

Review: Case Studies and the Statistical Worldview: Review of King, Keohane, and Verba's
Designing Social Inquiry: Scientific Inference in Qualitative Research
Reviewed Work(s): Designing Social Inquiry: Scientific Inference in Qualitative Research by
Gary King, Robert O. Keohane and Sydney Verba

Review by: Timothy J. McKeown

Source: *International Organization*, Vol. 53, No. 1 (Winter, 1999), pp. 161-190

Published by: The MIT Press

Stable URL: <http://www.jstor.org/stable/2601375>

Accessed: 08-03-2017 17:02 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://about.jstor.org/terms>



The MIT Press is collaborating with JSTOR to digitize, preserve and extend access to *International Organization*

Case Studies and the Statistical Worldview: Review of King, Keohane, and Verba's *Designing Social Inquiry: Scientific Inference in Qualitative Research*

Timothy J. McKeown

Introduction

Is there a single logic of explanation common to all empirical social scientific research? Is that logic a statistical one? Gary King, Robert Keohane, and Sidney Verba's *Designing Social Inquiry* answers "yes" to both questions. Although their book seems to be oriented primarily to the practical problems of research design in domains that have traditionally been the province of statistical analysis, its subtitle—"Scientific Inference in Qualitative Research"—reveals a much larger agenda. King, Keohane, and Verba assume at the outset that qualitative research faces the same problems of causal inference as quantitative research; that assumption, in turn, forms the basis for analyzing causal inference problems in qualitative research as if they were problems of classical statistical inference in a certain type of quantitative research. The solutions to the problems of qualitative research are therefore deemed to be highly similar to those in quantitative research. Although this is not an entirely new position—Paul Lazarsfeld and Morris Rosenberg outlined a similar position more than forty years ago—the treatment by King, Keohane, and Verba is the most extensive and theoretically self-conscious exposition of such a view.¹

I discuss the nature and implications of that assumption. I argue that it is problematic in ways that are not discussed by King, Keohane, and Verba, and that it is an

Special thanks to Janice Stein and Alex George for encouraging me to return to this subject. I received helpful comments on earlier versions of this article at two seminars hosted by the Center for International Security and Arms Control, Stanford University, in 1996, and at a presentation to Bob Keohane's graduate seminar at Duke University in 1997. My thanks to Lynn Eden and Alex George for inviting me to CISAC and providing an extraordinary opportunity to discuss the issues in this article in great detail, and to Bob Keohane for being so magnanimous and helpful. Thanks also to seminar participants for their probing questions and comments. I also received helpful comments, in many cases at great length, from the following individuals: Hayward Alker, Aaron Belkin, Andrew Bennett, David Collier, David Dessler, George Downs, James Fearon, Ronald Jepperson, William Keech, Robert Keohane, Catherine Lutz, Mike Munger, Thomas Oatley, Robert Powell, John Stephens, Sidney Tarrow, Isaac Unah, Peter VanDoren, the editors of *IO*, and an anonymous reviewer. I assume full responsibility for whatever errors remain.

1. Lazarsfeld and Rosenberg 1955, 387–91.

error to attempt to squeeze all empirical practice in the social sciences into a particular statistical mold. Because the statistical worldview embodied in King, Keohane, and Verba's assumption is usually not the worldview that animates case studies, their approach leads to a series of misconceptions about the objectives of case studies and their accomplishments. These misconceptions are constructive, however, in the sense that exposing them leads to a clearer notion not only of the underlying logic behind case studies but also of the importance of nonstatistical thinking and research activity in all research domains—even those dominated by classical statistical data analysis.

King, Keohane, and Verba's Philosophy of Science

Although they disclaim any interest in the philosophy of science, King, Keohane, and Verba adopt essentially Popperian positions on many important questions.² In particular, their emphasis on a clear distinction between forming or stating hypotheses and testing them, an accompanying reluctance to treat hypothesis formation as anything other than an art form,³ their stress on the need for simplicity in theories, and their insistence on subsuming each case within a class of cases are all highly consistent with logical positivism or Karl Popper's reworking of it.⁴ The King, Keohane, and Verba project—to delineate a theory of confirmation that specifies a priori rules for using observations to evaluate the truthfulness of hypotheses, regardless of the field of inquiry or the specifics of the hypotheses—is a project not only of Popper but also of logical positivism more broadly understood.⁵

Like Popper, King, Keohane, and Verba accept two departures from strict positivism. First, they treat observations as theory-laden, so the separation of theory and data is more a matter of degree and emphasis than of kind. Second, they argue that parsimony as an end is not very important and can often be abandoned as an objective. However, neither concession has much practical impact on the advice they offer, and they do not face squarely the inconsistencies that arise between their practical advice and their philosophical position.

Does it make any difference that King, Keohane, and Verba are Popperians? How one answers that depends on what one believes should be the evaluation criteria for their argument. If the design and execution of research are best understood as a pragmatic activity heavily informed by the substantive requirements of a particular field, then its philosophical underpinnings might seem unimportant. Research methods could be evaluated in terms of the quality of the research produced when the advice is followed (and the evaluation of quality, in turn, could be based on pragmatic, field-specific grounds). However, this viewpoint creates two problems for King, Keohane, and Verba. One is that the only justification they could then offer for

2. King, Keohane, and Verba 1994, 4.

3. *Ibid.*, 14.

4. Popper 1959.

5. Miller 1987, 162.

their approach is that “it works.” If it could also be shown that another approach “works” or that theirs does not always do so, there would be no basis for privileging King, Keohane, and Verba’s prescriptions, unless somehow we could demarcate a domain in which their prescriptions have a comparative advantage over others. The second problem is that the philosophical side of their argument would be reduced to a ceremonial role. Perhaps that is why, when the pragmatic requirements of doing research conflict with Hempelian notions, the authors are not averse to leaving Hempel behind. That this pragmatic viewpoint is itself the embodiment of a philosophy that is distinctly non-Hempelian does not concern them.

Although it is tempting to adopt such a task-oriented view of proper research methodology in order to move quickly to concrete issues, there are two reasons to resist doing so. First, an argument for a particular way of pursuing research would be more convincing if it could be shown that it were constructed on a firmer philosophical foundation. Second, to the extent that the foundation guides thinking about research methods, and to the extent that the foundation is deficient, the concrete conclusions might also be deficient.

King, Keohane, and Verba and the Popperian View of Theory

King, Keohane, and Verba are partisans of causal analysis. This is a departure from Popper, who was very skeptical of the idea that the mere identification of causal processes is sufficient to warrant the term *theory*. (His views on evolutionary theory, for example, were decidedly negative.⁶) Popper, in common with positivists more generally, wished to dispense with references to causation and restrict discussion to regularities and entailments.⁷ He wished, in other words, to make the relation between theory and observation one of *logic*. King, Keohane, and Verba take for granted that causal laws are readily accommodated within the “covering law” approach of positivism (their examination of this issue was apparently confined to consulting Daniel Little, who approvingly cites Hempel’s claim that causal explanations can be subsumed within such a framework).⁸ Such an explanation cites one or more general if-then propositions (“laws”) relating outcomes to antecedents in a given situation and then establishes that a given observation is of an event or situation specified in (that is, “covered” by) the general laws. However, the claim that causal analysis can be so accommodated is widely doubted, particularly by those who are partisans of causal analysis. Richard Miller, for example, contends that the positivist project of establishing general logical relationships between observations and theories is essentially unworkable and that a causal conception of explanation avoids the problems said to plague the former perspective.⁹

6. Popper 1959.

7. See Popper 1968, 59–62; and Miller 1987, 235.

8. See Little 1991, 18; and Hempel 1965, 300–301.

9. Miller 1987.

Although Miller's full argument is lengthy and complex, a sense of his criticisms can be gleaned from the following excerpt:

Causal explanation is not analyzed by the covering law model. Here, counter-examples really are sufficient. A derivation fitting one of the two basic patterns often fails to explain. When a barometer falls, a change for the worse in the weather is very likely to follow. The high probability is dictated by laws of meteorology. But the weather does not change because the barometer falls. In conjunction with basic and utterly general laws of physics and chemistry, the shift toward the red of spectral lines in spectra from more distant stars entails that the observed universe is expanding. [However, t]he red shift does not explain why the universe is expanding. . . . Because these examples fit the covering law model so well . . . and because the failure to explain is so obvious, they are overwhelming.¹⁰

For those who are aware of such criticisms, the simultaneous commitments of King, Keohane, and Verba to the notion of a general logic of inquiry founded on a covering law approach and to an account of explanation that stresses the role of causal mechanisms thus creates a strong tension that is never confronted, let alone resolved.

Is There More Than One Logic of Research?

Ironically, a powerful argument that there is *not* more than one logic is provided by a prominent critic of positivist methodological dicta. In discussing the research techniques of those who work within a hermeneutic mode of analysis, Paul Diesing corroborates the King, Keohane, and Verba claim that a unified logic of inference exists, at least at the most basic level:

The hermeneutic maxim here is: no knowledge without foreknowledge. That is, we form an expectation about the unknown from what we "know." Our foreknowledge may be mistaken, or partial and misleading, or inapplicable to this text; but in that case the interpretation will run into trouble. . . . Our foreknowledge directs our attention. . . . The passages that answer these questions point in turn to other passages. . . .

We form hypotheses about the meanings of a text based on our prior theory of the text, which in turn has emerged from our own experience. If our hypotheses are disconfirmed, then our prior theory is called into question.

In finding an analogue to external validity in hermeneutic approaches, Diesing sounds remarkably like King, Keohane, and Verba as he discusses how to pursue a qualitative research program:

We can call our foreknowledge into question if it sometimes produces an expected interpretation that cannot make a coherent message out of the text, in context. To question our own foreknowledge, we must first focus on it and become

10. Ibid., 34.

aware of what we are assuming; then we must devise a different assumption, perhaps one suggested by this case, and see whether it produces better hypotheses. This process does not produce absolute truth, but a validity that can be improved within limits.¹¹

Nothing in Diesing's account is inconsistent with the advice that King, Keohane, and Verba offer on how to do research aimed at uncovering and testing propositions about cause-and-effect relationships. Indeed, their most likely response would probably be "we told you so." (They come close to Diesing's position when they note that both science and interpretation rely on "formulating falsifiable hypotheses on the basis of more general theories, and collecting the evidence needed to evaluate these hypotheses."¹²) Since the epistemological "distance" between hermeneutics and the statistical analyses of survey research responses might be supposed to be large, we have powerful support for an important part of King, Keohane, and Verba's analysis in a place where we might least expect to find it, by an author who is notably unsympathetic to the King, Keohane, and Verba project. Diesing himself is not averse to these conclusions, suggesting that hermeneutic approaches are compatible with Popper's "conjectures and refutations" description of scientific activity.¹³ Inspecting what we know about the world in order to draw some tentative conclusions about the processes that govern that world and then examining how well those conclusions account for existing or newly acquired knowledge are fundamental to empirical research. What is less clear is whether this activity is always governed by the statistical logic proposed by King, Keohane, and Verba.

Is Inference Fundamentally "Statistical"?

The best description of how King, Keohane, and Verba view qualitative research is that it is "prestatistical" (my term): most of the time, it is undertaken because of the infeasibility of statistical methods, and it is governed by the same objectives as statistical research, uses procedures that are shadows of statistical procedures, and is evaluated by procedures that are shadows of those used to evaluate statistical research. They mention one situation where a case study is superior to quantitative research: When accurate measurement is too costly to be conducted repeatedly, an intensive research design (my term) in which a great deal of effort is expended on a single case is preferable to relying on measurements of doubtful validity collected in an extensive design for purposes of statistical analysis.¹⁴ Then one must either rely on the case study alone or else use the information gleaned from the case study to adjust one's measurements in a larger sample that is then subsequently subjected to statistical analysis.¹⁵ However, except for this single situation, qualitative research is viewed as a second-best research strategy, undertaken because quantitative strategies are

11. Diesing 1991, 108–10.

12. King, Keohane, and Verba 1994, 37.

13. Diesing 1991, 143.

14. King, Keohane, and Verba 1994, 67.

15. *Ibid.*, 68–69.

infeasible. Correspondingly, conclusions about causal processes in qualitative research are possible but are said to be “relatively uncertain.”¹⁶

King, Keohane, and Verba’s argument is founded on applying statistical logic to causal inference. *Every* concept they apply to empirical social science is borrowed from classical statistics:

1. At its most basic, empirical activity is viewed as the making of discrete “observations,” which are represented as values assigned to variables.
2. Their model of the representation of observations is a “statistic.”¹⁷
3. The three criteria that they apply to judging methods of making inferences are “unbiasedness, efficiency, and consistency”¹⁸—terms familiar to anyone who has ever studied statistics. The brief formalizations of important concepts—bias, efficiency, measurement error, endogeneity—are also familiar statistical territory.

King, Keohane, and Verba do not consistently apply any nonstatistical criteria. They explicitly mention construct validity¹⁹ once in a discussion of precision versus accuracy,²⁰ and they also seem to be discussing this subject in the guise of the “bias-efficiency trade-off,”²¹ but they do not devote any sustained attention to this or any other matters connected with the movement between the language and propositions of a theory and those of an empirical investigation. Thus, the question of assessing the adequacy of operationalizations—the defining of the empirical referents to theoretical concepts—seems to fall outside the scope of their inquiry. In spite of their sympathy for Eckstein’s idea that different hypothesis tests might possess different levels of stringency,²² they are likewise skeptical of the overall thrust of his brief for “critical cases,”²³ contending that

- (1) very few explanations depend upon only one causal variable; to evaluate the impact of more than one explanatory variable, the investigator needs more than one implication observed;
- (2) measurement is difficult and not perfectly reliable;
- and (3) social reality is not reasonably treated as being produced by deterministic processes, so random error would appear even if measurement were perfect.²⁴

Here their qualifying remarks about cases containing multiple observations are set aside (at least, that is how I read the reference to “one implication” and a passage just preceding this one, where “case” is defined as a single observation). So, too, is any sense that in some contexts what King, Keohane, and Verba would term the “reliabil-

16. *Ibid.*, 6.

17. *Ibid.*, 53.

18. *Ibid.*, 63.

19. As defined by Cronbach and Meehl 1955, construct validity refers to whether an empirical test can be shown to be an adequate measure of some theoretical term.

20. King, Keohane, and Verba 1994, 151.

21. *Ibid.*, 69.

22. *Ibid.*, 209.

23. Eckstein 1975.

24. King, Keohane, and Verba 1994, 210.

ity” of the observations is not an important consideration, because it is known to be quite high. Finally, the implicit Bayesianism of Eckstein’s call to focus on critical cases is not addressed by King, Keohane, and Verba. Their position seems to leave them with very little leeway for arguing that one case is superior to another one as a subject of research. Although they seem to accept Eckstein’s notion that some tests are more demanding than others, they provide no basis for making such an assessment.

What has happened in their argument is that the problem of making inferences about the correctness of a theoretical account of causal processes has been redefined without comment as the problem of making statistical or quasi-statistical inferences about the properties of a sample or of the universe that underlies that sample. Although at the outset the inferences they profess to consider are of the former type,²⁵ by the time they discuss the barriers to drawing correct inferences about a theory from the properties of the data,²⁶ they treat the entire problem as a statistical one. Later qualifications to the effect that negative empirical results need not entail the automatic rejection of a theory,²⁷ though useful practical advice, are not grounded in this discussion and certainly do not follow as a matter of “logic” from any preceding argument in the book.

Although King, Keohane, and Verba would have us believe that model acceptance or rejection rests on logical deductions from the results of data analysis, that does not seem to be what happens in several well-known domains. To claim that inferences are drawn and tested is not to claim that they are tested using a process that mimics classical statistics or relies only on the results of statistical tests.

Stephen Toulmin has suggested that legal proceedings be taken as an exemplar of how a community arrives at judgments about the truthfulness of various statements.²⁸ In such proceedings judges or juries are asked to make judgments about causation and intent based quite literally on a single case. Although statistical evidence is sometimes used in court, the only way that judicial judgments are statistical in any more general sense is if the term is meant to apply to the implicitly probabilistic conception of guilt that underlies an evidentiary standard such as “beyond a reasonable doubt.” Likewise, if one considers the standard set of successful scientific research programs that are commonly used as exemplars in discussions of the philosophy of science, one searches in vain among these cases from early modern chemistry, astronomy, or physics, from the germ theory of disease or the theory of evolution, for any instance where explicit statistical inference played a noticeable role in the development of these research programs.²⁹ If King, Keohane, and Verba are correct, how could any judge or jury ever convict anyone (unless perhaps the defendant were being tried for multiple crimes)? If there is a statistical logic to all scientific inference, what are we to make of situations in the physical or biological sciences where a

25. *Ibid.*, 7–8.

26. *Ibid.*, 63 et seq.

27. *Ibid.*, 104.

28. Toulmin 1972.

29. Genetics and psychometrics are exceptions to this generalization. Glymour et al. 1987, chap. 9.

few observations (or even a single one for Einstein's theory of relativity and the bending of light by gravity) in nonexperimental situations were widely perceived to have large theoretical implications?

King, Keohane, and Verba seek to accommodate the drawing of valid conclusions about causes in such situations by means of two claims. The first, noted earlier, is that causal inference is possible in such situations, though with a relatively lower degree of confidence. The second is their repeated acknowledgments that case studies often contain many observations, not just one.³⁰ The claims taken together may appear to offer a way to reconcile the drawing of causal conclusions in such situations with the overall thrust of their argument. However, this is so only if one finesses the issue of degree of confidence and ignores the implications of the fact that many observations within cases are generally made on many variables. If a case contains too few observations per variable to warrant statistical analysis, it is difficult to see how its observations could persuade any statistically inclined jury beyond a reasonable doubt. Although all sorts of criticisms are leveled against judicial systems, I am aware of no one who claims that judges and juries are literally incapable of coming to defensible judgments about guilt or innocence on the basis of a single case. Likewise, nobody seems to criticize the empirical work of pre-modern scientists for their seeming lack of concern about the need to repeat their observations often enough to attain statistically meaningful sample sizes.

How then can we make sense of what happens in trials or in fields like astronomy or biology—or in case studies? One way to speak statistically about some domains such as astronomy is to declare that they possess zero or near-zero sample variability—the members of the population are so similar on the dimensions of interest that the informational value of additional observations approaches zero. To the uninitiated, an *a priori* assumption of zero sample variability is no more and no less plausible than an assumption of some arbitrarily large sample variability. If observations are costly and sample variability is believed to be quite low, the case for more observations is hardly self-evident. However, it is probably not wise to attempt to proceed very far in political science on the assumption that sample variability can be neglected.

A more fundamental difficulty lies in King, Keohane, and Verba's contention that how one reacts to statistical results is a matter of "logic." The problem with this claim is revealed once we consider how researchers might respond to statistical results that are unexpectedly inconclusive or even disconfirming. When are poor statistical results to be viewed as (1) "bad luck"—that is, sampling from a tail of a distribution; (2) arising from problems of faulty observation or measurement (reflecting a faulty operationalization of key concepts); (3) suggesting the impact of previously ignored variables; (4) the result of a misspecification of the relationships among variables already included in the model; (5) due to overoptimistic assumptions about the degree of homogeneity of the cases under observation; or (6) evidence that the entire explanatory strategy is misconceived and ought to be abandoned? King, Keo-

30. King, Keohane, and Verba 1994, 47, 52, 212, 221.

hane, and Verba view *statistical* inference as but the most clear-cut form of *scientific* inference,³¹ which is perfectly consistent with their notion that decisions about which model to accept are a matter of “logic.” But if it were a matter of logic, what is the logic of the modeler’s decision in this situation? Although some of these questions yield to the application of various statistical diagnostics or to repeated analysis with different samples or specifications, even then researchers’ conceptions of what is a “reasonable” way to remeasure the data or to respecify the model are heavily dependent on their substantive understandings of the problem being investigated. If there is a “logic” of how to do this, it is not supplied by King, Keohane, and Verba. Perhaps their practical experience as researchers has convinced them that this sort of decision cannot be guided by abstract, general rules and must be based on a context-sensitive understanding of the adequacy of empirical methods, the theory in question, the plausibility of rival theories, and the level of confidence in the myriad “auxiliary hypotheses” that provide the mostly unspoken set of assumptions underlying the research task. That perspective is one that is both widely shared and possessed of articulate and persuasive defenders, but it is not consistent with their claim to present a general “logic” governing all social scientific research or the Hempelian approach that they believe to be the foundation for their inquiry. Here, what guides research is not logic but craftsmanship, and the craft in question is implicitly far more substantively rich than that of “social scientist without portfolio.” The latter’s lack of context-specific knowledge means that the researcher cannot call on information from outside the sample being analyzed to supplement the information gleaned from statistical analyses. (Just how qualitative information from outside a sample is weighted and combined with statistical information to produce a considered judgment on the accuracy of a theory is not well understood, but if the qualitative information is accurate, the resulting judgment ought to merit more confidence). For someone equipped with adequate contextual knowledge, a given statistical (or quasi-statistical) analysis still affects the evaluation of the accuracy of a theory, but it is only one consideration among several, and its preeminence at an early point in the research project is far from obvious.

If scientific inference is treated as essentially statistical, it is no wonder that King, Keohane, and Verba view case studies as chronically beset by what I term a “degree-of-freedom problem” or what they term “indeterminate” research designs:³² the number of “observations” is taken to be far fewer than the number of “variables.” (Here, again, the qualifiers about case studies containing many observations are set aside.) This situation precludes the statistical identification of models—hence King, Keohane, and Verba’s use of the “indeterminate” label.

James Fearon, like King, Keohane, and Verba, has argued that all causal inferences are statistically based.³³ Yet he provides a riposte to this contention in his discussion of what he terms “counterfactual” explanations:

31. *Ibid.*, 6.

32. *Ibid.*, 119–20.

33. Fearon 1991, 172.

Support for a causal hypothesis in the counterfactual strategy comes from *arguments* about what would have happened. These arguments are made credible (1) by invoking general principles, theories, laws, or regularities distinct from the hypotheses being tested; and (2) by drawing on knowledge of historical facts relevant to a counterfactual scenario.³⁴

What Fearon offers is a strategy for constructing a nonstatistical basis for causal inferences. However, if one can support causal inferences by means of arguments of the sort that Fearon mentions, there is no need for counterfactual speculation. One can just move directly from the arguments to the conclusions about causal processes operating in the case, without any need to construct counterfactuals. Fearon's strategy is always available, whether or not one is interested in constructing counterfactuals. (However, as discussed later, case study researchers might have good reasons to be interested in counterfactuals.)

As applied to a setting such as a trial or a case study, two types of arguments can be mustered in support of causal conclusions. The first are causal claims that are so uncontroversial that they operate essentially as primitive terms. If the jury views an undoctored videotape in which a suspect is seen pointing a gun at the victim and pulling the trigger, and the victim is then seen collapsing with a gaping hole in his forehead, it reaches conclusions about the cause of the victim's death and the intent of the suspect to shoot the victim that are highly certain. Barring the sort of exotic circumstances that a philosopher or a mystery writer might invoke (for example, the victim died of a brain aneurysm just before the bullet struck, or the gun was loaded with blank cartridges and the fatal shot was fired by someone else), the assessment of causation is unproblematic. Even if exotic circumstances are present, a sufficiently diligent search has a good chance of uncovering them, as any reader of detective fiction knows.

A second type of causal claim is weaker: It is the "circumstantial evidence" so often used by writers of murder mysteries. An observation may be consistent with several different hypotheses about the identity of the killer and rules out few suspects. No one observation establishes the identity of the killer, but the detective's background knowledge, in conjunction with a series of observations, provides the basis for judgments that generate or eliminate suspects. As the story progresses, we are usually presented with several instances in which "leads" (that is, hypotheses) turn out to be "dead ends" (that is, are falsified by new observations). Sometimes an old lead is revived when still more new observations suggest that previous observations were interpreted incorrectly, measures or estimates were mistaken, or low-probability events (coincidences) occurred. Typically, the detective constructs a chronology of the actions of the relevant actors in which the timing of events and the assessment of who possessed what information at what time are the central tasks. This tracing of the causal process leads to the archetypical final scene: All the characters and the evidence are brought together and the brilliant detective not only supplies the results of the final observation that eliminates all but one suspect but also

34. *Ibid.*, 176, emphasis in original.

proceeds to explain how the observations fit together into a consistent and accurate causal explanation of events. Rival theories are assessed and disposed of, generally by showing that they are not successful in accounting for all the observations. The suspect may attempt to argue that it is all a coincidence, but the detective knows that someone has to be the killer and that the evidence against the suspect is so much stronger than the evidence against anybody else that one can conclude beyond a reasonable doubt that the suspect should be arrested.

It may be objected that in this situation all that is happening is that the statistical basis for conclusions has merely been shifted back from the immediate case at hand to the formation of the prior beliefs. The hypothetical juror then deduces the correct verdict on the basis of those prior beliefs, which are themselves based on statistical inference. Although there is no reason why this is impossible, it is a less than satisfactory defense of the attempt to ground all conclusions on a foundation of statistical inference. First, it is merely an epistemological “IOU”—it does not resolve the issue, it merely displaces it back to the question of how the prior beliefs were formed. Moreover, it uses “statistical inference” metaphorically, as a catchall descriptor for the process of making sense of experience. As such, it attempts to finesse the need to use judgment (as, for example, in the earlier discussion of how to respond to negative statistical results, or in the question of deciding what rules or laws are relevant to a single case, or of classifying a single case as a member of one set and not another). Although Johannes von Kries argued more than one hundred years ago that conclusions about causal linkages in singular cases such as legal proceedings can be treated as resting on probabilistic “nomological knowledge” of links between events, each of a certain type,³⁵ formulations such as his fail to deal with the question of how events are to be sorted into types in the first place. Such an activity is one of judgment, not statistics. (Before deciding how likely it is to draw a red ball out of an urn, one has to know not only something about the contents of the urn but also from which urn the ball came.) The statistical view of prior knowledge also provides no way of making sense of operations that are nondeductive—it cannot, for example, make sense of its very own use of the “juror as statistician” metaphor, because creating or invoking metaphors is not an operation in statistical theory. Finally, it offers no defense of the Humean reliance on deduction from prior knowledge and current observations to a conclusion. “Beyond a reasonable doubt” and the “reasonable person” standard are not equivalent to certainty. Although they might be sometimes interpreted as the verbal equivalent of statistical significance levels with a very small “*p*” value, they are also terms that apply to operations of judgment and classification. If certainty is a better standard to use, the case for it ought to be made. Such a case would have to explain how certainty could ever be reached on questions of judgment and classification and what is to be done if it cannot.

The detective’s reconstruction of the case is what Wesley Salmon terms an “ontic” explanation. Although it rests on a foundation of observed regularities, the regularities themselves are only the basis for an explanation—they are not the explanation

35. Ringer 1997, 64.

itself. The explanation provides an answer to a “why” or “how” question in terms of mechanisms of (probabilistic) cause and effect:

Mere fitting of events into regular patterns has little, if any, explanatory force. . . . [Although] some regularities have explanatory power, . . . others constitute precisely the kinds of natural phenomena that demand explanation. . . .

To provide an explanation of a particular event is to identify the cause and, in many cases at least, to exhibit the causal relation between this cause and the event-to-be-explained.³⁶

The ontic conception is a more demanding standard than the following common strategy in statistical work in political science: (1) Positing a series of bivariate functional relationships between a dependent variable and various independent variables, rooted perhaps in intuition or in expectations formed from prior research; (2) demonstrating statistical regularities in a set of observations; and (3) claiming to have a satisfactory explanation of variation in the dependent variable because there is an adequate statistical accounting of covariation. From the ontic perspective, we do not have an adequate explanation of the phenomenon under study until we can say why the model works.³⁷ Moreover, if we can do this, we are much less likely to succumb to what Andrew Abbot has called “general linear reality”—the casual acceptance of the behavioral assumptions implicit in general linear statistical models in situations where they are not appropriate.³⁸

Equipped with this understanding of explanation, we can now make sense of Ronald Rogowski’s point that one case sometimes seems to have an impact on theorizing that is way out of proportion to its status as a nonquantitative, low-*n* “observation.”³⁹ He cites Arend Lijphart’s study of political cleavages in the Netherlands as an example of such a case study. Though it analyzed only one political system, its publication led to major changes in the way that political cleavages were theorized. A similar example from the study of international relations is Graham Allison’s study of the Cuban missile crisis, which had a large impact on the extant practice of theorizing the state as if it were a unitary, rational actor.⁴⁰ Understanding such situations from the standpoint of King, Keohane, and Verba’s analysis is difficult.

Does the reassessment of a theory require the replication of any anomalous finding first obtained in a case study? King, Keohane, and Verba seem usually to answer “yes,” as when they extol the value of various strategies to increase the number of observations;⁴¹ however, when they discuss the relation of Lijphart’s findings in his case study of social cleavages and social conflict in the Netherlands to the literature

36. See Salmon 1984, 121–22. For a very similar account in explaining why Darwin’s work was critical to the development of biology, see Rescher 1970, 14–16.

37. Aronson, Harré, and Way contend that the deductive-nomological framework drastically underestimates the importance of models for doing science and argue that the provision of adequate models rather than the writing of general laws is the primary activity of science. Aronson, Harré, and Way 1994.

38. Abbot 1988.

39. See Rogowski 1995, 468; and King, Keohane, and Verba 1994.

40. Allison 1971.

41. King, Keohane, and Verba 1994, 120 et seq.

on pluralism that preceded it, they seem to answer “no.”⁴² Although one could justify a “no” answer in terms of nonstatistical explanations of the sort mentioned by Fearon, that is not the path that King, Keohane, and Verba choose. Like Fearon, they seem to believe that case studies are beset by a degree-of-freedom problem; unlike him, they cannot offer any alternative to statistical evidence but the mimicking of statistical analyses in verbal, nonquantitative form. How then can a single case study alter our confidence in the truth or falsity of any theory?

One way is that when the existence of a phenomenon is in question, only one case is needed to establish it. Since Lijphart and Allison do just that, it is important, because it suggests that a phenomenon that previous theory had suggested could not exist does in fact occur. However, if it occurs only once, is that important statistically? King, Keohane, and Verba describe Lijphart’s study as “the case that broke the camel’s back.”⁴³ For that to be so, the statistical camel would already have to be under a great deal of strain due to the accumulation of previous anomalous findings. But no other anomalous findings are mentioned. They also note that there had been many previous studies of the relation between cleavages and democracy. If so, the mystery of why this one study should have such an impact only deepens. Unless one believes that this prediction failure is especially threatening to the previous pluralist theory, the presence of many previous studies that found the predicted association between cleavage structure and democracy would provide even more reason to write off Lijphart’s case study as an outlier. No statistical model is rejected because it fails to predict only one case, and the influence of any one case on judgments (or computations) about the true underlying distribution is a decreasing function of sample size—so more previous case studies would imply that Lijphart’s study would matter *less*. Unless the sample is quite small, adding just one “observation” (assuming for the moment that a case study is just an observation) is going to make very little difference. And, from a conventional statistical standpoint, small samples are simply unreliable bases for inferences—whether or not one adds one additional case.

If one accepts that the Lijphart and Allison studies had a pronounced impact on theorizing in comparative and international politics, and if one views this impact as legitimate and proper, there is no way to rationalize this through statistical thinking. Rogowski’s original suggestion for how to understand this situation—as an example of a clear theory being confronted with a clear outlier—is a step in the right direction. But if that were all that were happening, one would simply be presented with an unusually strong anomalous finding, to which one could respond in a large variety of ways.

If a case study can succeed in explaining why a case is an outlier by identifying causal mechanisms that were previously overlooked, it will have a much more pronounced impact. It is not the fact that the old theory is strongly disconfirmed that makes the Lijphart or Allison studies so important; rather, it is their provision of such mechanisms—connecting cleavage structure to democracy, or the state’s organiza-

42. See Lijphart 1975; and King, Keohane, and Verba 1995, 477.

43. King, Keohane, and Verba 1995, 477.

tional structure to observed outcomes—in empirical accounts that fit the data at least once. In the provision of alternative accounts of causation, perhaps relying on different concepts than formerly employed, one finds the primary reason for the impact of the single case.⁴⁴ John Walton assesses a set of “classic” cases in sociology similarly—their importance lies in their provision of “models capable of instructive transferability to other settings.”⁴⁵ In the same vein, Nicholas Rescher speaks of Darwin as providing a “keystone” for the development of modern biology; the keystone was not a missing piece of data, but a missing step in a causal argument. That missing step was developed from a combination of intense observation and theoretical arbitrage (his borrowing from Malthus).⁴⁶

Cases are often more important for their value in clarifying previously obscure theoretical relationships than for providing an additional observation to be added to a sample. In the words of one ethnographer, a good case is not necessarily a “typical” case, but rather a “telling” case, in which “the particular circumstances surrounding a case serve to make previously obscure theoretical relationships sufficiently apparent.”⁴⁷ Max Weber seems to have had a similar conception of ideal types—he saw them as deliberately “one-sided” constructs intended to capture essential elements of causation and meaning in a particular setting, without regard to whether they adequately represented all relevant situations.⁴⁸

John Walton and Arthur Stinchcombe offer an even stronger claim—that the process of constructing a case study is superior to other methods for the task of theory construction.⁴⁹ This is supposedly so because completing a case study requires the researcher to decide what exactly something is a case of and exactly how causation works. Although case studies do not seem to me be unique in this regard (at the very least, the same could be said about two other research strategies discussed later), it seems plausible that the activity of searching for and identifying sources of variation in outcomes is likely to lead to richer models than a research strategy that can easily use controls and randomization to build a firewall separating a larger causal mechanism from a small number of variables of immediate interest.

The issue of whether a causal mechanism must be provided in order for an argument to be considered a scientific theory is precisely the point at which King, Keohane, and Verba’s inattentiveness to the conflicts between Hempel’s deductive-nomological conception of theory and more recent philosophical accounts leads to confusion about what case studies are capable of accomplishing. It is not merely that a case provides an explanation for a particular set of events. Rather, the source of its potentially large impact is its capacity to incite us to reformulate our explanations of previously studied events.

44. See Laitin 1995; and Caporaso 1995.

45. Walton 1992, 126.

46. Rescher 1970, 15.

47. Mitchell 1984, 239.

48. See Burger 1976, 127–28; and Hekman 1983, 25.

49. See Walton 1992, 129; and Stinchcombe 1978, 21–22.

Toward a Methodology of Intensive Research: An Outline of an Alternative “Logic” for Case Studies

King, Keohane, and Verba’s choice of a statistical framework for thinking about all studies, and their attempt to distinguish descriptive from explanatory constructs and to privilege the latter,⁵⁰ leaves unclear the status of several research techniques in the study of international relations. What is the status of such projects as the construction of decision or game trees, or computer language or ordinary language representations of a decision-making process? These are possible end-products of a “process tracing” research strategy. Are they just “descriptions?” Or are these “theories” in any sense that King, Keohane, and Verba would recognize as legitimate? If a verbal description can be a “model,” are these other constructs also models? Are they explanations? More broadly, how (if at all) can we make sense of such activities from the standpoint of King, Keohane, and Verba’s explication of good research design? Diesing explicitly argues that these research activities cannot be subsumed within the statistical framework; is he right?

Claims that nonstatistical explanation is possible matter little if they cannot be substantiated with examples of how such explanations can be constructed and evaluated. Although the earlier examples of courts, hermeneutic readings, and theory building in the physical and biological sciences do substantiate the contention that such research is an important alternative to conventional statistical approaches, the philosophical and practical issues involved in such research have received far less attention within political science than these same issues in quasi-experimental research.

Although a complete explication of the philosophical and operational issues involved in intensive research could easily be as long as a book, we can identify some issues that such a methodology must address, as well as some ways of addressing them.

Understanding Existing Research

A substantial body of literature within the field of international relations is much more easily understood from within Salmon’s ontic conception of explanation than the modified Hempelian framework preferred by King, Keohane, and Verba. Examining two well-known research programs in terms of the language and concepts of King, Keohane, and Verba is helpful in revealing exactly how far one can extend the kind of framework they offer without encountering research practices that are not readily accommodated within their account. In each case the discussion parallels that of King, Keohane, and Verba: First, the elementary empirical “atom” is defined; second, how the “atoms” are assembled is described; third, how these assemblies are evaluated is addressed. I then note some problems in attempting to carry though King, Keohane, and Verba’s conception of research in these domains.

50. King, Keohane, and Verba 1994, chap. 2.

Cognitive mapping. An important research program in the study of foreign policy decision making builds on Richard Snyder, H. C. Bruck, and Burton Sapin's suggestion to construct a theory that captures decision-makers' "definition of the situation" and the decision-making process they use.⁵¹ If our project is to construct ordinary language or machine language representations of decision-making processes along the lines of "cognitive maps" in Robert Axelrod's sense or expert systems in Charles Taber's or Hayward Alker's sense, the basic "atom" of empirical work would be the sentence (in ordinary language) or the statement (in machine language) rather than the value (typically, though not necessarily, numerical) of a variable.⁵² There does not seem to be nor does there have to be any kind of representation of the atomic units in reduced form (something equivalent to the moments of a distribution in the statistical example). However, we can speak of the ensemble of empirical atoms as a "protocol" (in ordinary language) or a "program" (in machine language). There is little point in speaking of the output of a program as being "caused" by one line of computer code apart from the other lines of code; thus, the objective of apportioning causal weights to the various components of the model, an important part of the statistical project, has no counterpart in an artificial intelligence or a cognitive mapping context. (If translating such a model into a statistical framework is necessary, it would be akin to a statistical model in which each of the explanatory variables has no main effect, but rather enters the model only interactively.) After being appropriately initialized with assumptions deemed to capture essential aspects of a historical situation, the model is fitted to historical data, and this fitting exercise can be assessed quantitatively.⁵³ However, an assessment method such as comparing root mean squared errors can be undertaken without reference to a defined universe, samples, or significance and in this sense is not "statistical" at all.

Anders Ericsson and Herbert Simon have articulated the research strategy used in cognitive mapping in terms that are highly similar to the ontic conception outlined earlier: "A single verbal protocol is not an island to itself, but a link in a whole chain of evidence, stretching far into the past and the future, that gradually develops, molds, and modifies our scientific theories. It needs to be processed with full attention to these linkages."⁵⁴

For Ericsson and Simon, theories suggest data to acquire, while data suggest theories to investigate—one is not logically prior to or dependent on the other. Unlike Popper's world, where research is typified in terms of a single movement from the logic of discovery to the logic of falsification, the research process here cycles between theory (re)formulation and theory evaluation. Hypotheses and theory formulation are treated as activities amenable to normative guidance, rather than a completely subjective realm.

51. Snyder, Bruck, and Sapin 1954.

52. See Axelrod 1976; Taber 1992; and Alker 1996.

53. Cyert and March 1963, 320.

54. Ericsson and Simon 1984, 280.

Game theory applied to empirical situations. The story is much the same from a rational choice standpoint. Here a formal representation of the decision-making process involving strategic interaction is constructed, based on a relatively slender and simple set of postulates. The empirical accuracy of this game is then assessed by comparing its predictions with actual outcomes in a situation thought to be relevant to assessing the performance of the formal model. Bruce Bueno de Mesquita and David Lalman provide a good example of this approach.⁵⁵ (In many game-theoretic accounts the fit to empirical situations is addressed more cursorily, because the analyst's primary intention is to elucidate the consequences of a given set of initial assumptions, rather than to provide a good empirical fit *per se*.)⁵⁶ In a game-theoretic representation, there is not one kind of atom, but five: players, nodes (representing outcomes), branches (representing alternatives), utilities, and probabilities. The ensemble of atoms forms a tree, or a game in extensive form. The ensemble as a whole governs choice, and, again, framing queries about the relative causal weight of one atomic unit versus another one is pointless. Goodness of fit can be assessed as in the cognitive-artificial intelligence situation, or (more commonly) a statistical model is constructed based on the tree and auxiliary hypotheses ("operationalizations").⁵⁷ In this latter situation one can, if one wishes, assess the weight or influence of individual factors. Although the statistical evaluation of the performance of such models is an activity that raises no difficulties from King, Keohane, and Verba's point of view, the question of how one settles on a given cognitive map or tree for evaluation is not answerable from within the confines of their perspective.

Although games can be infinitely long, a game tree is often finite; it does not attempt to trace causation back beyond a starting point chosen by the analyst. Thus, King, Keohane, and Verba's objection that attempting to describe completely the causal mechanisms in a concrete situation leads to explanations that are in principle infinitely large⁵⁸ is irrelevant, since explanations do not aim at being complete, but merely at answering the question that the researcher asks.⁵⁹ Human decision making is inherently limited in the number of factors that impinge on the awareness of the decision maker, thus making possible the construction of trees that are reasonably complete representations of the decision situations facing historical actors, as those actors see them. As Alexander George and I argued, "Because the limitations on the perceptual and information-processing capabilities of humans are well known and pronounced, the process-tracing technique has a chance of constructing a reasonably complete account of the stimuli to which an actor attends."⁶⁰ Constructing such a tree is thus feasible, though in any given historical situation the limitations of the available evidence may create a situation where we are not confident that our tree representation of the decision-making situation is accurate and complete. (An addi-

55. Bueno de Mesquita and Lalman 1992.

56. I am grateful to Robert Powell (personal communication) for emphasizing this distinction.

57. Signorino 1998.

58. King, Keohane, and Verba 1994, 86.

59. Levi 1984, 51.

60. George and McKeown 1985, 36.

tional limitation to this approach is that once we leave the world of binary interaction and attempt to model three or more independent agents, the capacity of formal theories of optimizing behavior to provide solutions that are relevant to empirically encountered situations diminishes sharply without adopting many seemingly arbitrary restrictions.)⁶¹

The strategy of constructing a tree based on historical information can in principle also address two other problems that King, Keohane, and Verba rightly discuss as common failings of qualitative research: inattentiveness to selection bias and a failure to specify counterfactual claims with enough precision or accuracy to permit their intelligent use in an assessment of which factors really matter in shaping behavior in any given situation. King, Keohane, and Verba and their reviewers discuss selection bias as if it amounts to a statistical problem (that is, an error in sample construction), which is certainly one way to think of it. However, another way to view it is to say that it amounts to being unaware of the fact that the game that one has just analyzed is merely a subgame of a larger game. The difference in conceptualizations is important, because how one views selection bias determines how one evaluates work plagued by it. From a statistical standpoint, an improperly drawn sample will likely result in statistical findings that are useless for making inferences about an underlying population—particularly when the nature of the bias is not known. However, from a game-theoretic standpoint the analysis of a subgame, if conducted correctly, still provides a valid and useful result. If an analyst does not realize that the outcomes of interest can be reached from branches of the tree that occur prior to the node at which the analysis begins to investigate the decision-making process (as happens in the studies of deterrence mentioned by King, Keohane, and Verba),⁶² then the analyst will likely be mistaken in judgments about which factors are most important in reaching an outcome. Once the analyst has realized that the relevant tree for analyzing the outcome of interest is larger than initially realized, the initial results are still useful as part of a larger tree. What before were (mistakenly) viewed as unconditional probabilities are now seen as conditional ones. Although this change may destroy the case for policy prescriptions based on the old, incorrect view, the tree of the subgame survives intact and is now nested with a larger tree and a more complete explanation.

Another signal advantage of thinking in terms of trees is that they explicitly represent counterfactual situations. By doing so, they delineate which counterfactual situations among the infinite number available for consideration are the most theoretically relevant. Assuming we know the preferences attached by actors to these counterfactual outcomes, we can address the question of how changes in the payoffs—either of the outcome that occurred or of the outcomes that did not—affect the choices made in the given decision situation.⁶³

61. Ekeland 1988, especially chap. 1.

62. King, Keohane, and Verba 1994, 134–35.

63. Brady seems to suggest a similar treatment in discussing the implications of King, Keohane, and Verba's approach to causal analysis and counterfactuals. Brady 1995.

It has been objected that trees or other choice-theoretic representations are just as mechanistic a method as relying on statistical inference for the development of theories of cause and effect.⁶⁴ Both methods are seen to be fundamentally in error because they treat political phenomena as “clocklike,” when in reality there are aspects of political life that make the clock metaphor ultimately inappropriate—in particular, (imperfect) memory and learning. Such an argument fails to grasp that even with the use of some clocklike representation of decision making, the resulting explanation of behavior will still be incomplete. Although the problem of modeling preference change was addressed nearly forty years ago, very little progress has been made.⁶⁵ The “rules of the game” must also generally be analyzed the same way, since our current capacities to understand institutions as the outcome of strategic interaction are still quite limited. The use of trees, computer simulations, and so on should be understood as an attempt not to model political systems as if each were a single clocklike mechanism, but to extract the clocklike aspects from a social situation in which we possess “structural” knowledge (in Jon Elster’s sense) only of some features.⁶⁶

Although they are not typically described in this fashion, both the cognitive–artificial intelligence and choice-theoretic approaches can also be understood as implementations of Weber’s venerable concept of “ideal types.” (This family resemblance is seldom discussed in treatments of Weber’s methodology, but it becomes a good deal more understandable once one learns that his work on ideal types was in part a response to economic theory and that he persistently cited that theory to illustrate the uses of “ideal-typical” construction.)⁶⁷ They amount to ways of fusing a conception of actors’ definition of the situation with a conception of a social structure within which social action occurs. Although they part company with Weber on the question of whether a model can be empirically accurate (with Weber seemingly arguing that empirical accuracy is not a property usefully attached to an ideal type),⁶⁸ they share with Weber an interest in fusing the “subjective” and “objective” aspects of a social situation in a single model.

A “Folk Bayesian” Approach

We can use the notion that humans (particularly social scientists) are intuitive statisticians and view them as folk Bayesians.⁶⁹ This is a different statistical metaphor than was applied by King, Keohane, and Verba, who do not cover Bayesian approaches. Supplanting or replacing their more conventional understanding of statistics with a Bayesian one would improve their account in two ways. First, it would enable us to make sense of several previously inexplicable research activities, some of which

64. Almond and Genco 1977, 509.

65. Cohen and Axelrod apparently provide the only model of this process by political scientists. Cohen and Axelrod 1984. The *Social Sciences Citation Index* shows that Cohen and Axelrod’s paper has been cited once in the last ten years.

66. Elster 1983.

67. See Burger 1976, 140–53; and Ringer 1997, 110.

68. Burger 1976, 152–53.

69. Putnam 1981, 190–92.

King, Keohane, and Verba acknowledge and approve, some of which they do not. Second, it would extend and enrich the normative directives they provide by giving researchers guidance on how to think and act systematically about likelihoods and loss functions, rather than continuing to rely solely on their intuitions to guide them.

A Bayesian approach to the problem of explanation is not a panacea: There are important difficulties on both an operational⁷⁰ and a philosophical⁷¹ level. Moreover, to say of researchers that they are “folk Bayesians” implies that their application of Bayesian principles is largely intuitive—it has usually been more a matter of making research decisions in the spirit of Bayes than of consciously applying Bayesian techniques. It is therefore more useful in this context to view Bayesian statistical theory as a metaphor than as an algorithm.

The Bayesian metaphor comes to mind when one considers that researchers in the social sciences, even in the branches that rely heavily on classical statistics, are “interactive processors.” They move back and forth between theory and data, rather than taking a single pass through the data.⁷² As Edward Leamer notes, one can hardly make sense of such activity within the confines of a classical theory of statistics.⁷³ A theory of probability that treats it as a process involving the revision of prior beliefs is much more consistent with actual practice than one that views the information in a given data set as the only relevant information for assessing the truth of a hypothesis.

If we treat researchers as “folk Bayesians,” several research practices seen as anomalous by King, Keohane, and Verba become much easier to understand. I have already suggested that Eckstein’s ideas on critical cases seem to emanate from a folk Bayesian perspective: The selection of cases for investigation is guided by the researcher’s prior probabilities of a given explanation being correct in a certain kind of setting, coupled with that researcher’s assessment of the costs of being wrong in that assessment. A “hard case” for a theory (Stephen Van Evera’s “smoking gun” case) then would be one where the prior probability of a theory being a correct explanation is low but the degree of confidence in that prior assessment is not high.⁷⁴ A “critical case” would be one where the prior probability is an intermediate value, such that either a confirmation or a disconfirmation will produce a relatively large difference between the prior and posterior probabilities. One might also select a case in which the expected disutility of being wrong were low and then proceed to more demanding tests only if the initial results are encouraging. This would make good sense if investment in a large research project entailed substantial costs.

A Bayesian perspective can also make sense of King, Keohane, and Verba’s “file drawer problem”⁷⁵ in which negative research findings are relegated to researchers’ private files, and only positive findings are submitted for publication. From this perspective, King, Keohane, and Verba’s contention that a negative result is as useful as

70. Leamer 1994.

71. Miller 1987.

72. Gooding 1992.

73. Leamer 1994, x.

74. Van Evera 1996, 40.

75. King, Keohane, and Verba 1994, 105fn.

a positive one is only true if one originally thought that both results were equally likely. If one conjectured that a positive result were highly likely, then getting such a result would be minimally informative. Thus, a journal devoted to electoral behavior would probably not publish the “positive result” that white American evangelical Protestants in 1994 were more likely to vote for the Republicans than the Democrats, simply because nobody would view that as news. Conversely, the negative result that a sector-specific model of coalitions in U.S. trade politics does not account for the coalitional pattern surrounding NAFTA is news indeed, simply because the prior model had become so well accepted.⁷⁶

A Bayesian perspective also yields a different normative judgment about the preconceptions of researchers than that offered by King, Keohane, and Verba. Whereas in their view having a preconception makes one “slightly biased,”⁷⁷ from a Bayesian perspective having a preconception is *necessary* in order to make sense of one’s research results. One cannot do Bayesian analysis without an intelligible prior probability for the outcomes in question. (The difficulty in forming such priors in some cases is an important criticism of the indiscriminate use of Bayesian analysis.)⁷⁸ King, Keohane, and Verba’s position on preconceptions is not unreasonable—succumbing to motivated perceptual bias is always a danger, and it is well that it should be flagged. However, thinking that a researcher has no preconceptions is unrealistic, and ignoring the useful role that preconceptions can play is not at all “conservative.”

If a Bayesian begins a case study with a prior estimate of some variable that is close to zero, but with a prior estimate of the variance of that estimate that is relatively large—because the number of prior observations has been zero or very small—the observation of the first anomalous result is going to raise the posterior estimate of the anomalous finding a very considerable distance above zero. Thus, the change in the subjective assessment on the basis of just a single case would be quite large, but it would be understood as a simple application of Bayesian statistical theory, rather than as a finding that poses any unusual challenge to a statistical understanding of cases.

The Bayesian perspective is also implicit in King, Keohane, and Verba’s own advice to begin with theories that are “consistent with prior evidence about a research question.”⁷⁹ This seemingly amounts to *de facto* acceptance of the prior evidence (that is, assigning it a relatively high prior probability of being based on a correct theory of observation), and this too is far from innocuous. A concrete example illustrates what is at stake. In studying U.S. foreign policy decision making, one confronts a raft of studies by diplomatic historians and political scientists that purport to explain foreign policy decision making by what amounts to a “realist” model—one in which the geostrategic environment drives decisions, and other factors intrude at most in a secondary way. These studies take as their evidence a moun-

76. See Magee 1980; and Commins 1992.

77. King, Keohane, and Verba 1994, 71.

78. Miller 1987, 269.

79. King, Keohane, and Verba 1994, 19.

tain of declassified government documents that offer geostrategic justifications for various foreign policy decisions. However, the decision of these researchers about where to search for evidence about the motivations of U.S. central decision makers is itself driven by their theoretical conception of what motivates decision makers and how they decide. The resulting studies are vulnerable to criticism because (1) they generally fail to consider whether policy options that were not chosen also have plausible geostrategic justifications, (2) they generally offer no method for distinguishing between plausible post hoc rationalizations for policy and the reasons why a policy is adopted, and (3) they inadequately address rival hypotheses or theories. As a result, the research program is liable to criticism that it creates a circular argument.⁸⁰ Whether this argument is always true is less important than the broader and more general implication that “prior evidence” is unproblematic only to the extent that one accepts the theoretical preconceptions that generated it. If one disagrees with those preconceptions, it makes no sense to assign the evidence generated on the basis of those preconceptions a high prior probability of being correct. In such situations it would not be surprising or improper if those who propose a new theory respond to an inconsistency between their theory and existing data by criticizing the “form of these data.”⁸¹

The sharpest difference between folk Bayesians and King, Keohane, and Verba is in their differing assessments of ex post model fitting. King, Keohane, and Verba’s view is that ex post model revisions to improve the fit of the model to the data “demonstrate nothing” about the veracity of the theory.⁸² Some disagree. For example, Ericsson and Simon argue that the time when a hypothesis was generated is not, strictly speaking, relevant to assessing the posterior probability of it being true. However, they concede that having the data before the hypothesis should probably incline us to place less credence in it.⁸³

Similarly, Richard Miller contends that

When a hypothesis is developed to explain certain data, this can be grounds for a charge that its explanatory fit is due to the ingenuity of the developer in tailoring hypotheses to data, as against the basic truth of the hypothesis. If an otherwise adequate rival exists, this charge might direct us to a case for its superiority. But such a rival does not always exist, and the advantages of having first been developed, then tested against the data are not always compelling. As usual, positivism takes a limited rule of thumb for making a fair argument of causal comparison, and treats it as a universal, determinate rule, functioning on its own. . . .

While confirmation often does exist in such cases, it is usually weaker than it would be on a basis of discovery. . . . A theory of confirmation that makes . . . questions of timing invisible neglects phenomena that are clearly relevant to the comparison of hypotheses—and that ought to be if confirmation is fair causal comparison.⁸⁴

80. Gibbs 1994.

81. Tanner and Swets 1954, 40.

82. King, Keohane, and Verba 1994, 21.

83. Ericsson and Simon 1984, 282–83.

84. Miller 1987, 308–309.

These viewpoints are sensitive to King, Keohane, and Verba's concern about "fiddling" with models solely to improve statistical goodness of fit, but they do not view that concern as dispositive because they value having a fragile model much more highly than having no model. From a Bayesian standpoint, any attempt to retrofit a model onto data that does not use a model that is plausible on other grounds will likely begin with the assignment to that model of a low prior probability of being correct. If the objective is to find a model that has a high posterior probability of being correct, in light of the fact that it fits the data, it is far better to begin with a model that has a high prior probability. In that sense the Bayesian perspective incorporates a safeguard against the sort of abuse that King, Keohane, and Verba fear, without being categorical in its rejection of *ex post* fitting.

In contemporary American political science a Bayesian conception of probability has only recently begun to receive attention;⁸⁵ in the discussion of case study methodology it has received no attention at all (except for a fleeting mention in George and McKeown.)⁸⁶ Given its capacity for linking preobservation to postobservation beliefs about the world, and its explicit consideration of the costs of being wrong, greater attention to Bayesian approaches seems sensible, both for case study researchers and for practitioners of statistical analysis.

Heuristics for Theory Construction

An unfortunate practical consequence of the Popperian perspective and positivism more generally is that they fixate on testing theory at the expense of constructing it. If the extent of one's knowledge about political science were the tables of contents of most research methods books, one would conclude that the fundamental intellectual problem facing the discipline must be a huge backlog of attractive, highly developed theories that stand in need of testing. That the opposite is more nearly the case in the study of international relations is brought home to me every time I am forced to read yet another attempt to "test" realism against liberalism. If only for this reason, a philosophy of science that took seriously the task of prescribing wise practices for constructing theories would be quite refreshing and genuinely helpful.

Such a prescriptive body of theory has been produced piecemeal by researchers who are in contact with the problems that arise in the performance of intensive research. However, to the extent that its existence is even acknowledged, the nature of that theory is often misconstrued. Rather than constituting a set of surefire methods, guaranteed to work because they harness deductive logic to the task of theory construction, these prescriptions are a series of highly useful heuristics. Intended for the boundedly rational inhabitants of a messy world, they provide guidance on how to generate theories or frame problems and where to search for evidence that is relevant to assessing extant theories.

85. See Western, Jackman, and Marks 1994; Jackman and Marks 1994; van Deth 1995; and Bartels 1996 and 1997.

86. George and McKeown 1985, 38.

Case selection heuristics. Case studies are often undertaken because the researcher expects that the clarification of causal mechanisms in one case will have implications for understanding causal mechanisms in other cases. Indeed, it is precisely for that reason that heuristics for case selection—from Mill’s methods of difference and agreement, to Eckstein’s discussion of critical cases, to George and McKeown’s discussion of typological sampling—have been proposed. Pointing out that such heuristics do not guarantee statistical control⁸⁷ and that the generalization of case study findings is problematic is correct but unimportant in this context. Whether a causal account that fits one historical circumstance will fit others is an open question. What matters here is that a causal mechanism has been identified, and the researcher has some framework within which to begin to investigate the external validity of the causal claims. Such a framework permits initial judgments about which cases are theoretically “near” the case in question and whether similarities and dissimilarities in causal patterns in different cases are in line with or diverge from initial understandings of how similar the cases are.

Thought experiments and counterfactuals. Some philosophers and political scientists⁸⁸ have argued that developing and exploring counterfactuals is an important part of the research process. The assertion of counterfactuals is typically associated with attempts to find a causal pattern or to explore the implications of a causal pattern that one believes to be present in the situation being analyzed. In the latter case, an explicit and complete theory (such as the earlier-mentioned completed game tree) generates conclusions about counterfactual circumstances while accounting for the outcomes that did occur. Although such counterfactual conclusions, if valid, may be an important and valuable guide to action, the counterfactual statements themselves merely help the analyst to see the implications of a previously developed theory. In situations where theory is ill-formed and immature, thought experiments reveal “latent contradiction and gaps” in theories and direct the analyst’s search toward nodes in the social interaction process where action might plausibly have diverged from the path that it did follow.⁸⁹ Although in principle there is no reason to associate counterfactual analysis with case studies any more than with other empirical methods, the frequent concern of case study researchers with theories that are relatively immature means that they probably use counterfactuals as a heuristic guiding the search for causal patterns more than those who work with highly developed theories where causality is better understood.

Exploiting Feedback from Observation to Design

Although in a general way all empirical research relies on feedback from empirical work to modify theory and to redirect subsequent research, in case study designs the

87. King, Keohane, and Verba 1995, 134.

88. See Tetlock and Belkin 1996b; and Gooding 1992.

89. Tetlock and Belkin 1996a.

feedback loop often operates within the case besides affecting the theory and methods applied to subsequent empirical research. As King, Keohane, and Verba have noted, doing this is quite difficult to reconcile with a classical conception of statistical inference, but that conception is not well suited to a research environment in which the costs of an inappropriate research design are quite high and relying on the next study to correct the mistakes of the current one is impractical. The latter are often true when conducting fieldwork. In a common fieldwork situation the researcher arrives at the site and quickly learns that certain key assumptions of the research design were based on a mistaken understanding of the case. Perhaps the envisioned data-gathering technique is not feasible. Or the ministry thought to be central to decision making in the issue of interest turns out to be a rubber stamp for another less visible set of interests. This leads to a redesign of the field work, which, as was noted, consumes degrees of freedom. However, the weeks or months of fieldwork that follow this redesign are not rendered worthless simply because they capitalized on information learned early in the research process.

Identifying Causal Processes Rather than Testing

If the investigator is searching empirical evidence to identify causal processes, terming this activity “identification” seems preferable. We can then reserve the term *test* for those situations where more than one substantive⁹⁰ model has been developed and brought to bear, and there is a comparative assessment of the success of the models in explaining the outcomes of interest. The advantage of speaking in this fashion is that it allows us to discuss model identification as an activity that is conceptually distinct from hypothesis formation and testing, and then to address in a systematic way the process involved in doing this well rather than poorly. This saves model identification from being thrown in with hypothesis formation, where it would succumb to the Popperian prejudice against the possibility of saying anything helpful about any other part of the research enterprise than testing. The issue of the generalizability of the model can thus be separated from the question of whether the model is an accurate explanation of cause and effect in the situation in which it has been putatively identified.

Superficially, this may seem to concede an advantage to the statistical view, because a statistical model is always “tested” when its performance is compared with a null model. However, this is an advantage of little importance if one accepts, as King, Keohane, and Verba seem to, the goal of finding the model of a causal mechanism that best accounts for the observations. Given a choice between a null (that is, random) model of planetary motion and one developed by Ptolemy, we would choose the Ptolemaic model every time, because it would perform significantly better than the null model. As long as the relevant statistical tests justified it, we would keep adding epicycles to the model (“variables”) to improve our R^2 . Although hypothetically it is possible that a latter-day Copernicus would write an entirely different

90. A null model is not considered here to be a substantive model.

specification that would succeed in producing a significantly better goodness of fit, given the paltry theoretical weaponry of most empirical investigations (typically, lists of bivariate relations between a dependent variables and other variables that specify the signs of the coefficients, with little or no theoretical guidance on interactions among independent variables, or the precise nature of feedback from the dependent variable to the independent variables), this cannot be relied on. Clark Glymour, Richard Scheines, Peter Spirtes, and Kevin Kelly provide a telling example of the difficulties involved in hitting on the correct representation of an underlying causal mechanism in their brief but sobering analysis of the combinatorics of a six-variable system.⁹¹ Assume that there are four different relations applicable to each pair of variables x and y (x affects y but is not affected by it, y affects x but is not affected by it, they each affect the other, neither affects the other). Given that six variables create fifteen possible variable pairs, there are 415 possible path diagrams one may draw and hence 415 different models to test in order to identify the one that fits the data best. Showing that a model performs significantly better than a null model does little to settle the question of whether it is the best model of the observations that can be written. Accepting “significantly better than null” as the criterion for a successful explanation leads to a perverse, tacit stopping rule for quantitative empirical research: search the universe of plausible model specifications bounded by prior theoretical restrictions until you find one that yields results better than null, then publish. If there are something like 415 specifications from which to select, it would not be at all surprising to find that published models are inferior in terms of statistical goodness of fit to hitherto undiscovered models (which is precisely what Glymour and his colleagues repeatedly show). Thus, the fact that a model can be identified in a statistical sense—and that a computer program embodying the model will indeed run⁹²—is no guarantee that the model is the best account of causal processes that can be written.

How then does model identification proceed? Glymour and his colleagues propose the systematic application of explicit search heuristics to the task of finding models.⁹³ Gerg Gigerenzer claims that researchers often work in just this fashion.⁹⁴ He argues that between the alternatives of treating discovery of models either as a matter of logic or as entirely idiosyncratic there are intermediate possibilities in which search may be guided by one or more heuristics. One possibility that Gigerenzer finds to have been repeatedly applied in research in cognitive psychology is what he terms the “tools-to-theories heuristic”—enlisting methods of justifying claims about models to the cause of organizing the exploration of empirical events. Thus, statistics became not merely a method for evaluating hypotheses, but an organizing concept that affected how psychologists came to think about human thought: The heuristic of decision maker as intuitive statistician has become a central perspective in work on human cognition.

91. Glymour et al. 1987, 7.

92. King, Keohane, and Verba 1994, 118.

93. Glymour et al. 1987.

94. Gigerenzer 1991.

Conclusions

King, Keohane, and Verba are experienced and skilled researchers, and the most successful and original parts of their book are their discussions and recommendations based on their practical experience. The more theoretically self-conscious aspects of their argument—using classical statistics as an exemplar for all questions of research design, and their rather perfunctory attempt to ground such an argument in a philosophical framework of Popper and Hempel—are problematic when they are employed to provide a basis for assessing research practices that rely on intensive investigation of a small number of cases rather than extensive investigation of as many cases as sampling theory suggests are needed. Simply stated, the disparities between case study research and classical statistical hypothesis testing are too great to treat the latter as an ideal typical reconstruction of the former. Rather than treating that disparity as a reason for abandoning case studies or regarding them as pointlike observations, it is just as reasonable to treat it as a reason for rethinking the usefulness of methodological advice founded on such bases as classical statistics and a Hempel-Popper view of epistemology.

What would be an alternative basis for methodological advice? In contrast to King, Keohane, and Verba's definition of science as "primarily [its] rules and methods"—and not its subject matter—Paul Diesing quotes approvingly the hermeneutic maxim "No knowledge without foreknowledge," suggesting that what researchers already know has a decisive impact on how they conduct research.⁹⁵ Indeed, the relationship between a researcher's knowledge of the system being studied to the choice of research method and the interpretation of research findings is a central issue in a variety of contexts: in the choice of subjects to be investigated, in the choice of the case study method rather than a statistical method, in the selection of statistical models to be applied to the data and in the interpretation of statistical findings, in the choice of counterfactuals to be assessed, and in the interpretation of the findings of a single case. Although thinking of researchers as folk Bayesians in their approach to these topics is helpful in making sense of some practices that otherwise appear puzzling or just mistaken, there is little to be gained and much to be lost by insisting on attempting to interpret everything that a researcher does or thinks from a purely statistical standpoint, Bayesian or otherwise.

A more general point is that researchers almost never begin from the starting point envisioned by Descartes or Hume—their thought experiments involving radical doubt radically misstate the situation facing the researcher. Typically, the research task is not how to move from a position of ignorance to one of certainty regarding the truth of a single proposition. Rather, it is how to learn something new about a world that one already knows to some degree. Framed in this fashion, the basic tasks of research are then (1) to devise ways of leveraging existing understanding in order to extend our knowledge, and (2) to decide what are sensible revisions of prior understandings in light of the knowledge just acquired. Bayesian statistics, case selection heuristics,

95. See King, Keohane, and Verba 1994, 9; and Diesing 1991, 108.

counterfactual speculation, and “interactive processing”—moving back and forth between theory formulation and empirical investigation—are all strategies that take into account the mutual dependence of understanding and observation. They are all consistent with a pattern model of explanation, in which the research task is viewed as akin to extending a web or network, while being prepared to modify the prior web in order to accommodate new findings.⁹⁶ Seen in this light, the test of a hypothesis—the central theoretical activity from the standpoint of classical statistics—is but one phase in a long, involved process of making sense of new phenomena.

Recent developments in the history and philosophy of science, artificial intelligence, and cognitive psychology provide a more useful foundation for thinking about the problems of knowledge inherent in performing and evaluating case studies than can be found in Hempel or Popper. Unfortunately, interest in these developments among case study researchers or their statistically inclined critics has been minimal. The result has been a discourse dominated by the classical statistical metaphor, which is often adopted even by those who wish to defend the value of case studies. What is needed if the theory and practice of case study research are to move forward is to explicate case studies from a foundation that is more capable than logical positivism of dealing with the judgments involved in actual research programs. Such a method will not discard or devalue the genuine advances that more positivistic research methodologies have brought to the study of clocks, but will supplement them with better advice about how to cope with the clouds.

References

- Abbot, Andrew. 1988. Transcending General Linear Reality. *Sociological Theory* 6:169–86.
- Alker, Hayward R. 1996. *Rediscoveries and Reformulations: Humanistic Methodologies for International Studies*. Cambridge: Cambridge University Press.
- Allison, Graham T. 1971. *Essence of Decision: Explaining the Cuban Missile Crisis*. Boston: Little, Brown.
- Almond, Gabriel A., and Stephen J. Genco. 1977. Clouds, Clocks, and the Study of Politics. *World Politics* 29 (4):489–522.
- Aronson, Jerrold L., Rom Harré, and Eileen Cortnell Way. 1994. *Realism Rescued: How Scientific Progress Is Possible*. London: Duckworth.
- Axelrod, Robert, ed. 1976. *Structure of Decision: The Cognitive Maps of Political Elites*. Princeton, N.J.: Princeton University Press.
- Bartels, Larry M. 1996. Pooling Disparate Observations. *American Journal of Political Science* 40 (3): 905–42.
- . 1997. Specification Uncertainty and Model Averaging. *American Journal of Political Science* 41 (2):641–74.
- Brady, Henry E. 1995. Symposium on Designing Social Inquiry. Part 2. *The Political Methodologist* 6 (2):11–20.
- Bueno de Mesquita, Bruce, and David Lalman. 1992. *War and Reason: Domestic and International Imperatives*. New Haven, Conn.: Yale University Press.

96. George and McKeown 1985, 35–36.

- Burger, Thomas. 1976. *Max Weber's Theory of Concept Formation: History, Laws, and Ideal Types*. Durham, N.C.: Duke University Press.
- Caporaso, James A. 1995. Research Design, Falsification, and the Qualitative-Quantitative Divide. *American Political Science Review* 89 (2):457–60.
- Cohen, Michael D., and Robert Axelrod. 1984. Coping with Complexity: The Adaptive Value of Changing Utility. *American Economic Review* 74 (1):30–42.
- Commins, Margaret M. 1992. From Security to Trade in U.S.–Latin American Relations; Explaining U.S. Support for a Free Trade Agreement with Mexico. Paper presented to the 17th International Congress of the Latin American Studies Association, Los Angeles, 24–27 September.
- Cronbach, Lee J., and Paul E. Meehl. 1955. Construct Validity in Psychological Tests. *Psychological Bulletin* 52 (4):281–302.
- Cyert, Richard M., and James G. March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs, N.J.: Prentice-Hall.
- Diesing, Paul. 1991. *How Does Social Science Work? Reflections on Practice*. Pittsburgh, Pa.: University of Pittsburgh Press.
- Eckstein, Harry. 1975. Case Study and Theory in Political Science. In *Handbook of Political Science*. Vol. 7, *Strategies of Inquiry*, edited by Fred I. Greenstein and Nelson W. Polsby, 79–137. Reading, Mass.: Addison-Wesley.
- Ekeland, Ivar. 1988. *Mathematics and the Unexpected*. Chicago: University of Chicago Press.
- Elster, Jon. 1983. *Explaining Technical Change: A Case Study in the Philosophy of Science*. Cambridge: Cambridge University Press.
- Ericsson, K. Anders, and Herbert A. Simon. 1984. *Protocol Analysis: Verbal Reports as Data*. Cambridge, Mass.: MIT Press.
- Fearon, James D. 1991. Counterfactuals and Hypothesis Testing in Political Science. *World Politics* 43 (2):169–95.
- George, Alexander L., and Timothy J. McKeown. 1985. Case Studies and Theories of Organizational Decision-Making. In *Advances in Information Processing in Organizations*, edited by Robert F. Coulam and Richard A. Smith, 21–58. Greenwich, Conn.: JAI Press.
- Gigerenzer, Gerd. 1991. From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology. *Psychological Review* 98 (2):254–67.
- Gibbs, David N. 1994. Taking the State Back Out: Reflections on a Tautology. *Contention* 3 (3):115–37.
- Glymour, Clark, Richard Scheines, Peter Spirtes, and Kevin Kelly. 1987. *Discovering Causal Structure: Artificial Intelligence, Philosophy of Science, and Statistical Modeling*. Orlando, Fla.: Academic Press.
- Gooding, David. 1992. The Procedural Turn; or, Why Do Thought Experiments Work? In *Mimesota Studies in the Philosophy of Science*. Vol. 15, *Cognitive Models of Science*, edited by Ronald N. Giere, 45–76. Minneapolis: University of Minnesota Press.
- Hekman, Susan J. 1983. *Weber, the Ideal Type, and Contemporary Social Theory*. Notre Dame, Ind.: University of Notre Dame Press.
- Hempel, Carl Gustav. 1965. *Aspects of Scientific Explanation*. New York: Free Press.
- . 1966. *Philosophy of Natural Science*. Englewood Cliffs, N.J.: Prentice-Hall.
- Jackman, Simon, and Gary N. Marks. 1994. Forecasting Australian Elections—1993, and All That. *Australian Journal of Political Science* 29:277–91.
- Kaplan, Abraham. 1964. *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler.
- King, Gary, Robert O. Keohane, and Sydney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, N.J.: Princeton University Press.
- . 1995. The Importance of Research Design in Political Science. *American Political Science Review* 89:475–81.
- Laitin, David. 1995. Disciplining Political Science. *American Political Science Review* 89:454–56.
- Lazarsfeld, Paul F., and Morris Rosenberg. 1955. *The Language of Social Research*. Glencoe, Ill.: Free Press.
- Leamer, Edward E. 1994. *Sturdy Econometrics*. Brookfield, Vt.: Edward Elgar.

- Levi, Isaac. 1984. *Decisions and Revisions: Philosophical Essays on Knowledge and Value*. Cambridge: Cambridge University Press.
- Lijphart, Arend. 1975. *The Politics of Accommodation: Pluralism and Democracy in the Netherlands*. 2d ed. Berkeley: University of California Press.
- Little, Daniel. 1991. *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science*. Boulder, Colo.: Westview Press.
- Magee, Stephen P. 1980. Three Simple Tests of the Stolper-Samuelson Theorem. In *Issues in International Economics*, edited by Peter Oppenheimer, 138–53. London: Oriol Press.
- Miller, Richard W. 1987. *Fact and Method: Explanation, Confirmation, and Reality in the Natural and the Social Sciences*. Princeton, N.J.: Princeton University Press.
- Mitchell, J. Clyde. 1984. Case Studies. In *Ethnographic Research: A Guide to General Conduct*, edited by R. F. Ellen, 237–41. Orlando, Fla.: Academic Press.
- Popper, Karl. 1959. Prediction and Prophecy in the Social Sciences. In *Theories of History*, edited by Patrick Gardiner, 276–85. Glencoe, Ill.: Free Press.
- . 1968. *The Logic of Scientific Discovery*. New York: Harper and Row.
- Putnam, Hilary. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.
- Rescher, Nicholas. 1970. *Scientific Explanation*. New York: Free Press.
- Ringer, Fritz. 1997. *Max Weber's Methodology: The Unification of the Cultural and Social Sciences*. Cambridge, Mass.: Harvard University Press.
- Rogowski, Ronald. 1995. The Role of Theory and Anomaly in Social-Scientific Inference. *American Political Science Review* 89 (2):467–70.
- Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, N.J.: Princeton University Press.
- Signorino, Curtis S. 1998. Statistical Analysis of Finite Choice Models in Extensive Form. Paper presented at the 94th Annual Meeting of the American Political Science Association, September, Boston.
- Snyder, Richard C., H. W. Bruck, and Burton Sapin. 1954. Decision-making as an Approach to the Study of International Politics. Foreign Policy Analysis Series No. 3. Princeton, N.J.: Organizational Behavior Section, Princeton University.
- Stinchcombe, Arthur L. 1978. *Theoretical Methods in Social History*. New York: Academic Press.
- Taber, Charles S. 1992. POLI: An Expert System Model of U.S. Foreign Policy Belief Systems. *American Political Science Review* 86 (4):888–904.
- Tanner, W. P., Jr., and J. A. Swets. 1954. A Decision-Making Theory of Visual Detection. *Psychological Review* 61 (6):401–409.
- Tetlock, Philip E., and Aaron Belkin. 1996a. Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives. In *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*, edited by Philip E. Tetlock and Aaron Belkin, 1–38. Princeton, N.J.: Princeton University Press.
- Tetlock, Philip E., and Aaron Belkin, eds., 1996b. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. Princeton, N.J.: Princeton University Press.
- Toulmin, Stephen E. 1972. *Human Understanding*. Princeton, N.J.: Princeton University Press.
- van Deth, J. W. 1995. Comparative Politics and the Decline of the Nation-State in Western Europe. *European Journal of Political Research* 27:443–62.
- Van Evera, Stephen. 1996. Guide to Methodology for Students of Political Science. Working Paper, Defense and Arms Control Studies Program. Cambridge, Mass.: MIT.
- Walton, John. 1992. Making the Theoretical Case. In *What Is a Case? Exploring the Foundations of Social Inquiry*, edited by Charles C. Ragin and Howard S. Becker, 121–37. New York: Cambridge University Press.
- Western, Bruce, Simon Jackman, and Gary N. Marks. 1994. Bayesian Inference for Comparative Research. *American Political Science Review* 88:412–23.