechanisms, while fallible, are not mere guesswork. Some of the sources of inference that Hume pointed to—constant conjunction and congruity (similarity in size)—relate to statistical methods, but others—temporal succession and contiguity—relate more directly to process tracing.¹⁸ The detailed tracing of sequential processes among spatially and/or temporally contiguous entities in a single case is a fundamentally different source of inference from the assessment of correlations among cases. Process tracing involves examining the hypothesized causal sequences that a theory and its associated causal mechanisms predict should have taken place in a case, then determining whether the intervening variables along these pathways, or those predicted by alternative explanations, were in fact extant in the case. This provides a basis for inference on whether the hypothesized explanation can or cannot be ruled out as a historical explanation for the case, which in turn allows inferences on the more general scope conditions of the theories under investigation (a theory that fails to explain a “most likely case,” for example, is strongly impugned).

It is the demand for a high level of detail and continuity in explaining an individual historical case that distinguishes process tracing from statistical analysis. As noted previously, the finding that 98 of 100 dominoes have knocked one another over in sequence is not enough to establish that the 99th domino caused the 100th to fall. In this regard, process tracing is quite different from the “manipulation account” of causal inference, in which the value of one variable is manipulated in a controlled experiment to provide a basis for causal inference. Process tracing is useful primarily in nonexperimental studies of historical cases, where controlled experiments are impossible. It can still be useful even in experimental settings, however, as a check on possible sources of error or failure to fully control all of the differences between two trials of an experiment. It can also be useful as a supplement to statistical studies of nonexperimental data by providing a check on possibly spurious interferences and giving evidence on causal direction, or helping to discern which of two correlated variables appears to

be causing the other by temporally preceding it. By combining deductive inquiry—what should I expect to see in the detailed processes in a case if a theory is true?—and inductive inquiry—how might I explain the unanticipated sequences or processes that I find in the case?—process tracing is a powerful source of inference. The inductively derived insights that arise in a case can be distinguished from mere storytelling if they can be explained by extant theories or if they lead to additional novel predictions about the processes in the case or in other cases that are then empirically verified.

To take one example from the medical sciences, scientists have been confident for many years on the basis of statistical analysis of nonexperimental data in humans and experimental data in animal studies that smoking cigarettes increases the likelihood of contracting lung cancer. But this data did not provide much insight into the microlevel causal mechanisms that linked the act of smoking to the outcome of cancer. Nor did it offer explanations of individual cases of lung cancer, as nonsmokers can contract lung cancer as well. Only recently has an improved understanding of cellular-level mechanisms begun to fill in the missing linkages between smoking and cancer. This knowledge has been fostered by pathology studies of individual cases of both human and animal subjects, analogous to process tracing, and of how healthy and cancerous cells and organs changed over time. This improved understanding may eventually improve our ability to predict which individuals are most likely to contract cancer if they smoke, who may be at low risk of cancer despite smoking, and who may be at high risk of cancer despite not smoking. An individual who defied the odds in either direction would be a prime candidate for closer pathology studies (process tracing) that might lead to new insights about the underlying mechanisms.

Historical Explanation of Cases

Conversely, not only can we use a case study to explore causal mechanisms, we can use causal mechanisms to give historical explanations of cases. Historical explanation is quite different from the development and testing of variable-centered theories based on the statistical study of a large number of cases. In historical explanation, the researcher uses theories at each step of a historical process to show how the variables made subsequent steps and the ultimate outcome likely under the historical circumstances of the case (Roberts 1996). This is quite different from establishing statistical gener-

alizations. As statistical researchers readily acknowledge, correlation does not imply causality, and a statistically significant correlation does not necessarily “explain” any or all of the cases upon which it is based. It is not enough to know, for example, that an individual fitting specified relationships to a murder victim is more likely than most to have committed the murder. The prosecutor needs to establish empirically that means, motive, and opportunity existed in this particular case. Ideally, they need a complete and uninterrupted chain of evidence, using forensic, psychological, and other theories to bolster each point in the chain, establishing how the crime was likely have been done by the accused, together with evidence and theoretical explanations that help rule out other likely suspects. Process tracing allows this kind of analysis in individual cases.

Process tracing is thus similar in some respects to standard techniques of writing diplomatic or political history, and there has been an active and growing dialogue between case study researchers and historians.¹⁹ Historians often use theories implicitly to explain rather than merely describe events, and they frequently generalize, though usually only to limited domains of time and space. Yet the purposes, methods, and writings of historians and political scientists remain quite different. As Jack Levy argues, historians seek to understand single unique events, the *milieu et moment*, while political scientists aim to generalize about classes of events; historians tend to favor complex explanations, while political scientists aim for elegant and parsimonious explanations. Historians construct narrative-based explanations; political scientists construct theory-based explanations. Political scientists are explicit about their theoretical assumptions and causal argument; historians are more implicit.²⁰

Levy notes that these distinctions are best understood as lying on a continuum, and case study methods are closer to the writing of history than are other political science methods. Yet the difference remains that case study researchers in political science are interested in the theory-based explanation of individual cases for the purposes of generalizing to other cases, while for historians the explanation of individual cases is a primary goal in itself.

Addressing Complex Causal Relations

A final advantage of case studies is their ability to accommodate complex causal relations such as equifinality, complex interactions effects, and path

dependency (Ragin 1987).²¹ If equifinality holds and there are several paths or combinations that can lead to the same outcome, a typological theory can provide contingent generalizations on each path or combination, and case studies can examine the processes of each. Similarly, by treating cases as configurations of variables, rather than seeking partial correlations among specified variables, case studies can capture complex interactions effects and model path-dependent relationships. The ability to address complexity comes at a price, however, as the more contingent and fine-grained a typological theory, the less parsimonious it becomes and the fewer the cases to which it applies.

Limitations and Trade-offs in the Use of Case Studies

Case Selection Biases and Confirmation Biases

One of the most common critiques of case study methods is that they are prone to “selection bias” (Achen and Snidal 1989; Geddes 1990). Selection bias, in statistical terminology, occurs “when some form of selection process in either the design of the study or the real-world phenomena under investigation results in inferences that suffer from systematic error” (Collier and Mahoney 1996, 59). Such biases can occur when the researcher selects cases that represent a truncated sample along the dependent variable of the relevant universe of cases (Collier and Mahoney 1996, 60; King, Keohane, and Verba 1994, 128–32). In statistical research, the standard presentation of selection bias suggests that a truncated sample typically understates the strength of the relationship between the independent and dependent variables. In other words, it reduces the magnitude of the estimated beta coefficients; Huth and Allee note that in some instances selection biases can also reverse the sign of the coefficients in statistical studies (see chap. 9, this vol.). This is why statistical researchers are recurrently admonished not to select cases on the dependent variable (Collier and Mahoney 1996, 60).

Practitioners and analysts of case study methods, however, have argued that cases selected on the dependent variable can test whether a variable is necessary for the selected outcome (Dion 1997; Collier 1995; Goertz and Starr 2003, 30). If a variable hypothesized to be necessary for a specified outcome can be shown to have been absent in even a single case in which the outcome occurred, then this case can disprove the claim that the vari-

able is a necessary condition for the outcome. In addition, in the early stages of a research program, selection on the dependent variable can serve the heuristic purpose of identifying the potential causal paths and variables leading to that dependent variable. Later, when this first stage of research has clarified the causal model, this model can be tested against cases in which there is variation on the dependent variable.²² Of course, ideally, researchers would have the functional equivalent of a controlled experiment, with controlled variation in independent variables and resulting variation in dependent variables. However, the requisite cases for such research designs seldom exist.²³

Statistical views of the problem of selection bias also understate the most severe and the most common kinds of selection biases in qualitative research. The potential case study selection bias with the most damaging consequences arises from a form of confirmation bias: selecting only those cases whose independent *and* dependent variables vary as the favored hypothesis suggests and ignoring cases that appear to contradict the theory. This type of selection bias can occur even when the traditional warnings against selection bias have not been violated; that is, even when there is variation on both independent and dependent variables, and even when this variation covers the full range of values that these variables can assume. Rather than understating the relationship between independent and dependent variables, this selection bias can understate or overstate the relationship, and it is particularly misleading when the results are overgeneralized to wider populations (Collier and Mahoney 1996, 71–72). Thus, researchers need to be extremely careful in generalizing results from case study designs that include only “present-present” and “absent-absent” cases; that is, they should look hard for similar cases in which the independent variable of interest is present but the predicted effect is absent and for those in which the independent variable is absent but the dependent variable is present.

While this is the most dangerous kind of selection bias, it is also usually easy to identify and avoid. Several other potential biases are more common in case study selection. These include selection of cases based on extreme values of the variables, on the availability of evidence, or on cases’ “intrinsic” historical importance. Each of these criteria for case selection has value for some research goals. Looking at cases with extreme values on the variables, for example, can allow studying particular causal mechanisms in especially stark or obvious forms (Van Evera 1997, 42–49). However, there

is also a risk in emphasizing these criteria to the exclusion of other standards. Selection of cases based on extreme values may lead to overgeneralization if researchers are not vigilant in reminding others (and themselves) that they are working on an extremely truncated sample (Collier and Mahoney 1996, 71). Selection of historically “important” or easily researched cases is less useful for theory building than the selection of cases that are likely to be the most theoretically informative such as deviant, most likely, or most similar cases.²⁴

In addition to contributing to case selection biases, confirmation biases can affect the selection and interpretation of evidence within cases. This can lead to competing or contradictory interpretations by different researchers studying the same case. It is important to guard against this problem by explicitly considering a wide range of alternative explanations for a case and doing systematic process-tracing on these alternatives. Also, whenever researchers modify a historical explanation to better fit a case, they should endeavor wherever possible to find some novel facts that the new explanation also fits and to place more confidence in modifications that do lead to new and empirically verified facts.²⁵

Potential Indeterminacy

Particular case studies may suffer from indeterminacy, or an inability to exclude all but one explanation of a case on the basis of the available process-tracing evidence from that case (Njolstad 1990). When this occurs, it may still be possible to narrow the number of plausible explanations, and it is also important to indicate as clearly as possible the extent to which the remaining hypotheses appear to be complementary, competing, or incommensurate in explaining the case.

One version of the problem of indeterminacy has been widely misapplied to case study methods. This is the “degrees of freedom” problem, which is one kind of indeterminacy that can afflict statistical studies. The degrees of freedom problem arises in statistical work when there are more independent variables than cases, so that it becomes impossible to find coefficient estimates for the variables. Thus, when a researcher has many independent variables but only one or a few observations on the dependent variable, the research design is indeterminate. Some analysts have thus suggested that case studies inherently suffer from a degrees of freedom problem since they have many variables and few “cases” (Achen and Snidal

1989, 156–57). An important misinterpretation arises on this issue, however, from using definitions of *case*, *variable*, and *observation* that are excessively narrow. Earlier I criticized the definition of a case as a phenomenon in which we report only one measure on any pertinent variable. It is this misguided definition, plus inattention to the potential for process tracing, that leads to the conclusion that case studies suffer from an inherent degrees of freedom problem. In fact, as noted previously, an entity may have many different dimensions or contrast classes rather than providing a “single observation.” An apple, for example, has a certain color, texture, sugar content, flavor, and so on; we might aggregate these into a single index defining a “good” apple, but this is different from capturing the many distinct qualities of the apple. In addition, within a single case there are many possible process-tracing observations along the hypothesized causal paths between independent and dependent variables. A causal path may include many necessary steps, and they may have to occur in a particular order. Defining and observing the steps along the hypothesized causal path can lead to “a plethora of new observable implications for a theory” and circumvent the degrees of freedom problem (King, Keohane, and Verba 1994, 119–20; Campbell 1975). There is still the possibility, noted earlier, that a particular case study will be indeterminate in discerning which of several competing hypotheses apply. This is more appropriately described as an indeterminacy problem rather than a degrees of freedom problem, however, as it is more a matter of how the evidence in a particular case matches up with competing hypotheses than a mechanical issue of the number of cases and the number of variables.

Lack of Representativeness

Statistical methods require a large sample of cases that is representative of and allows inferences about an even wider population of cases. To get a representative sample, such studies often rely on random selection of cases. While useful and necessary in statistical studies, these requirements and practices are inappropriate and counterproductive when extended to case study methods (King, Keohane, and Verba 1994, 124–27).

Case study researchers do not aspire to select cases that are “representative” of large and diverse populations, and they ordinarily cannot make claims that their findings are applicable to such populations, with the partial exception of case studies that show that a theory failed to explain its

most likely case or that disprove purported necessary or sufficient conditions (McKeown 1999). Case study researchers are usually more interested in finding out the conditions under which specified outcomes occur and the mechanisms through which they occur than the frequency with which those conditions and their outcomes arise (George and Bennett forthcoming). Researchers often select cases with the goal of providing the strongest possible inferences on particular theories or of using deviant cases to help identify left-out variables. In either research design, the cases selected are intentionally and necessarily unrepresentative of wider populations, and researchers must be careful to point out that they seek only contingent generalizations that apply to cases that are similar to those under study (George and Smoke 1989; George and Bennett forthcoming). To the extent that there is a “representativeness” problem, it is more accurately presented as a problem of overgeneralization that arises if case study researchers or their readers extend research findings to types of cases unlike those actually studied.²⁶

In this regard, case studies involve a trade-off between generalizability and specificity. Rich generalizations in the social sciences often apply only to small and well-defined populations or subtypes, whereas theories that apply to broader populations are usually not very specific. In part, choices between rich but narrow generalizations and less specific but broadly applicable generalizations depend on aesthetic decisions about the kind of theory one prefers and pragmatic considerations such as whether the theory is to focus on “manipulable variables” that policymakers can change to affect outcomes. Choices between broad or deep theorizing can also reflect theoretical assumptions about the complexity of the world. If the researcher believes that similar causal relations hold for large populations and that there are limited interactions effects, then broad theories may prove fruitful, and they may even be fairly rich as well. If multiple interactions effects are present, on the other hand, then only highly contingent theorizing for small and well-defined subpopulations may be possible.²⁷

Potential Lack of Independence of Cases

Another issue concerns whether cases are independent of one another. Here again, there is a particular statistical version of this problem that does not apply to case studies, and a more fundamental version that does. In a statistical study, if a correlation is the result not of the hypothesized relation-

ship under consideration but of learning or diffusion from one case to the others, then the additional cases do not provide as much new information as if they were fully independent of one another, so in effect the sample size is smaller than if the cases were independent (George 1982, 19–23; King, Keohane, and Verba 1994, 222; and see Huth and Allee, chap. 9, this vol., for a related discussion of this issue in the context of statistical methods). This is sometimes referred to as Galton's problem. In case studies, there is a danger that the researcher will fail to identify a lack of independence between cases, but this danger does not manifest itself as a problem related to the sample size or number of cases studied, and it is not necessarily amplified by the intentional selection of cases based on a preliminary knowledge of their variables (indeed, such intentional selection may be designed specifically to assess the independence of cases or the diffusion processes among them). As Alexander George has argued, the question of whether the independence of cases is a relevant consideration is not a question that can be answered "on a priori grounds; the answer surely depends on the research objectives of a particular study, what theory or hypothesis is being developed, and how the comparison of cases is structured" (1982, 21). As George notes, process tracing can inductively uncover linkages between cases and reduce the dangers of any unanticipated lack of independence of cases. When learning or diffusion processes are anticipated or uncovered and taken into account, they need not undercut the value of studying partially dependent cases. Indeed, only cases that are perfectly dependent provide no additional information (King, Keohane, and Verba 1994, 222). Moreover, as George points out, case study methods can be particularly effective at examining precisely the kinds of path-dependent learning and diffusion processes that give rise to Galton's problem (George 1982, 21).

Other limitations of case study methods require only brief mention. Case studies are better at determining the scope conditions under which variables have an effect than estimating the magnitude of that effect. This latter task of assessing the causal "weight" or causal effect of variables is better performed through statistical studies. Case study researchers also face a trade-off between doing richly detailed studies of a small number of cases versus seeking broader generalizations across a larger number of cases. Often the best approach is for each researcher to focus in detail on a small but well-defined subset of cases or types of cases, while making compar-

isons to existing research in the same research program so that the field as a whole incrementally fills out the typological space.

In sum, critiques of case study methods through the prism of statistical concepts have often misconstrued the strengths and weaknesses of case studies. On the issues of degrees of freedom, “representativeness,” independence of cases, and the use of Mill’s methods, case studies are generally stronger than their critics have suggested. On the question of case selection and selection bias, standard statistical critiques have overstated some methodological problems but understated others. The two most constraining limits of case study methods are the problem of getting a range of cases for study that covers many of the possible causal paths or types and the problem of interpreting outcomes and processes that are consistent with more than one theory. Both of these problems have received less attention because they do not fit as readily into statistical terms (for exceptions see Little 1998; Ragin 1987; Lieberman 1992; Njolstad 1990).

Comparative Strengths and Limitations of Formal Models and Statistical Methods

To underscore the essential complementarity of the leading methods in political science, it is useful to review briefly the comparative advantages of formal models and statistical methods. The comparative advantages of formal models center on their rigorous deductive logic. Deductive logic can be useful in elucidating the dynamics of causal mechanisms, and it can lead to counterintuitive hypotheses that can then be tested. Well-known examples include the literatures on collective action dilemmas, principal-agent relations, problems of credible commitment, two-level games, gatekeeping, veto points, and tipping points. Limitations of formal models include presence of multiple equilibria, the potential for path dependencies, and the possibility of self-denying prophecies (that is, understanding of the model itself can lead to changes in behavior, though this problem is not unique to formal models). Of course, formal modeling is not an empirical method and must be linked to either case studies or statistical studies to provide empirical tests.²⁸

The primary advantages of statistical methods include their ability to estimate the average explanatory effects of variables, their ability to analyze the representativeness or frequency of subsets of the data collected, their

visual display, and the high degree of replicability of studies using the same database. Limitations of standard statistical methods include the challenges they face in identifying new variables, dealing with multiple conjunctural causality or equifinality, devising conceptually valid operationalizations of qualitative variables, and providing or testing historical explanations of individual cases. Some of these limitations may be inherent in statistical methods, while others may involve trade-offs that could ease somewhat with the development of more sophisticated statistical techniques. Notably, these advantages and limitations are almost precisely the converse of the ones associated with case study methods, which are poor at partial correlations and measures of frequency but good at identifying new variables, dealing with complex causal relations, and devising and testing historical explanations.

The Outlook for Increased Multimethod Collaborative Research

As the editors to this volume conclude, the increasingly evident complementary relationship between case studies, statistical methods, and formal modeling has begun to lead toward more multimethod and collaborative research. Because case studies, statistical methods, and formal modeling are all increasingly sophisticated, however, it is difficult for a single researcher to be adept at more than one set of methods while also attaining a cutting-edge theoretical and empirical knowledge of his or her own field. As a result, much multimethod work is collaborative. Encouraging such cooperative efforts will require that political science departments do not discriminate against multiauthored works in their hiring and promotion decisions, as they currently often do. This will raise the problem of evaluating the work of individual contributors to multiauthored works, but in many other fields (particularly the medical and physical sciences) multiauthored works are common. One means of apportioning credit for such works, which is often done in books but could be extended to articles, is to provide a brief footnote that outlines which parts of a work were done primarily by one author or another and which were fully collaborative.

Collaboration can also take place sequentially if researchers work to build on findings generated by those using different methods. For example, statistical analysis might identify outliers or deviant cases, and case studies can investigate why these cases are deviant (Ness 1985). Case studies can also look at the “average” or “representative” cases identified in sta-

tistical studies to test and refine the hypothesized causal mechanisms behind the correlations or patterns observed and provide a check on whether a correlation is spurious and on potential endogeneity. Statistical studies, in turn, can assess the general applicability of causal mechanisms uncovered by case studies, and statistical studies might identify strong patterns that can be used to structure the study of individual cases. Similarly, proposed formal models can be tested in case studies to see if their hypothesized causal mechanisms were in fact in operation. This eschews the “as if” assumption made by some formal modelers, namely, the argument that actors did not need to actually go through the posited decision process as long as the model correctly predicts outcomes “as if” actors went through the hypothesized processes. Such “as if” assumptions are inconsistent with causal explanations that refer to causal mechanisms, as these explanations are open to challenge if they can be shown to be inconsistent with processes observed at a lower level of analysis or in finer detail.²⁹ Case studies can also inductively identify variables and theories that can then be formalized in models.

Both kinds of collaboration require that even as they become expert in one methodological approach, scholars must also become aware of the strengths and limits of other methods and capable of an informed reading of their substantive results. If proficiency in the cutting-edge techniques of all three methods is an unrealistic goal for most mortals, an achievable goal would be proficiency as a producer in one method and an informed reader capable of using and critiquing research using the other two methods. This requires that graduate curricula offer sequences of courses leading up to the highest levels of current professional practice in all three methods and that departments require a reading proficiency in all three methods.³⁰

Conclusions

At a high level of generality, the successors of the positivist tradition who employ case study methods, statistical methods, and formal models share an epistemological logic of inference. They all agree on the importance of testing theories empirically, generating an inclusive list of alternative explanations and their observable implications, and specifying what evidence might infirm or affirm a theory. On the methodological level, however, what is useful or necessary for one method, such as random selection of cases in a statistical study, may be unnecessary or even counterproduc-

tive in another, such as case studies. This creates an obligation for researchers to learn how to translate between the various methods and to understand their respective strengths and limitations. More important, it creates an opportunity that has not yet been fully or efficiently realized. The comparative advantages of the respective methods are profoundly different, but this allows the strengths of one method to compensate for the weaknesses of another.

Recommended Readings

- Brady, H., and D. Collier. 2003. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Berkeley, CA: Berkeley Institute of Governmental Studies Press; and Latham, MD: Rowman and Littlefield.
- George, A. L., and A. Bennett. Forthcoming. *Case Studies and Theory Development*. Cambridge: MIT Press.
- Little, D. 1998. *Microfoundations, Method, and Causation: On the Philosophy of the Social Sciences*. New Brunswick, NJ: Transaction.
- Mahoney, J., and D. Rueschemeyer, eds. 2003. *Comparative Historical Analysis in the Social Sciences*. Cambridge: Cambridge University Press.
- Ragin, C. 2000. *Fuzzy Set Social Science*. Chicago: University of Chicago Press.
- Van Evera, S. 1997. *Guide to Methods for Students of Political Science*. Ithaca: Cornell University Press.

Notes

1. In statistical terms, this problem of “too many variables, too few observations” is known as the “degrees of freedom problem.”
2. For further discussion of the issues of defining “what is this event a case of?” and “given this phenomenon, is this event a case of it?” see Ragin and Becker (1992).
3. Strictly speaking, congruence testing may involve explicit or implicit comparisons to other cases, and counterfactual analysis involves comparison to a hypothetical case, so process tracing is the only method that is purely within-case.
4. This is also a standard practice in the physical sciences. For example, we might use plant records to develop a theory that a large meteorite caused a case of mass dinosaur extinction, and then test this theory against other geological evidence on the possible impact of a large meteor at the time in question (King, Keohane, and Verba 1994). Similarly, we might develop a historical explanation of a

political process from available evidence in a case and then predict what evidence we should find from archives once they are opened (Wohlforth 1998).

5. Bennett (1999). I choose my own work in part because it is difficult to evaluate fully the research design decisions made by others on substantive issues in which one is not an expert.

6. The book's chapter on learning theory includes a typological theory on the different paths that learning can take depending on such factors as the strength of the state and of civil society (Bennett 1999, 108–12).

7. For excellent advice on carrying out these kinds of field research, see Thies (2002) and Murphy (1980).

8. Rogowski (1995) gives several examples, including Lijphart (1975).

9. For a similar view, see Van Evera (1997). In practice, few cases pose such clearly decisive tests, but it is still important for researchers to indicate the severity of the test a case poses for a theory.

10. Eckstein also suggests that a “plausibility probe” might be undertaken to give a preliminary estimate of a theory's explanation of a case before a more intensive and costly study is undertaken. This should not be misinterpreted, however, as a means of “lowering the bar” for a new hypothesis; rather, it is an opportunity to judge if further study is warranted and to adapt the theoretical framework for a more systematic test against additional evidence from within the case.

11. Another of Mill's methods, “concomitant variation,” relies upon observed covariations in the strength of variables. In Mill's example of this method, one might observe the covariation between the level of the tides and the phases of the moon and assume that there is some causal connection. Concomitant variation is thus related to the statistical logic of partial correlations, and, like Mill's other methods, it is vulnerable to spurious inferences unless restrictive conditions are satisfied.

12. A more flexible variant of Mill's methods is Charles Ragin's method of qualitative comparative analysis (QCA) (1987). This method relies on Boolean algebra to make pairwise comparisons of cases, or types of cases, and relaxes some of the assumptions necessary for the direct use of Mill's methods. QCA allows for the possibility of equifinality, a key advance over Mill's methods, but QCA still requires sufficiency at the level of conjunctions of variables to reach definitive results, and it requires the inclusion of all causally relevant variables to prevent spurious inferences. In addition, the results of QCA are unstable in that adding a single new case or changing the coding of one variable can radically change the results of the analysis (Goldthorpe 1997, 20 nn. 8, 9). For these reasons, Ragin warns against the “mechanical” use of QCA for causal inference (1987, 98), and his later work on fuzzy logic is in some sense an effort to relax the requirements of QCA. In short, with QCA, as with Mill's methods, it is necessary to supplement case comparisons with process tracing of cases in order to relax the restrictive and

unrealistic assumptions necessary for definitive results from comparisons alone (Rueschemeyer and Stephens 1997, 60–61).

13. For an example of using process tracing to rule out residual differences between cases as being causal, see James Lee Ray (1995), *Democracies and International Conflict*, 158–200.

14. King, Keohane, and Verba (1994, 59 n. 12) suggest that most statisticians use deterministic working assumptions in this regard.

15. Similarly, Paul Huth and Todd Allee (chap. 9, this vol.) note the trade-off between careful measurement of variables in a small number of cases versus less precise measurement over a large number of cases. Ronald Mitchell and Thomas Bernauer (chap. 4, this vol.) addresses the closely related issue of internal validity, or the ability to discern that an observed correlation between variables in a case is in fact causal (in part by achieving high construct validity, reducing measurement error, and excluding alternative explanations). They note that “internal validity is a precondition for external validity” and argue that case studies can achieve high internal validity through process tracing. They also concur, however, that selecting cases to control for certain variables limits the range of cases to which one can generalize, except to the extent that the results of “hard cases” or tough tests, such as the failure of a theory to fit its most likely case or the ability of a theory to explain even a least likely case, can be generalized.

16. Questionnaires with open-ended questions may also turn up new variables, and scholars critiquing a piece of statistical research may suggest possible omitted variables as well. There are thus some means of identifying omitted variables in statistical research, though they are quite limited in studies that use only preexisting databases.

17. On this and other definitions of causal mechanisms, see James Mahoney (2001).

18. On Hume and sources of causal inference, see Marini and Singer (1988).

19. Elman and Elman (2001).

20. Jack Levy (2001, 40).

21. This advantage may be relative rather than absolute. Statistical methods can model several kinds of interactions effects, although they can do so only at the cost of requiring a larger sample size, and models of nonlinear interactions rapidly become complex and difficult to interpret. It is possible that new statistical methods may be able to improve upon the statistical treatment of equifinality and interactions effects, and at least narrow the gap in the treatment of this issue. (Braumoeller 2002)

22. Case study researchers in many instances should make comparisons between the subset of cases or types studied and the larger population, where there is more variance on the dependent variable (Collier and Mahoney 1996, 63). Some-

times, such comparisons can be made to existing case studies in the literature, or the researcher might include “mini” case studies, or less in-depth studies, of a wide number of cases in addition to full studies of the cases of greatest interest. To say that such comparisons are often useful for many research goals, however, is very different from arguing that they are always necessary for all research goals.

23. A related issue is whether foreknowledge of the values of variables in cases, and perhaps researchers’ cognitive biases in favor of particular hypotheses, necessarily slant the selection of case studies. However, selection with some preliminary knowledge of cases allows much stronger research designs, as cases can be selected with a view toward whether they are most likely, least likely, or crucial. Selecting cases in this way can strengthen, rather than undermine, the severity of the process-tracing test of a theory. Also, within-case analysis often leads to the finding that the researcher’s (or the literature’s) preliminary knowledge of the values of the independent and dependent variables was incomplete or simply wrong, and case study researchers sometimes conclude that none of the proposed theories is an adequate explanation of a case (Campbell 1975). In addition, intentional selection of cases can benefit from knowledge of the findings of existing studies, and it can be guided by estimations of whether the theories of interest are strong and previously tested or new and relatively weak (Laitin 1995, 456).

24. Van Evera in fact offers many criteria for selecting the most theoretically informative cases, including cases with large within-case variance, cases about which competing theories make opposite and unique predictions, cases that are well-matched for controlled comparisons, outlier cases, and cases whose results can be replicated. These criteria present fewer complications than those of extreme values, data availability, and intrinsic importance.

25. This is the methodological standard emphasized by Lakatos (1970).

26. In some instances, critiques of particular case studies have overstated the problems of representativeness and selection bias by assuming that these studies have purported to offer generalizations that cover broad populations, whereas in fact these studies carefully circumscribed their claims to apply them only to cases similar in well-specified respects to those studied. Collier and Mahoney (1996, 80–87) make this critique of Barbara Geddes’s (1990) review of case studies and selection bias.

27. This is similar to the question of whether the “unit homogeneity” assumption is theorized to be applicable to a large and diverse population or only to small and well-defined populations with regard to the processes under study. Two units are defined as homogenous whenever “the expected values of the dependent variables from each unit are the same when our explanatory variable takes on a particular value” (King, Keohane, and Verba 1994, 91).

28. Additional strengths and limitations attend to that subset of formal mod-

els that focus on rational choice theories. There is a tendency in the literature to conflate formal and rational choice models, but it is possible to have formal models based on cognitive processes other than rational ones, such as prospect theory.

29. See, for example, Bates, Greif, Levi, Rosenthal, and Weingast (1998).

30. A reading proficiency does not require equal numbers of courses in each method; one course each in case study methods and formal modeling might be sufficient, while several courses in statistics may be necessary. According to a survey of graduate curricula that I have undertaken, however, many top departments require one or more courses in statistics, but courses in qualitative methods and formal modeling are seldom required and sometimes not even offered (Bennett, Barth, and Rutherford 2003). In part to address this imbalance, the interuniversity Consortium on Qualitative Research Methods now sponsors a two-week training institute in these methods at Arizona State University, analogous to the University of Michigan's summer institute in statistical methods (see <http://www.asu.edu/clas/polisci/cqrm/>), 2003.

References

- Achen, C., and D. Snidal. 1989. Rational Deterrence Theory and Comparative Case Studies. *World Politics* 41 (2): 143–69.
- Bates, R., A. Greif, M. Levi, J. Rosenthal, and B. Weingast. 1998. *Analytic Narratives*. Princeton: Princeton University Press.
- Bennett, A. 1999. *Condemned to Repetition? The Rise, Fall, and Reprise of Soviet-Russian Military Interventionism, 1973–1996*. Boston: MIT Press.
- Bennett, A., A. Barth, and K. Rutherford. 2003. Do We Preach What We Practice? A Survey of Methods in Political Science Journals and Curricula. PS: *Political Science and Politics* 36 (3): 373–78.
- Braumoeller, B. F. 2002. Causal Complexity and the Study of Politics. Manuscript, Harvard University.
- Campbell, D. T. 1975. “Degrees of Freedom” and the Case Study. *Comparative Political Studies* 8 (2): 178–94.
- Collier, D. 1993. *The Comparative Method. in Political Science: The State of the Discipline II*. Edited by A. W. Finifter. Washington, DC: APSA.
- . 1995. Translating Quantitative Methods for Qualitative Researchers: The Case of Selection Bias. *American Political Science Review* 89:461–66.
- . 1998. Comparative Method in the 1990s. APSA-CP Newsletter 9 (1).
- Collier, D., and J. Mahoney. 1996. Insights and Pitfalls: Selection Bias in Qualitative Research. *World Politics* 49 (1): 56–91.
- Dion, D. 2003. Evidence and Inference in the Comparative Case Study. In *Necessary Conditions: Theory, Methodology, and Applications*, edited by G. Goertz and H.

- Starr, 95–112. Boulder: Rowman and Littlefield.
- Eckstein, H. 1975. Case Study and Theory in Political Science. In *Handbook of Political Science*, vol. 7, *Strategies of Inquiry*, edited by F. Greenstein and N. Polsby. Reading, MA: Addison-Wesley Press.
- Elman, C., and M. F. Elman, eds. 2001. *Bridges and Boundaries: Historians, Political Scientists, and the Study of International Relations*. Cambridge: MIT Press.
- Fearon, J. D. 1991. Counterfactuals and Hypothesis Testing in Political Science. *World Politics* 43 (2): 169–95.
- Geddes, B. 1990. How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics. *Political Analysis* 2:31–50.
- George, A. L. 1979a. Case Studies and Theory Development: The Method of Structured, Focused Comparison. In *Diplomacy: New Approaches in History, Theory, and Policy*, edited by P. G. Lauren. New York: Free Press.
- . 1979b. The Causal Nexus between Cognitive Beliefs and Decision-Making Behavior. In *Psychological Models in International Politics*, edited by L. Falkowski. Boulder: Westview.
- . 1982. Case Studies and Theory Development. Manuscript, Stanford University.
- George, A. L., and A. Bennett. Forthcoming. *Case Studies and Theory Development*. Cambridge: MIT Press.
- George, A. L., and T. J. McKeown. 1985. Case Studies and Theories of Organizational Decision Making. In *Advances in Information Processing in Organizations*, edited by R. Coulam and R. Smith. Greenwich, CT: JAI Press.
- George, A. L., and R. Smoke. 1989. Deterrence and Foreign Policy. *World Politics* 41 (2): 170–82.
- Goertz, G., and H. Starr, eds. 2003. *Necessary Conditions: Theory, Methodology, and Applications*. Lanham, MD: Rowman and Littlefield.
- Goldthorpe, J. 1997. Current Issues in Comparative Macrosociology. *Comparative Social Research* 16:1–26.
- King, G., R. O. Keohane, and S. Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Laitin, D. D. 1995. Disciplining Political Science. *American Political Science Review* 89 (2): 454–56.
- Lakatos, I. 1970. Falsification and the Methodology of Research Programmes. In *Criticism and the Growth of Knowledge*, edited by I. Lakatos and A. Musgrave. Cambridge: Cambridge University Press.
- Levy, J. S. 2001. Explaining Events and Developing Theories: History, Political Science, and the Analysis of International Relations. In *Bridges and Boundaries*, edited by C. Elman and M. F. Elman. Cambridge: MIT Press.
- Lieberson, S. 1992. Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases. In *What Is*

- a Case? Exploring the Foundations of Social Inquiry*, edited by C. Ragin and H. Becker. Cambridge: Cambridge University Press.
- Lijphart, A. 1971. Comparative Politics and the Comparative Method. *American Political Science Review* 65 (2): 682–93.
- . 1975. *The Politics of Accommodation: Pluralism and Democracy in the Netherlands*. Berkeley: University of California Press.
- Little, D. 1998. *Microfoundations, Method, and Causation*. New Brunswick, NJ: Transaction.
- Lock, R., and K. Thelen. 1998. Problems of Equivalence in Comparative Politics: Apples and Oranges Again. APSA-CP: 9–15.
- Mahoney, J. 1999. Nominal, Ordinal, and Narrative Appraisal in Macro-Causal Analysis. *American Journal of Sociology* (3): 1154–96.
- . 2001. Beyond Correlational Analysis: Recent Innovations in Theory and Method. *Sociological Forum* 16 (3): 575–93.
- Marini, M., and B. Singer. 1988. Causality in Social Sciences. In *Sociological Methodology 1988*, edited by C. Clogg. Washington, DC: American Sociological Science Association.
- McKeown, T. J. 1999. Case Studies and the Statistical World View. *International Organization* 53 (1): 161–90.
- Murphy, J. 1980. *Getting the Facts: A Fieldwork Guide for Evaluators and Policy Analysts*. Santa Monica: Goodyear.
- Ness, G. 1985. Managing Not-So-Small Numbers: Between Comparative and Statistical Methods. *International Journal of Comparative Sociology* 26:1–13.
- Njolstad, O. 1990. Learning from History? Case Studies and the Limits to Theory-Building. In *Arms Races: Technological and Political Dynamics*, edited by O. Njolstad. Los Angeles: Sage.
- Przeworski, A., and H. Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley.
- Ragin, C. 1987. *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- Ragin, C., and H. Becker. 1992. Introduction. In *What Is a Case? Exploring the Foundations of Social Inquiry*, edited by C. Ragin and H. Becker. Cambridge: Cambridge University Press.
- Ray, J. L. 1995. *Democracy and International Conflict: An Evaluation of the Democratic Peace Proposition*. Columbia: University of South Carolina Press.
- Roberts, C. 1996. *The Logic of Historical Explanation*. University Park: Pennsylvania State University Press.
- Rogowski, R. 1995. The Role of Theory and Anomaly in Social-Scientific Inference. *American Political Science Review* 2:467–70.
- Rueschemeyer, D., and J. Stephens. 1997. Comparing Historical Sequences: A Powerful Tool for Causal Analysis. *Comparative Social Research* 16:55–72.

- Sartori, G. 1970. Concept Misformation in Comparative Politics. *American Political Science Review* 64:1033–53.
- Tetlock, P. E., and A. Belkin, eds. 1996. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. Princeton: Princeton University Press.
- Thies, C. 2002. A Pragmatic Guide to Qualitative Historical Analysis in the Study of International Relations. *International Studies Perspectives* 3 (4): 351–72.
- Van Evera, S. 1997. *Guide to Methods for Students of Political Science*. Ithaca: Cornell University Press.
- Wohlforth, W. 1998. Reality Check: Revising Theories of International Politics in Response to the End of the Cold War. *World Politics* 50 (4): 650–80.