RESEARCH DESIGN, FALSIFICATION, AND THE QUALITATIVE-QUANTITATIVE DIVIDE

JAMES A. CAPORASO *University of Washington*

While disagreement may be more interesting than agreement, I preface my remarks by saying that I am broadly sympathetic to the arguments of *Designing Social Inquiry* by King, Keohane, and Verba. The authors have tried, with considerable success, to provide unifying principles and research strategies for qualitative and quantitative research. The central argument is that the rules for descriptive and causal inference have been unnecessarily restricted to quantitative designs. Good qualitative designs also profit from variance in the explanatory variables, proper measurement strategies, and control of extraneous variation. While there are legitimate differences between qualitative and quantitative research, KKV debunk the polarized images of the systematic quantitative researcher reducing politics to rows of equations versus the qualitative scholar giving solo performances with nonreproducible insight and *Fingerspitzengefühl*. In short, the authors' "reconciliation project" provides a methodological bridge connecting qualitative and quantitative research. Some may see reconciliation as conquest, since the unity achieved does take place on a particular turf, with particular standards. Yet the results are impressive.

KKV place strategies of inquiry, or research design, at the center of the book. In teaching research methods, what I find most useful are books that encourage us to construct research strategies that bring some probing value to our questions. Some books see all problems as resolvable by sophisticated statistical manipulation. The spirit of this book is quite different. With design at the center, the central issues are choosing appropriate units, ensuring variation in explanatory variables, and controlling for confounding influences. Weak (or indeterminate) designs cannot be salvaged by clever data analysis (p. 120). Research that is structurally defective in the sense that there is no variation in explanatory variables (or more explanations than observations) is doomed to fail, no matter how insightful the analyst. Without an appropriate organizing structure, additional data and even sophisticated analysis can tell us little.

My differences with the work are framed by the fact that my methods education was heavily influenced by two people never mentioned in this book: Hubert Blalock and Donald Campbell. Blalock's sociological contributions are premised on the notion that the ecology of social science is characterized by many independent variables, all intercorrelated and imperfectly measured, with feedback effects from the dependent variable. Many methodological problems are diagnosed in this notion: overdetermination, multicollinearity, error in variables (fallible measures), and endogeneity. Each of these problems is dealt with by KKV. Thus, while Blalock's enormous contributions are not noted (perhaps because they have been assimilated into modern statistical theory), the spirit of his work is well represented in this book. On the other hand, Campbell's quasi-experimental orientation is not only omitted but rejected early in the book (p. 7). Since Campbell and Stanley's *Experimental and Quasi-experimental Designs for Research* (1963) was—and to some extent still is—an important reference for empirical researchers spanning sociology, education, psychology, political science, and policy analysis, I take up KKV's categorical rejection of this type of research.

My reactions focus on three points: (1) the nature of qualitative research, (2) the meaning of *falsificationism*, and (3) the usefulness of quasi-experimental designs. The first two points are largely agreeable elaborations of positions taken in the book. The third represents a disagreement.

The Nature of Qualitative Research

KKV strenuously argue that the same rules of inference apply to qualitative and quantitative research. While I am inclined to agree, it is because I share the authors' definition of qualitative research as research based on *in-kind* rather than *in-degrees* differences. With this distinction, variance can be of two types: across categories (e.g., types of government, gender) and across quantities of the same variable (income, degree of labor repression). In measurement theory, qualities are represented as nominal variables, and quantities, as ordinal, interval, and ratio measures. Qualitative variation is not variation in magnitude, quantitative variation is. This characterization shows that it is not really numbers that are at issue (nominal measures are assigned numbers, too) but the issue of magnitude versus quality.

With this definition of qualitative research in place, the authors easily show that a sound qualitative research strategy requires attention to the same rules of inference as a quantitative strategy ("if x, then y" is not logically different from "as x increases, y increases"). But qualitative work can be conceived differently and in ways that are more resistant to KKV's reconciliation project.

For some, qualitative research signifies something different from explanations of in-kind variation. Indeed, the whole idea of systematic research harnessed to the goal of explanation is put into question. Thick description and interpretation may serve as ends, not merely as spadework preparatory to explanation. Scholars may be interested in empathetic understanding, the interpretation of meanings, and detailed investigation of single (nonvarying) cases. Some of the book's arguments (e.g., the rules of descriptive inference) still hold. Others (e.g., the rules of causal inference) are less relevant, despite the authors' attempt to square Geertzian analysis with their project (pp. 38–41).

A related point is that KKV's arguments about differences and similarities between qualitative and quantitative research take place in a variable-centered world. This is not the only starting point. A variable-centered approach is already one in which variable properties have been abstracted from things, concrete names, and places. In the classroom, I find that the most difficult argument to make is not the unity of qualitative and quantitative research once a variable-centered model has been accepted but how one makes the transition from instances and concretely experienced sense data to variables. On this crucial issue, I know of no methodological guides. Between "Jumbo the elephant sliding down a grassy hill at Gasworks Park" and "a certain mass moving down an inclined plane with a given coefficient of friction" there is a gap.[1] Neither logic nor observation obliges us to accept the second statement once we accept the first. Yet the leap has to be taken to reach the abstract world of variables. Hitler's Reich as totalitarian regime; Austria, Norway, and Sweden as small, open, corporatist social democracies; and Brazil, Argentina, and South Korea as late-developing bureaucratic-authoritarian polities all represent examples of concept formation not forced by deductive or inductive logic (assuming one believes in the latter).

It may be that the urge to abstract is irresistible. Campbell was fond of arguing that theory and concept formation are "hard-wired in our retina," reflecting the absence of theoretical innocence in our sensory equipment. In the end, I am in agreement with the authors' but they have more careful work to do before the quantitative-qualitative gaps are bridged—and some will never be.

The Meaning of Falsificationism

Science proceeds not only by hypothesis and conjecture but also by relentless attempts to reject our own theories. This does not mean that we hope our theories are wrong but that we believe them to the extent they survive difficult tests. The falsificationist perspective is important because it emphasizes the pruning-editing-winnowing side of science (Campbell and Stanley 1963, 35) in contrast to the confirmatory perspective that attempts to assess hypotheses by discovering confirming instances.

I accept KKV's starting point—that confirmation and rejection are logically asymmetric. But the authors tend to see falsificationism in terms of deriving many implications of a theory, to increase the theory's exposure to evidence. The problem with this criterion is that there is no guarantee (or greater likelihood) that the additional derivations will be any riskier than the initial hypotheses. A developed falsificationist perspective would add three points.

First, we should consider which of our theories' implications are least likely to be confirmed if the theory is not true. This is another way of asking what the most distinctive explanatory-predictive content of the theory is. To predict that it will rain in Seattle during November is not risky.[2] Similarly, to explain why strong states win out over weaker ones (using standard definitions of capabilities) is not risky. Anomalies are those outcomes which go against the grain. They are not what our prevailing intuitions and theories would have us believe. A recent case study of bargaining outcomes illustrates the point. Lisa Martin and Kathryn Sikkink (1993) compared U.S. pressure on Argentina and Guatemala to improve their human rights records. The puzzle motivating the study was that Argentina (larger, more powerful, more autonomous) caved in to U.S. pressure, while Guatemala successfully resisted (p. 332). The authors' theory relied on a number of factors, among which was the strength of transnational human rights lobbies and organizations. Their theory is riskier in the sense that it explains an outcome different from what we would otherwise expect.

The intuition embedded in this example is that many theories are compatible with a particular outcome. Outcomes are overdetermined. In this sense, confirmation is highly equivocal in theoretical terms. Theoretically consistent outcomes are a necessary but hardly sufficient aspect of a good research strategy. Instead of finding data that correspond to theory, why not first ask which of the outcomes implied by the theory are least likely to be true if the theory is not true? This question forces us to find the "reduced set" of outcomes that are most distinctively implied by the theory. The art of good research design is to identify those cases which can tell us the most in terms of distinct theoretical content.

The second point is derivative. The authors argue that testing our theories in alternative settings is a good idea. But what guides do we have for how to conduct these tests? The falsificationist perspective provides a criterion. Elaborate the implicative core of the theory in such a way that the multiple tests reduce the set of rival hypotheses (competing explanations) as much as possible. Carrying out the same test in the same setting provides little additional support for the theory. The same test in a different setting expands the scope of a theory and may add confirmatory weight if additional factors thought to influence the outcome are taken into account. But this is hit or miss. The researcher should isolate the set of implications that has the greatest nonoverlap in competing explanations. If a theory holds across highly diverse settings, this is more impressive than confirmation under similar conditions. The presumption is that rival explanations have a greater opportunity to register their influence under diversity. This point is crucial to the most different system design (Przeworski and Teune 1970). Using Durkheim's theory of suicide as an example, Stinchcombe convincingly outlines the logic of this procedure (1968, 15–22).

Third, KKV could improve their argument by drawing out the links between falsification, quantitative reasoning, and the theoretical development of our discipline. In part, this relates to Rogowski's argument about strong theory (in the present symposium). One advantage of quantitative research is that it generates more precise predictions (often numerical values or ranges within which such values fall), which increase the difficulty of a test. Much of social science is at least implicitly about expectational standards. In statistics, one weak standard is the nondirectional null model. Findings departing from chance expectations in any direction are sufficient to reject the model. Another standard is provided by substantive theory. Do the results differ from what one expects after taking into account x, y, and z (this does not rule out a differently specified null model)?

My overall point is as follows. Improvements in measurement accuracy, theoretical specification, and research should yield a smaller range of allowable outcomes consistent with the predictions made. Cumulative improvements in knowledge should make our predictions riskier, more falsifiable. This seems to me to happen all too rarely in political science, in part because we are anxious to move on to new topics (skimming the cream from little investigated areas) and in part because we are more interested in presenting a "fresh look" or "new paradigm" than in using our collective achievements to define novel yet cumulative departures. We rarely report results in incremental (value-added) fashion, as additions to the existing capital stock. Instead, our results are presented as separate "findings." We are confronted with perverse incentives. To take seriously the *acquis* of social sciences has the effect of increasing the difficulty of our tests in the sense of raising the "observational hurdles" required to accept a hypothesis (Meehl 1967, 103). Conversely, to ignore past achievements makes our hypotheses easier to accept—but at great costs in terms of lowered standards and cumulation of knowledge.

Quasi-experimental Analysis

The experimental method is often considered too narrowly as a battery of techniques applicable in a laboratory but irrelevant to the "real" world. By elevating experimental *procedures* over its *logic*, we lose the opportunity to learn what experimentation implies for ex post facto research. In broad terms, the biggest achievement of experimental design is the preexperimental equivalence of groups through random assignment (Campbell and Stanley 1963, 2). The power of random assignment is often not fully appreciated in social science research. The important distinction between random assignment and random sampling is elided. Random sampling does not solve the problems of drawing inferences when numerous causal factors are associated with outcomes.[3] By contrast, the capacity of the experimenter to assign units (usually people) to treatment and control groups neutralizes nearly all subject-centered threats to validity.[4] Experimental control over the "how much" of x assures adequate variation in the independent variables. Control over the timing (the *when*) of exposure implies a solution to the endogeneity problem (since values of the independent variable can occur independently of the dependent variable).

The logic of experimental research provides guidelines in ex post facto settings. The random assignment technique directs us to find ways to control extraneous variables, for example, by using stratified designs that reduce variation in confounding variables or by building in variation and doing partial correlation and regression analysis. The manipulation procedure translates into the scheduling of units and observations so as to assure variation on the independent variable. Ex post facto research is the "continuation of experimental logic through other means." On this important philosophical point, I do not think there are differences with KKV. Why, then, do they reject quasi-experimental analysis?

They say "We reject the concept, or at least the word, 'quasi-experiment' " (p. 7n.). They further state that "investigator control over observations and values of the key causal variables" is the determining factor in deciding whether something is an experiment. Two points need to be made. First, researcher control over values of the independent variables is not enough to define experimentation. The ability to assign randomly is also crucial as is experimental isolation (a lab). In a pure experiment, the three properties go together. Without manipulating the independent variables, we cannot be sure that hypothesized effects will have a chance to occur. Without random assignment and laboratory isolation, we cannot be sure we would detect such effects even if they did occur.

The second point is more nuanced. If KKV mean that quasi-experimental designs do not represent a logically distinct category, I agree. However, the numerous designs pioneered by Campbell and Stanley (1963) were possible because they "unpacked" three properties that merge in pure experiments (manipulation of the independent variable, random assignment, and lab setting). These properties were then combined in various ways to produce various hybrid designs (see Achen 1986; Cook and Campbell 1979). For example, a field experiment allows for some ability to manipulate the independent variable but no control over random assignment and setting. Other designs allow random assignment (e.g., of court cases to different processing procedures) but no ability to affect the independent variables.

Quasi-experimental designs strive for three things: (1) natural settings in which abrupt variation occurs in independent variables, (2) some natural controls as one is likely to find when adjacent units of analysis experience different "treatments," and (3) a checklist of concepts and techniques to use to address internal and external validity. The interrupted time-series design and multiple control-series design provide valuable controls in many situations of interest to the political science researcher. Indeed, Campbell's (1969) "Reforms as Experiments" is a model for the policy analyst attempting to unravel complex interactions in the policy process. Because these designs are meant for a world of biased selection, differential exposure to threats to validity, measurement error, and researcher wish fulfillment, I find them helpful.

Finally, I find Campbell and Stanley's distinction between internal and external validity helpful—indeed almost necessary—to assessment of the value of any design. Internal validity is fundamental; it is the starting point: "Did x have an impact on y?" External validity asks the question, "To what other groups, units, populations can these results be extended?" The former question concerns us all, but the second is often reduced to a sampling instability issue, which it is not. It is a contextual (hence theoretical) issue: "If it is true in the Bronx, will it also hold in Cook County?," rather than "How are elements 1–30, sampled randomly from a population, different from 31–60, sampled from the same population?" The distinction is all the more helpful in that it maps nicely onto main effects versus interaction effects (main effects are threats to internal validity, interactions, to external validity). Thus, to take one example, if selection biases operate independently of one's hypothesized causal variable, it is a threat to internal validity; if these same selection factors interact with the causal variable, it is a threat to external validity. In the former case, x (causal variable) had no impact. In the second case, x did have an impact, but only because of its conjunction with selection factors.

Let me bring my comments back to the starting point. I regard this as a fine book that will be widely used in methodology courses. I am persuaded that much of what goes under the label of qualitative research is concerned with explanation and causality and must therefore be attentive to the main arguments of this book. By outlining a research strategy applicable in both descriptive and causal settings and relevant to qualitative and quantitative research, KKV hold the promise of unifying previously fragmented parts of our discipline. At the very least, *Designing Social Inquiry* encourages us to talk to one another and to learn more precisely where our differences lie.

Notes

I am grateful for the comments of Anthony Gill.

1. I believe this phrase, or something like it, is attributable to Sir Arthur Eddington.

2. Prediction, unlike forecast, requires a theoretical structure. The logical structure of a prediction is, "If x occurs, then y will occur." By contrast, a forecast simply asserts that y will occur in the future as a result of extrapolation ("casting forth") of y.

3. It does to some degree. Random sampling helps to eliminate chance as a factor explaining an association. However, if many variables are correlated (confounded) in the population, random sampling will only provide a more accurate assessment of this confounding. It will not control these variables in the sense of neutralizing their influence.

4. By *subject-centered* threats to validity, I mean those differences among groups which are the result of differences among the individuals that compose the groups. Random assignment does not control for differences in the environment of the groups (differences irrelevant to the treatment) or variation that the experimenter may introduce by treating the two groups differently.

TRANSLATING QUANTITATIVE METHODS FOR QUALITATIVE RESEARCHERS: THE CASE OF SELECTION BIAS

DAVID COLLIER *University of California, Berkeley*

King, Keohane, and Verba's *Designing Social Inquiry: Scientific Inference in Qualitative Research* (KKV) is an ambitious attempt to translate for the qualitative researcher a series of insights derived from quantitative methodology. The authors build on the basic framework of regression analysis—which has been enriched over the past two decades with innovations introduced by econometricians and statisticians—to make recommendations about how qualitative analysts should confront a variety of methodological problems. KKV are strongly committed to the premise that the underlying logic of quantitative and qualitative research is basically the same (p. ix). At the same time, they are attentive to the specific dilemmas that arise in actually carrying out qualitative research, and they provide many useful examples and employ clear, nontechnical language. This is a book that moves the discussion forward, and therefore merits close attention.

Selection bias is one of the important topics into which quantitative methodologists have recently offered significant new insights. Hence, the assessment of KKV's treatment of this topic[1] provides a useful window for evaluating their effort to transpose complex issues of quantitative method to the sphere of qualitative research. This is also an interesting topic to address because KKV are centrally concerned with selection bias resulting from deliberate selection by the investigator. Their recommendations are consequently of special importance: if their diagnosis is correct, a small improvement in methodological self-awareness can yield a large improvement in scholarship. Finally, KKV's recommendations merit examination precisely because they are quite emphatic. Given their emphatic character, readers may desire assurance that they are, in fact, receiving sound advice.

The question of how to situate the problem of selection bias in relation to a spectrum of other methodological and theoretical issues is not an easy one. At one pole, in discussions of selection bias in quantitative *sociology*, one finds an influential article suggesting that the impact of selection bias is not as serious as has been believed, that efforts to introduce statistical corrections for selection bias may create more problems than they solve, and that among the many problems of quantitative analysis, this one does not merit special attention (Stolzenberg and Relles 1990). In the present context, the appropriate point of entry into the problem is different. First, in the examination of selection bias in qualitative *political research*, it is much more difficult to assess its precise impact, so that conclusions about its importance are inevitably more tentative. Second, KKV's recommendations about selection bias are centrally concerned with deliberate selection by the investigator. Hence, relevant corrections do not involve statistical procedures, but rather basic choices about case selection that are relatively easy to achieve. Third, one of KKV's central goals is to explore the interrelations among a series of different methodological problems. They are not singling out selection bias as a paramount problem: in their view, it is one of many.

I conclude that KKV offer useful recommendations regarding selection bias. Yet they subsume under this term various issues with which qualitative researchers may already be familiar, but under different labels. Obviously, in a complex field it is common to find that a given phenomenon is named in different ways. For example, what was probably the first paper ever published on selection bias referred to it as a problem of spurious correlation (Berkson 1946, 51). Nevertheless, the overlap of labels raises the question of whether KKV's methodological insights really offer something new to the qualitative analyst.

In fact, some of the important recommendations offered by KKV can just as well be viewed not as insights derived from advanced quantitative methods, but rather as part of a long-standing effort to encourage qualitative scholars to be more methodologically and theoretically aware of which cases they are analyzing. This self-consciousness can also be encouraged by insistently posing a question that, according to the traditional lore of the comparative politics subfield, should often be asked at doctoral dissertation defenses: "What is this a case of?"

Many issues that underlie this question and that are highly relevant to KKV's discussion of selection bias have previously been raised in discussions of the comparative method, that is, the branch of methodology concerned with the systematic, qualitative analysis of relatively small numbers of cases (a "small N").[2] In assessing methodological claims about selection bias, it is useful to take these earlier discussions as a base line. With regard to issues of case selection, they include the ongoing evaluation of J. S. Mill's methods of experimental inquiry, the related distinction between "most similar" and "most different" systems designs, and a new perspective on case selection in small-N studies arising from counterfactual analysis. Regarding the problem of applying concepts and indicators across diverse contexts, an issue that arises in KKV's discussion, relevant insights from work on comparative method include the traditional concern with "conceptual stretching" and the use of "system-specific," as opposed to "common," indicators. Regarding the issue of generalization, relevant insights include the argument that it may at times be appropriate for scholars to limit severely the scope of generalizations from a given set of cases, an argument with important implications for what it means to think of these cases as being selected from a larger population.

The following discussion devotes central attention to the relationship between KKV's arguments and these familiar issues in the field of comparative method. After presenting an overview of selection bias as a methodological problem, along with some initial caveats, I shall consider the potential strategy of avoiding selection bias through random sampling, examine KKV's treatment of selection bias in descriptive inference, and then explore two issues that arise in their discussion of causal inference. Finally, I shall consider how qualitative researchers understand the role of generalization in social inquiry and the implications for selection bias. Although the discussion generally supports the thesis of a convergence in the logic of quantitative and qualitative methods, it is evident that qualitative researchers at times have different priorities in designing research.

Overview of Selection Bias

Selection bias is commonly understood as occurring when the *non*random selection of cases results in inferences, based on the resulting sample, that are not statistically representative of the population. The focus of the present discussion is on selection bias deriving from deliberate selection by the investigator.[3] A common problem arising from such selection is that it may overrepresent cases at one or the other end of the distribution on a key variable. When this specifically involves the selection of cases that fall above or below a particular value on the distribution of that variable, it is referred to as a form of *truncation*.[4]

The statistical insight crucial to understanding the consequences of such selection is the observation that selecting cases so as to constrain variation toward high or low values of the *dependent* variable tends to reduce the slope estimate produced by regression analysis, whereas an equivalent mode of selection on the *explanatory* variable does not have this effect. If, for example, the analyst selects a sample that is truncated to include only cases that have higher scores on the dependent variable, the sample will tend to overrepresent cases *above* the regression line that is derived from the full data set. This mode of selection therefore gives disproportionate weight in the calculation of the slope to cases for which factors *in addition to* the principal explanatory variable play an important role in producing higher scores on the dependent variable (or lower scores, in the case of a negative relationship). As a consequence, unless the investigator can identify missing variables that explain the position of these cases above the regression line, the bivariate relationship within this subset of selected cases will appear to be *weaker* than in the larger set of cases from which they are selected. A corresponding effect occurs if selection is biased toward the lower end of the dependent variable. By contrast, if selection is biased toward the higher or lower end of the *explanatory* variable, then, as long as the underlying relationship is linear, the expected value of the slope will not change (although it may vary in particular data sets).

This asymmetry is the basis for warnings about the hazards of "selecting on the dependent variable." This expression refers not only to the deliberate selection of cases according to their scores on this variable but to any mode of selection correlated with the dependent variable (i.e., tending to select cases that have higher, or lower, values on that variable) once the effect of the explanatory variables is removed (pp. 138–39). If such a correlation exists, causal inferences will tend to be biased.

Initial Caveats

Selection bias deriving from truncation is in some respects *less* serious—and in other respects *more* serious—than might initially appear to be the case. First, as KKV note, on average it will lead analysts to underestimate the strength of causal effects, and they suggest that estimates derived from the sample may be understood as a "lower bound" in relation to the true causal effect (pp. 130, 139). In qualitative research, where the inductive character of the analysis may entail a kind of ad hoc hypothesizing that can lead to "overfitting" the data, this kind of constraint might be a useful corrective.

Second, the basic asymmetry that calls for warnings about selecting on the dependent variable applies to the slope, but *not* the correlation. The correlation is a "symmetric" measure, and constraining variance on the explanatory variable *can* affect the correlation, potentially making researchers even more vulnerable to selection bias. In quantitative research the slope is more widely used for causal inference than the correlation, in part for this reason. Yet for some specific kinds of causal analysis, the correlation (or the standardized slope, which is a closely related coefficient) is more appropriate (Achen 1982, 74–76). To the extent that it is, the warning about selection bias must be extended to the problem of selecting on the explanatory variable. It remains a topic for further investigation whether the intuitive assessment of causal relationships by qualitative researchers should be understood as more nearly analogous to the slope or the correlation—and hence whether they should be concerned with selection on the independent variable, as well as the dependent variable.

Third, advice about the distinctive problems associated with the dependent variable should be qualified in another sense as well. Selecting on the explanatory variable can affect the slope estimate under some circumstances.[5] With a bivariate linear relationship, sampling toward the high or low end of the explanatory variable does not affect the expected value of the slope. However, if the underlying relationship is nonlinear, selecting different parts of the distribution on the explanatory variable can yield different slope estimates. This is *not* due to selection bias, but it *does* involve the more general issue that case selection can influence findings. Hence, this more generic form of the problem does not arise only with selection on the dependent variable.

The Option of Random Sampling

Given that selection bias is conventionally understood as deriving from the nonrandom selection of cases, one option for the qualitative researcher might be to engage in random sampling and thereby (hopefully) avoid the whole problem. Yet KKV suggest that random sampling is not the solution.

In the statistical literature, it has been traditional to treat selection bias and sampling error that results from random selection as separate issues.[6] By contrast, KKV argue that sampling error can produce selection bias. They offer a hypothetical small-N example of three cases that have high, medium, and low scores on the dependent variable, from among which only two cases will be selected. They discuss three alternative selection rules and point out that of the three combinations of cases that can result, two will constitute a sample that is biased either toward the high or low end of the dependent variable. They conclude that "since random selection of observations is equivalent to a random choice of one of these three possible selection rules, *random selection of units in this small-n example will produce selection bias with two-thirds probability!*" (p. 126, italics added).

KKV's statement that sampling error produces selection bias with two-thirds probability entails a usage that will surprise many readers familiar with standard statistical terminology.[7] However, their goal here is to point out that with a small N and only *one* random sample, the same kind of error that is associated with selection bias is quite likely to occur. Despite the presumed virtues of random sampling, with a very small N it can produce the same kind of error that is identified with selection bias deriving from a truncated sample. Hence, they argue that the investigator is much better off engaging in a carefully planned form of nonrandom selection (pp. 125–26). This is good advice.

Descriptive Inference

KKV's discussion of selection bias in descriptive inference includes two interesting examples that will be examined here. The first concerns case studies of deterrence in international relations that focus primarily on deterrence failure (pp. 134–35). This focus has an important consequence: inferring an overall success rate of deterrence from such case studies is a big mistake. Achen and Snidal, who are cited by KKV, give the example of a prominent social scientist who used a study of 12 cases of conventional deterrence to reach the surprising conclusion that conventional deterrence fails 83.3% of the time, i.e., 10 out of 12 cases (1989, 162). Yet it does not require a deep knowledge of modern regression analysis to grasp this problem, which can readily be understood within the conventional framework of the comparative method. Thus, it is essential for researchers to keep in mind how and why the cases were originally selected, and scholars following any variant of what J. S. Mill called the "method of agreement" (1974, 388–90) should not use the cases thereby selected as a basis for descriptive generalizations concerning the dependent variable.

In another example concerned with descriptive inference, KKV consider a hypothetical assessment of support for the Liberal party in New York State, based on votes for candidates endorsed by the Liberals in elections for the State Assembly. Because the Liberals do not endorse candidates in many districts where they believe they will lose, such a study would provide an inadequate assessment of support for the party. Thus, KKV argue, descriptive inferences derived from this study would suffer from selection bias (p. 135).

Within the tradition of work on the comparative method, qualitative researchers might more readily understand this example as raising issues linked to the problem of conceptual stretching and also as a complex measurement problem that may call for system-specific indicators. Central issues here are the definition of a political party and the problem of measuring party support. If one accepts a Sartori-type definition, according to which parties are political groups that present candidates in elections to public office (1976, 64), then the Liberals are not acting as a party in those districts where they do not present candidates. This doubtless construes the definition too narrowly, but it is appropriate to emphasize that in multiparty, as opposed to two-party, systems, the practice of not running candidates in selected districts is more common, and the question of what is and is not a party is often more complex. Some conceptual reflection would seem essential here. Second, analysts who use elections as a source of data on political support would normally devote close attention to which candidates run in a particular year, because alternative candidates for a given party will generate different profiles of support. Hence, it is *not* the case that elections usually provide a straightforward measure of party support. If elections are, nonetheless, used to measure party support, some other system-specific measure should be used in the districts where the Liberals do not run candidates.

Thus, one could argue that in the case of the Liberal party in New York, using the popular vote to measure party support will produce descriptive inferences that suffer from selection bias. Alternatively, this can be viewed as a problem of conceptualization and of developing system-specific measures, topics not covered in KKV's book but that are familiar themes of comparative method.

Mild Versus Extreme Selection Bias

An important element in KKV's discussion of selection bias in causal inference is their distinction between a "milder" form of selection bias, that results from merely constraining variation on the dependent variable, and what they refer to as an "extreme" form that results from selecting only *one value* on the dependent variable—according to them, a grave mistake (p. 130). This strategy of selecting only one value may be adopted by scholars who are analyzing an outcome of exceptional interest (e.g., deterrence failure, revolutions, or high growth rates) and who wish to focus only on this outcome, out of a belief that they will thereby achieve greater insight into the phenomenon itself and into its causes. Alternatively, they may be dealing with an outcome about which previous theories, conceptualizations, measurement procedures, and empirical studies provide limited insight. Hence, they may be convinced that a carefully contextualized analysis of one or a few cases of the outcome will be more analytically productive than a broader study that compares cases of its occurrence and nonoccurrence.

This distinction between mild and extreme selection bias conflates two distinct issues with which KKV are concerned. The first is the core issue of investigator-induced bias, involving the fact that the greater the constraint on the variance of the dependent variable, the more severe the bias in inference is likely to be. The second issue is that at the outer limit, when variance on the dependent variable disappears and the investigator focuses on only one outcome on that variable, a shift to a different kind of research design has occurred. Where there is no variance, selection bias certainly may be present in that the sample may well overrepresent cases for which factors other than the main explanatory variable play an important role in accounting for their higher scores on the dependent variable. But that outcome can more usefully be treated as a different issue from the switch in research design.

Selecting One Value on the Dependent Variable

Some of the strongest criticisms regarding selection bias have been leveled against studies that focus on a single outcome on the dependent variable, which KKV characterize as an extreme form of selection bias. In such studies, according to them, "nothing whatsoever can be learned about the causes of the dependent variable without taking into account other instances when the dependent variable takes on other values" (p. 129). They suggest that the need for variation on the dependent variable "seems so obvious that we would think it hardly needs to be mentioned" and that research designs lacking such variation "are easy to deal with: avoid them!" (pp. 129, 130).

On the one hand, given that they advance a definition of "causal effect" that requires the observation of at least two different values on the dependent variable (pp. 81–82), within their own framework their position can be seen as making sense. Yet other perspectives on this question are available. In the field of comparative method, a traditional way of thinking about this design is in terms of J. S. Mill's method of agreement, a perspective that KKV note (p. 134) but do not develop. This label is used by Mill because all cases under investigation "agree" on the dependent variable. Many authors have examined the strengths and weaknesses of this design, and a standard view, expressed by Mill himself, is that this design fails to provide a positive demonstration of causation, and rather should be viewed as a "method of elimination," which can exclude causal factors if they are consistently not present when a given outcome occurs (1974, 392). As Jervis has suggested, this design may serve to assess the necessary conditions of a given outcome, or, to put it more precisely, to eliminate some hypothesized necessary conditions (1989, 194). In this sense, KKV's assertion that this type of design makes it "impossible to evaluate any individual causal effect" (p. 134) seems incomplete: it can serve to *eliminate* some hypothesized causes, which can be a useful first step in causal analysis.

A second perspective on this design becomes relevant if analysts compare cases that are matched on the dependent variable but are extremely different from one another in other respects, in which case this can be called a "most different systems" design (Przeworski and Teune 1970). One of the merits of such a design is that the challenge of distilling a common set of explanatory factors out of this diversity can push scholars to discover new explanations that might not have emerged from the analysis of a more homogeneous set of cases (Collier 1993, 112).

A third perspective is found in Fearon's discussion of counterfactuals as a means of testing hypotheses within the framework of small-N analysis (1990, 179–80). He suggests that one can make "methodological sense" of designs with no variance on the dependent variable by recognizing that scholars can employ counterfactual analysis to introduce variance, and he goes on to present a detailed discussion of how such counterfactual analysis can be carried out.

Given these three alternative perspectives, KKV's claim that "nothing whatsoever can be learned about the causes of the dependent variable" if it does not vary within a given study would seem to be excessively limiting.

Two final observations may be made about this type of design. First, it appears that studies lacking variation on the dependent variable may be less common than scholars concerned with selection bias have sometimes implied, and studies that appear to lack it may have it after all. Michael Porter's (1990) book on industrial competitiveness, analyzed by KKV, is a case in point. The authors argue that Porter focuses on 10 nations that share a common outcome on the dependent variable of "competitive advantage," a research design that "made it impossible to evaluate any causal effect" (134). However, as Porter repeatedly points out, a central concern of his study is with explaining success and failure not at the level of nations, but rather at the level of firms and industrial sectors (e.g., pp. 28–29, 33, 69, 577, 735), of which he considers both successful and unsuccessful cases. At this level, it is incorrect to state that he lacks variance on the dependent variable. He may or may not take full advantage of the variance that is present in his study, but that is a different issue. Studies can doubtless be found in which such variance is completely lacking. Yet on closer inspection, one may at times discover some variation after all. In fact, due to a scholarly instinct for "variation seeking," analysts may have a strong tendency to find some variation on the dependent variable.

The other observation concerns the real trade-offs between these different designs. If little is known about a given outcome, then the close analysis of one or two cases of its occurrence may be more productive than a broader study, focused on positive and negative cases, in which the researcher never becomes sufficiently familiar with the phenomenon under investigation to make good choices about conceptualization and measurement, which in turn can lead to conclusions of dubious validity. On the other hand, by not utilizing the comparative perspective provided by the examination of negative cases, the researcher gives up a lot. In general, it is productive to build in a comparison of contrasting outcomes.

Samples, Populations, and the Role of Generalization

Another area in which issues of selection bias intersect with discussions of comparative method concerns the relationship among samples, populations, and the issue of generalization. Discussions of selection bias by definition presume the existence of a larger set of cases, from among which the cases under analysis have in some sense been chosen. Indeed, the claim that one is selecting cases that tend toward the high end of the dependent variable is not meaningful apart from the identification of a larger set of cases that define a range for this variable. Although in some domains of research the definition of the population is clear, in many domains, as KKV (p. 125) and others have noted, it may not be clearly specified, or its definition may be a matter of debate.

In qualitative comparative studies, a central issue in the definition of the population is a fundamental ambivalence about the process of generalization to additional cases. On the one hand, the generalization of empirical findings from an initial set of cases is a basic priority of social science research, and findings that cannot be generalized are routinely considered less important. On the other hand, over the past couple of decades a concern with sensitivity to context has been stimulated by a diverse spectrum of authors.[8] This concern has led many analysts to conclude that even important theories may sometimes apply only to limited domains. If the cases under study in fact constitute the full set of theoretically relevant cases, then an issue of selection bias in relation to a larger set of cases does not arise.[9]

Within the framework of a single piece of research, a critical issue is the appropriate balance between a legitimate process of delimiting the scope of findings and a degree of particularism that excessively limits the contribution of the study. A further issue of balance arises when other analysts become interested in the findings of a given study and wish to extend them to additional cases. On the one hand, these analysts should be alert to the limitations on the scope of claims that the original author sought to impose. On the other hand, from a different theoretical or comparative perspective these other analysts might make a different decision about the appropriate scope and seek to extend the analysis to additional cases. Hence, for them a problem of selection bias could arise that was not an issue for the original author. This kind of shift can occur in any sphere of research. However, it may have special importance in areas of qualitative comparative research in which investigators are particularly concerned about imposing constraints on the scope of their findings.

KKV's arguments about selection bias are most usefully understood as pushing qualitative researchers to think about a spectrum of selection issues. These include: (1) the core problem of selection bias that has been illuminated by advanced quantitative methods, that is, the specific impact on causal inference of certain kinds of deliberate case selection; (2) other issues already familiar to many qualitative researchers, including broader questions of case selection and their implications for various approaches to descriptive and causal analysis; and (3) additional areas in which the priorities of quantitative and qualitative researchers may sometimes be quite different—as with the issue of selecting matched versus contrasting cases and the implications for selection bias of severe restrictions on claims about scope.

These points of convergence lend support to KKV's claim that the underlying logics of quantitative and qualitative research are similar. The convergence also underscores the fact that some of their important recommendations do *not* provide qualitative researchers with new methodological insights. Finally, the divergences remind us that these two traditions sometimes make different choices about underlying trade-offs entailed in the design of research.

From the perspective of qualitative researchers, the core concern that should emerge out of these discussions can again be expressed in terms of the question, "What is this a case of?" If the debate on selection bias stimulates qualitative researchers to address this question more frequently and successfully, it will have accomplished a lot.

Notes

This work grows out of an analysis of selection bias in the field of comparative politics that I am carrying out jointly with James Mahoney. Christopher Achen, Henry Brady, Gary King, Mark Lichbach, and Laura Stoker provided exceptionally helpful comments on earlier drafts. Valuable suggestions were also made by Jake Bowers, Ruth Collier, Neil Fligstein, David Freedman, Lynn Gayle, Lincoln Moses, Michael Pretes, Thomas Romer and Mark Turner. My work on this analysis at the Center for Advanced Study in the Behavioral Sciences was supported by National Science Foundation Grant No. SBR-9022192.

1. Chapter 4 discusses the overall problem of selecting cases, or observations (as KKV call them), and one third of it is specifically concerned with selection bias.

2. Most of these issues were raised in such "classic" statements on comparative method as Bendix 1963; Lijphart 1971; Przeworski and Teune 1970; Sartori 1970; and Smelser 1976.

3. On other specific contexts in which selection bias arises, see Achen 1986; Geddes 1990, 145–48; King 1989; and Przeworski and Limongi 1992, 1993.

4. Truncation can take other forms as well; see p. 142 and Moses 1968.

5. KKV make a parallel point on p. 137.

6. See, for example, the definition of sampling error in the classic *Dictionary of Statistical Terms* prepared for the International Statistical Institute (Kendall and Buckland 1960, 255–56).

7. This formulation is stated in such a way that it appears to overlook a key theoretical idea about sampling. With truncated samples, the expected value of the estimate is biased, whereas with random samples, the expected value of the estimate is unbiased in that, if the sample is drawn a sufficient number of times, the average value of the estimates provided by the samples will be equal to the parameter one is estimating. Thus in their example it would be more helpful to say that there is a two-thirds probability that *any one sample* will contain this kind of error.

8. See Geertz 1973; Przeworski and Teune 1970; Ragin 1987; and Skocpol and Somers 1980; see also Walker and Cohen's (1985) discussion of "scope statements."

9. Moses (1968, 197) and Stolzenberg and Relles (1990, 407–08) likewise argue that problems of selection bias depend on the definition of the relevant population.

THE IMPORTANCE OF RESEARCH DESIGN IN POLITICAL SCIENCE

GARY KING, ROBERT O. KEOHANE, and SIDNEY VERBA *Harvard University*

Receiving five serious reviews in this symposium is gratifying and confirms our belief that research design should be a priority for our discipline. We are pleased that our five distinguished reviewers appear to agree with our unified approach to the logic of inference in the social sciences, and with our fundamental point: that good quantitative and good qualitative research designs are based fundamentally on the same logic of inference. The reviewers also raised virtually no objections to the main practical contribution of our book—our many specific procedures for avoiding bias, getting the most out of qualitative data, and making reliable inferences.

However, the reviews make clear that although our book may be the latest word on research design in political science, it is surely not the last. We are taxed for failing to include important issues in our analysis and for dealing inadequately with some of what we included. Before responding to the reviewers' most direct criticisms, let us explain what we emphasize in *Designing Social Inquiry* and how it relates to some of the points raised by the reviewers.

WHAT WE TRIED TO DO

Designing Social Inquiry grew out of our discussions while coteaching a graduate seminar on research design, reflecting on job talks in our department, and reading the professional literature in our respective subfields. Although many of the students, job candidates, and authors were highly sophisticated qualitative and quantitative data collectors, interviewers, soakers and pokers, theorists, philosophers, formal modelers, and advanced statistical analysts, many nevertheless had trouble defining a research question and designing the empirical research to answer it. The students proposed impossible fieldwork to answer unanswerable questions. Even many active scholars had difficulty with the basic questions: What do you want to find out? How are you going to find it out? and, above all, How would you know if you were right or wrong?

We found conventional statistical training to be only marginally relevant to those with qualitative data. We even found it inadequate for students with projects amenable to quantitative analysis, since social science statistics texts do not frequently focus on research *design* in observational settings. With a few important exceptions, the scholarly literatures in quantitative political methodology and other social science statistics fields treat existing data and their problems as given. As a result, these literatures largely ignore research design and, instead, focus on making valid inferences through statistical corrections to data problems. This approach has led to some dramatic progress; but it slights the advantage of improving research design to produce better data in the first place, which almost always improves inferences more than the necessarily after-the-fact statistical solutions.

This lack of focus on research design in social science statistics is as surprising as it is disappointing, since some of the most historically important works in the more general field of statistics are devoted to problems of research design (see, e.g., Fisher (1935) *The Design of Experiments*). Experiments in the social sciences are relatively uncommon, but we can still have an enormous effect on the value of our qualitative or quantitative information, even without statistical corrections, by improving the design of our research. We hope our book will help move these fields toward studying innovations in research design.

We culled much useful information from the social science statistics literatures and qualitative methods fields. But for our goal of explicating and unifying the logic of inference, both literatures had problems. Social science statistics focuses too little on research design, and its language seems arcane if not impenetrable. The numerous languages used to describe methods in qualitative research are diverse, inconsistent in jargon and methodological advice, and not always helpful to researchers. We agree with David Collier that aspects of our advice can be rephrased into some of the languages used in the qualitative methods literature or that used by quantitative researchers. We hope our unified logic and, as David Laitin puts it, our "common vocabulary" will help foster communication about these important issues among all social scientists. But we believe that any coherent language could be used to convey the same ideas.

We demonstrated that "the differences between the quantitative and qualitative traditions are only stylistic and are methodologically and substantively unimportant" (p. 4). Indeed, much of the best social science research can combine quantitative and qualitative data, precisely because there is no contradiction between the fundamental processes of inference involved in each. Sidney Tarrow asks whether we agree that "it is the *combination* of quantitative and qualitative" approaches that we desire (p. 473). We do. But to combine both types of data sources productively, researchers need to understand the fundamental logic of inference and the more specific rules and procedures that follow from an explication of this logic.

Social science, both quantitative and qualitative, seeks to develop and evaluate theories. Our concern is less with the development of theory than *theory evaluation*—how to use the hard facts of empirical reality to form scientific opinions about the theories and generalizations that are the hoped for outcome of our efforts. Our social scientist uses theory to generate *observable implications,* then systematically applies publicly known procedures to infer from evidence whether what the theory implied is correct. Some theories emerge from detailed observation, but they should be evaluated with new observations, preferably ones that had not been gathered when the theories were being formulated. Our logic of theory evaluation stresses maximizing leverage—explaining as much as possible with as little as possible. It also stresses minimizing bias. Lastly, though it cannot eliminate uncertainty, it encourages researchers to report estimates of the uncertainty of their conclusions.

Theory and empirical work, from this perspective, cannot productively exist in isolation. We believe that it should become standard practice to demand clear implications of theory and observations checking those implications derived through a method that minimizes bias. We hope that *Designing Social Inquiry* helps to "discipline political science" in this way, as David Laitin recommends; and we hope, along with James Caporaso, that "improvements in measurement accuracy, theoretical specification, and research should yield a smaller range of allowable outcomes consistent with the predictions made" (p. 459).

Our book also contains much specific advice, some of it new and some at least freshly stated. We explain how to distinguish systematic from nonsystematic components of phenomena under study and focus explicitly on trade-offs that may exist between the goals of unbiasedness and efficiency (chap. 2). We discuss causality in relation to counterfactual analysis and what Paul Holland calls the "fundamental problem of causal inference" and consider possible complications introduced by thinking about causal mechanisms and multiple causality (chap. 3). Our discussion of counterfactual reasoning is, we believe, consistent with Donald Campbell's "quasi-experimental" emphasis; and we thank James Caporaso for clarifying this.[1]

We pay special attention in chapter 4 to issues of what to observe: how to avoid confusion about what constitutes a "case" and, especially, how to avoid or limit selection bias. We show that selection on values of explanatory variables does not introduce bias but that selection on values of dependent variables does so; and we offer advice to researchers who cannot avoid selecting on dependent variables.

We go on in chapter 5 to show that while random measurement error in dependent variables does not bias causal inferences (although it does reduce efficiency), measurement error in explanatory variables biases results in predictable ways. We also develop procedures for correcting these biases even when measurement error is unavoidable. In that same chapter, we undertake a sustained analysis of endogeneity (i.e., when a designated "dependent variable" turns out to be causing what you thought was your "explanatory variable") and omitted variable bias, as well as how to control research situations so as to mitigate these problems. In the final chapter, we specify ways to increase the information in qualitative studies that can be used to evaluate theories; we show how this can be accomplished without returning to the field for additional data collection. Throughout the book, we illustrate our propositions not only with hypothetical examples but with reference to some of the best contemporary research in political science.

This statement of our purposes and fundamental arguments should put some of the reviewers' complaints about omissions into context. Our book is about doing empirical research designed to evaluate theories and learn about the world—to make inferences—not about generating theories to evaluate. We believe that researchers who understand how to evaluate a theory will generate better theories—theories that are not only more internally consistent but that also have more observable implications (are more at risk of being wrong) and are more consistent with prior evidence. If, as Laitin suggests, our singlemindedness in driving home this argument led us implicitly to downgrade the importance of such matters as concept formation and theory creation in political science, this was not our intention.

Designing Social Inquiry repeatedly emphasizes the attributes of good theory. How else to avoid omitted variable bias, choose causal effects to estimate, or derive observable implications? We did not offer much advice about what is often called the "irrational nature of discovery," and we leave it to individual researchers to decide what theories they feel are worth evaluating. We do set forth some criteria for choosing theories to evaluate—in terms of their importance to social science and to the real world—but our methodological advice about research design applies to any type of theory. We come neither to praise nor to bury rational-choice theory, nor to make an argument in favor of deductive over inductive theory. All we ask is that whatever theory is chosen be evaluated by the same standards of inference. Ronald Rogowski's favorite physicist, Richard Feynman, explains clearly how to evaluate a theory (which he refers to as a "guess"): "If it disagrees with [the empirical evidence], it is wrong. In that simple statement is the key to science. It does not make any difference how beautiful your guess is. It does not make any difference how smart you are, who made the guess, or what his name is—if it disagrees with [the empirical evidence] it is wrong. That is all there is to it" (1965, 156).[2]

One last point about our goal: we want to set a high standard for research but not an impossible one. All interesting qualitative and quantitative research yields uncertain conclusions. We think that this fact ought not to be dispiriting to researchers but should rather caution us to be aware of this uncertainty, remind us to make the best use of data possible, and energize us to continue the struggle to improve our stock of valid inferences about the political world. We show that uncertain inferences are every bit as scientific as more certain ones so long as they are accompanied by honest statements of the degree of uncertainty accompaning each conclusion.

OUR ALLEGED ERRORS OF OMISSION

The major theme of what may seem to be the most serious criticism offered above is stated forcefully by Ronald Rogowski. He fears that "devout attention" to our criteria would "paralyze, rather than stimulate, scientific inquiry." One of Rogowski's arguments, echoed by Laitin, is that we are too obsessed with increasing the amount of information we can bring to bear on a theory and therefore fail to understand the value of case studies. The other major argument, made by both Rogowski and Collier is that we are too critical of the practice of selecting observations according to values of the dependent variable and that we would thereby denigrate major work that engages in this practice. We consider these arguments in turn.

Science as a Collective Enterprise

Rogowski argues that we would reject several classic case studies in comparative politics. We think he misunderstands these studies and misses our distinction between a "single case" and a collection of observations. Consider two works that he mentions, *The Politics of Accommodation*, by Arend Lijphart (1968), and *The Nazi Seizure of Power*, by William Sheridan Allen (1965). Good research designs are rarely executed by individual scholars isolated from prior researchers. As we say in our book, "A single observation can be useful for evaluating causal explanations if it is part of a research program. If there are other observations, perhaps gathered by other researchers, against which it can be compared, it is no longer a single observation" (p. 211; see also secs. 1.2.1, 4.4.4, the latter devoted entirely to this point). Rogowski may have overlooked these passages. If we did not emphasize the point sufficiently, we are grateful for the opportunity to stress it here.

Lijphart: The Case Study that Broke the Pluralist Camel's Back

What was once called *pluralist theory* by David Truman and others holds that divisions along religious and class lines make polities less able to resolve political arguments via peaceful means through democratic institutions. The specific causal hypothesis is that the existence of many cross-cutting cleavages increases the level of social peace and, thus of stable, legitimate democratic government.

In *The Politics of Accommodation*, Arend Lijphart (1968) sought to estimate this causal effect.[3] In addition to prior literature, he had evidence from only one case, the Netherlands. He first found numerous observable implications of his descriptive hypothesis that the Netherlands had deep class and religious cleavages, relatively few of which were cross-cutting. Then—surprisingly from the perspective of pluralist theory—he found considerable evidence from many levels of analysis that the Netherlands was an especially stable and peaceful democratic nation. These descriptive inferences were valuable contributions to social science and important in and of themselves, but Lijphart also wished to study the broader causal question.

In isolation, a single study of the Netherlands, conducted only at the level of the nation at one point in time, cannot produce a valid estimate of the causal effect of cross-cutting cleavages on the degree of social peace in a nation. But Lijphart was *not* working in isolation. As part of a community of scholars, he had the benefit of Truman and others having collected many prior observations. By using this prior work, Lijphart could and did make a valid inference. Prior researchers had either focused only on countries with the same value of the explanatory variable (many cross-cutting cleavages) or on the basis of values of the dependent variable (high social conflict). Previous researchers therefore made invalid inferences. Lijphart measured social peace for the other value of the explanatory variable (few cross-cutting cleavages) and, by using his data in combination with that which came before, made a valid inference.

Lijphart's classic study is consistent with our model of good research design. As he stressed repeatedly in his book, Lijphart was contributing to a large scholarly literature. As such, he was not trying to estimate a causal effect from a single observation; nor was he selecting on his dependent variable. Harvesting relevant information from others' data, although often overlooked, may often be the best way to obtain relevant information.

By ignoring the place of Lijphart's book in the literature to which it was contributing, Rogowski was unable to recognize the nature of its contribution. Rogowski's alternative explanation for the importance of this book and the others he mentions—that "(1) all of them tested, relied on, or proposed, clear and precise *theories*; and (2) all focused on *anomalies*" (p. 469)—suggests one of many possible strategies for choosing topics to research; but it is of almost no help with practical issues of research design or ascertaining whether a theory is right or wrong. Indeed, the only way to determine whether something is an anomaly in the first place is to follow a clear logic of scientific inference and theory evaluation, such as that provided in *Designing Social Inquiry*.

Allen: Distinguishing History From Social Science

The Nazi Seizure of Power is an account of life in an ordinary German community during the Nazi seizure of power. Allen is not a social scientist: In his book, he proposes no generalization, evaluates no theory, and does not refer to the scholarly literatures on Nazi Germany; rather, he zeroes in on the story of what happened in one small place at a crucial moment in history, and he does so brilliantly. In our terms, he is describing historical detail and occasionally also conducting very limited descriptive inference. We emphasize the importance of such work: "Particular events such as the French Revolution or the Democratic Senate primary in Texas may be of intrinsic interest: they pique our curiosity, and if they were preconditions for subsequent events (such as the Napoleonic Wars or Johnson's presidency) we may need to know about them to understand those later events" (p. 36).

In our view, social science must go further than Allen. The social scientist must make descriptive or causal inferences, thus seeking explanation and generalization. Indeed, we think even Rogowski would not accept Allen's classic work of history as a dissertation in political science. Allen's work is, however, not irrelevant to the task of explanation and generalization that is of interest to us. In the hands of a good social scientist, who could place Allen's work within an intellectual tradition, it becomes a single case study in the framework of many others. This, of course, suggests one traditional and important way in which social scientists can increase the amount of information they can bring to bear on a problem: read the descriptive case study literature.

The Perils of Avoiding Selection Bias

We agree with David Collier's observation that, if our arguments concerning selection bias are sustained, then "a small improvement in methodological self-awareness can yield a large improvement in scholarship" (p. 461). Indeed, because qualitative researchers generally have more control over the selection of their observations than over most other features of their research designs, selection is an especially important concern (a topic to which we devote most of our chap. 4).[4]

Rogowski believes that we would criticize Peter Katzenstein's (1985) *Small States in World Markets* or Robert Bates's (1981) *Markets and States in Tropical Africa* as inadmissibly selecting on the dependent variable. We address each book in turn.

Katzenstein: Distinguishing Descriptive Inference from Causal Inference

Peter Katzenstein's (1985) *Small States in World Markets* makes some important descriptive inferences. For example, Katzenstein shows that small European states responded flexibly and effectively to the economic challenges that they faced during the 40 years after World War II; and he distinguishes between what he calls "liberal and social corporatism" as two patterns of response. But many of Katzenstein's arguments also imply causal claims—that in Western Europe "small size has facilitated economic openness and democratic corporatism" (p. 80), and that in the small European states, weak landed aristocracies, relatively strong urban sectors, and strong links between country and city led to cross-class compromise in the 1930s, creating the basis for postwar corporatism (chap. 4).

Katzenstein seeks to test the first of these causal claims by comparing economic openness in small and large states (1985, table 1, p. 86). To evaluate the second hypothesis, he compares cross-class compromise in six small European states characterized by weak landed aristocracies and strong urban sectors, with the relative absence of such compromise in five large industrialized countries and Austria, which had different values on these explanatory variables. Much of his analysis follows the rules of scientific inference we discuss—selecting cases to vary the value of the explanatory variables, specifying the observable implications of theories, seeking to determine whether the facts meet theoretical expectations.

But Katzenstein fudges the issue of causal inference by disavowing claims to causal validity: "Analyses like this one cannot meet the exacting standards of a social science test that asks for a distinction between necessary and sufficient conditions, a weighting of the relative importance of variables, and, if possible, a proof of causality" (p. 138). However, estimating causal inferences does not require a "distinction between necessary and sufficient conditions, a weighting of the relative importance of variables," or an absolute "proof" of anything. Katzenstein thus unnecessarily avoids causal language and explicit attention to the logic of inference which results. As we explain in our book, "Avoiding causal language when causality is the real subject of investigation either renders the research irrelevant or permits it to remain undisciplined by the rules of scientific inference" (p. 76).

Remaining inexplicit about causal inference makes some of Katzenstein's claims ambiguous or unsupported. For example, his conclusion seems to argue that small states' corporatist strategies are responsible for their postwar economic success. But because of the selection bias induced by his decision to study only successful cases, Katzenstein cannot rule out an important alternative causal hypothesis—that any of a variety of other factors accounts for this uniform pattern. For instance, the postwar international political economy may have been benign for small, developed countries in Europe. If so, corporatist strategies may have been unrelated to the degree of success experienced by small European states.

In the absence of variation in the strategies of his states, valid causal inferences about their effects remain elusive. Had Katzenstein been more attentive to the problems of causal inference that we discuss, he would have been able to claim causal validity in some limited instances, such as when he had variation in his explanatory and dependent variables (as in the 1930s analysis). More importantly, he would also have been able to improve his research design so that valid causal inferences were also possible in many other areas.

Rogowski is not correct in inferring that we would dismiss the significance of *Small States in World Markets*. Its descriptions are rich and fascinating, it elaborates insightful concepts such as liberal and social corporatism, and it provides some evidence for a few causal inferences. It is a fine book, but we believe that more explicit attention to the logic of inference could have made it even better.

Bates: How to Identify a Dependent Variable

Rogowski claims that Robert Bates's purpose in *Markets and States* was to explain economic failure in tropical African states and that by choosing only states with failed economies and low agricultural production, Bates biased his inferences. If agricultural production were Bates's dependent variable, Rogowski would be correct, since (as we describe in *Designing Social Inquiry* and as elaborated by Collier), using—but not correcting for—this type of case selection does bias inferences. However, low agricultural production was, in fact, not Bates's dependent variable.

Bates's book makes plain his two dependent variables: (1) the variations in *public policies* promulgated by African states and (2) differences in the *group relations* between the farmer and the state in each country. Both variables vary considerably across his cases. Bates also proposed several explanatory variables, which he derived from his preliminary descriptive inferences. These include (1) whether state marketing boards were founded by the producers or by alliances between government and trading interests, (2) whether urban or rural interests dominated the first postcolonial government, (3) the degree of governmental committment to spending programs, (4) the availability of nonagricultural sources for governmental funds, and (5) whether the crops produced were for food or export. These explanatory variables do vary, and they helped account for the variations in public policy and state-farmer relations that Bates observed.

As such, Bates did not select his observations so they had a constant value for his dependent variable. Moreover, he did not stop at the national level of analysis, for which he had a small number of cases and relatively little information. Instead, he offered numerous observable implications of the effects of these explanatory variables at other levels of analyses within each country. As with many qualitative studies, Bates had a small number of cases but an immense amount of information. We believe one of the reasons Bates's study is—and should be—so highly regarded is that it is an excellent example of a qualitative study that conforms to the rules of scientific inference. In sum, Rogowski says that Bates had an excellent book that we would reject. If the book were as Rogowski describes it, we very well might reject it. Since it is not—and indeed is a good example of our logic of research design—we join Rogowski in applauding it.[5]

TRIANGULAR CONCLUSIONS

We conclude by emphasizing a point that is emphasized both in *Designing Social Inquiry* and in the reviews. We often suggest procedures that qualitative researchers can use to increase the amount of information they bring to bear on evaluating a theory. This is sometimes referred to as "increasing the number of observations." As all our reviewers recognize, we do not expect researchers to increase the number of full-blown case studies to conduct a large-*n* statistical analysis: our point is not to make quantitative researchers out of qualitative researchers. In fact, most qualitative studies already contain a vast amount of information. Our point is that appropriately marshaling all the thick description and rich contextualization in a typical qualitative study to evaluate a specific theory or hypothesis can produce a very powerful research design. Our book demonstrates how to design research in order to collect the most useful qualitative data and how to restructure it even after data collection is finished, to turn qualitative information into ways of evaluating a specific theory. We explain how researchers can do this by collecting more observations on their dependent variable, by observing the same variable in another context, or by observing another dependent variable that is an implication of the same theory. We also show how one can design theories to produce more observable implications that then put the theory at risk of being wrong more often and easily.

This brings us to Sidney Tarrow's suggestions for using the comparative advantages of both qualitative and quantitative researchers. Tarrow is interested specifically in how unsystematic and systematic variables and patterns interact, and seems to think that principles could be derived to determine what unsystematic events to examine. We think that this is an interesting question for any historically-sensitive work. Many unsystematic, nonrepeated events occur, a few of which may alter the path of history in significant ways; and it would be useful to have criteria to determine how these events interact with systematic patterns. We expect that our discussions of scientific inference could help in identifying which apparently random, but critical, events to study in specific instances, and we are confident that our logic of inference will help determine whether these inferences are correct; but Tarrow or others may be able to use the insights from qualitative researchers to specify them more clearly. We would look forward to a book or article that presented such criteria.

Another major point made by Tarrow is that all appropriate methods to study a question should be employed. We agree: a major theme of our book is that there is a single unified logic of inference. Hence it is possible effectively to combine different methods. However, the issue of triangulation that Tarrow so effectively raises is not the use of different logics or methods, as he argues, but the triangulation of diverse *data sources* trained on the same problem. Triangulation involves data collected at different places, sources, times, levels of analysis, or perspectives, data that might be quantitative, or might involve intensive interviews or thick historical description. The best method should be chosen for each data source. But more data are better. Triangulation, then, is another word for refering to the practice of increasing the amount of information to bear on a theory or hypothesis, and that is what our book is about.

Notes

The table of contents, preface, and chapter 1 of *Designing Social Inquiry* are available via Gopher from hdc-gopher.harvard.edu.

1. To clarify further, we note that the definition of an "experiment" is investigator control over the assignment of values of explanatory variables to subjects. Caporaso emphasized also the value of random assignment, which is desirable in some situations (but not in others, see pp. 124–8) and sometimes achievable in experiments. (Random selection and a large number of units are also desirable and also necessary for relatively automatic unbiased inferences, but experimenters are rarely able to accomplish either.) A "quasi-experiment" is an observational study with an exogenous explanatory variable that the investigator does not control. Thus, it is not an experiment. Campbell's choice of the word "quasi-experiment" reflected his insight that observational studies follow the same logic of inference as experiments. Thus, we obviously agree with Campbell's and Caporaso's emphases and ideas and only pointed out that the word "quasi-experiment" adds another word to our lexicon with no *additional* content. Its a fine idea, much of which we have adopted; but it is an unnecessary category.

2. Telling researchers to "choose better theories" is not much different than telling them to choose the right answer: it is correct but not helpful. Many believe that deriving rules for theory creation is impossible (e.g., Popper, Feynman), but we see no compelling justification for this absolutist claim. As David Laitin correctly emphasizes, "the development of formal criteria for such an endeavor is consistent with the authors' goals."

3. Lijphart also went to great lengths to clarify the precise theory he was investigating, because it was widely recognized that the concept of pluralism was often used in conflicting ways, none clear or concrete enough to be called a theory. Ronald Rogowski's description of pluralism as a "powerful, deductive, internally consistent theory" (p. 10) is surely the first time it has received such accolades.

4. Selection problems are easily misunderstood. For example, Caporaso claims that "if selection biases operate independently of one's hypothesized causal variable, it is a threat to internal validity; if these same selection factors interact with the causal variable, it is a threat to external validity" (p. 9). To see that this claim is false, note, as Collier reemphasizes, that Caporaso's "selection factors" can also be seen as an omitted variable. But omitted variables cannot cause bias if they are independent of your key causal variable. Thus, although the distinction between internal and external validity is often useful, it is not relevant to selection bias in the way Caporaso describes.

5. Subsequently, Bates pursued the same research program. For example, in *Essays on the Political Economy of Rural Africa* he evaluated his thesis for two additional areas—colonial Ghana and Kenya (1983, chap. 3). So Bates does exactly what we recommend: having developed his theory in one domain, he extracts its observable implications and moves to other domains to see whether he observes what the theory would lead him to expect.

Symposium References

Achen, Christopher H. 1982. *Interpreting and Using Regression Analysis*. University Paper series on Quantitative Applications in the Social Sciences, no. 29. Beverly Hills: Sage.

Achen, Christopher H. 1986. *The Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.

Achen, Christopher H., and Duncan Snidal. 1989. "Rational Deterrence Theory and Comparative Case Studies." *World Politics* 41:143–69.

Allen, William Sheridan. 1965. *The Nazi Seizure of Power: The Experience of a Single German Town, 1930–1935*. New York: Watts.

Almond, Gabriel. 1990. *A Divided Discipline*. Newbury Park, CA: Sage.

Arendt, Hannah. 1958. *The Origins of Totalitarianism*. Cleveland, OH: World.

Ayer, A. J. 1946. *Language, Truth, and Logic*. 2d ed. London: Gollancz.

Bakhtin, M. M. 1986. *Speech Genres and Other Late Essays*. Trans. Vern M. McGee. Austin: University of Texas Press.

Bates, Robert H. 1981. *Markets and States in Tropical Africa: The Political Basis of Agrarian Policies*. Berkeley: University of California Press.

Bates, Robert H. 1983. *Essays on the Political Economy of Rural Africa*. New York: Cambridge University Press.

Bendix, Reinhard. 1963. "Concepts and Generalizations in Comparative Sociological Studies." *American Sociological Review* 28:532–39.

Berkson, Joseph. 1946. "Limitations of the Application of Fourfold Table Analysis to Hospital Data." *Biometrics Bulletin* 2:47–53.

Blalock, Hubert M. 1964. *Causal Inferences in Non-experimental Research*. Chapel Hill: University of North Carolina Press.

Blalock, Hubert M. 1984. *Basic Dilemmas in the Social Sciences*. Beverly Hills, CA: Sage.

Bunce, Valerie. 1981. *Do New Leaders Make a Difference? Exclusive Succession and Public Policy under Capitalism and Socialism*. Princeton: Princeton University Press.

Bourdieu, Peirre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Trans. Richard Nice. Cambridge: Harvard University Press.

Campbell, Donald T. 1969. "Reforms as Experiments." *American Psychologist* 24:409–29.

Campbell, Donald T., and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin.

Collier, David. 1993. "The Comparative Method." In *Political Science: The State of the Discipline II*, ed. Ada W. Finifter. Washington: American Political Science Association.

Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimentation*. Chicago: Rand-McNally.

Durkheim, Emile. 1938. *The Rules of Sociological Method*. Trans. Sarah A. Solovay and John H. Mueller. New York: Free Press.

Eckstein, Harry. 1975. "Case Study and Theory in Political Science." In *Handbook of Political Science*, vol. 7, *Strategies of Inquiry*, ed. Fred I. Greenstein and Nelson Polsby. Reading, MA: Addison-Wesley.

Fearon, James D. 1990. "Counterfactuals and Hypothesis Testing in Political Science." *World Politics* 43:169–95.

Feynman, Richard Phillips. 1965. *The Character of Physical Law*. Cambridge: MIT Press.

Fisher, Sir Ronald Aylmer. 1935. *The Design of Experiments*. Edinburgh: Oliver & Boyd.

Foucault, Michel. 1972. *The Archaeology of Knowledge*. Trans. A. M. Sheridan Smith. New York: Pantheon Books.

Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." In *Political Analysis*, vol. 2, ed. James A. Stimson. Ann Arbor: University of Michigan Press.

Geertz, Clifford. 1973. "Thick Description: Toward an Interpretative Theory of Culture." In his *Interpretation of Cultures*. New York: Basic Books.

George, Alexander, and Timothy J. McKeown. 1985. "Case Studies and Theories of Organizational Decision Making." *Advances in Information Processes in Organizations* 2:21–58.

Gourevitch, Peter Alexis. 1978. "The International System and Regime Formation: A Critical Review of Anderson and Wallerstein." *Comparative Politics* 10:419–38.

Griffin, Larry J. 1992. "Temporality, Events, and Explanation in Historical Sociology: An Introduction." *Sociological Methods and Research* 20:403–27.

Heberle, Rudolf. 1970. *From Democracy to Nazism: A Regional Case Study on Political Parties in Germany*. New York: Grosset & Dunlap.

Heberle, Rudolf. 1963. *Landbevölkerung und Nationalsozialismus: Eine soziologische Untersuchung der politischen Willensbildung in Schleswig–Holstein 1918 bis 1932*. Rev. ed. Stuttgart: Deutsche Verlags–Anstalt.

Hempel, Carl, G. 1966. *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice–Hall.

Jervis, Robert. 1989. "Rational Deterrence: Theory and Evidence." *World Politics* 41:183–207.

Katzenstein, Peter J. 1985. *Small States in World Markets: Industrial Policy in Europe*. Ithaca, NY: Cornell University Press.

Kendall, Maurice G., and William R. Buckland. 1960. *A Dictionary of Statistical Terms*. 2d ed. New York: Hafner.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Cambridge: Cambridge University Press.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Kohli, Atul. 1987. *The State and Poverty in India*. New York: Cambridge University Press.

Kornhauser, William. 1959. *The Politics of Mass Society*. New York: Free Press of Glencoe.

Kriesi, Jan. 1994. *New Social Movements in Western Europe*. Minneapolis: University of Minnesota Press.

Kubik, Jan. 1994. *The Power of Symbols against the Symbols of Power: The Rise of Solidarity and the Fall of State Socialism in Poland*. University Park: Pennsylvania State University.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Laba, Roman. 1991. *The Roots of Solidarity: A Political Sociology of Poland Working-class Democratization*. Princeton: Princeton University Press.

Laitin, David. 1994. "The Tower of Babel as a Coordination Game." *American Political Science Review* 88:622–34.

Lave, Charles, and James March. 1975. *An Introduction to Models in the Social Sciences*. New York: Harper & Row.

Lederer, Emil. 1940. *State of the Masses*. New York: Norton.

Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65:682–93.

Lijphart, Arend. 1975. *The Politics of Accommodation: Pluralism and Democracy in the Netherlands*. Berkeley: University of California Press.

Lipton, Michael. 1976. *Why Poor People Stay Poor: Urban Bias in World Development*. Cambridge: Harvard University Press.

McAdam, Doug. 1981. *Freedom Summer*. New York: Oxford University Press.

Martin, Lisa L. 1992. *Coercive Cooperation: Explaining Multilateral Economic Sanctions*. Princeton: Princeton University Press.

Martin, Lisa L., and Kathryn Sikkink. 1993. "U.S. Policy and Human Rights in Argentina and Guatemala, 1973–1980." In *Double-edged Diplomacy* ed. Peter B. Evans, Harold K. Jacobson, and Robert D. Putnam. Berkeley: University of California Press.

Meehl, Paul E. 1967. "Theory-Testing in Psychology and Physics: A Methodological Paradox." *Philosophy of Science*, June, pp. 103–15.

Mill, John Stuart. 1974. "Of the Four Methods of Experimental Inquiry." In his *A System of Logic*. Toronto: University of Toronto Press.

Moore, Barrington, Jr. 1967. *Social Origins of Dictatorship and Democracy*. Boston: Beacon.

Moses, Lincoln E. 1968. "Truncation and Censorship." In *International Encyclopedia of the Social Sciences*, vol. 15; ed. David L. Sills. New York: Macmillan.

Perrot, Michelle. 1986. "On the Formation of the French Working Class." In *Working Class Formation*, ed. Ira Katznelson and Aristide Zolberg. Princeton: Princeton University Press.

Porter, Michael E. 1990. *The Competitive Advantage of Nations*. New York: Free Press.

Przeworski, Adam, and Fernando Limongi. 1992. "Selection, Counterfactuals, and Comparisons." University of Chicago. Typescript.

Przeworski, Adam, and Fernando Limongi. 1993. "Political Regimes and Economic Growth." *Journal of Economic Perspectives* 7:51–69.

Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley.

Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.

Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.

Russell, Bertrand. 1969. *The Autobiography of Bertrand Russell, 1914–1944*. New York: Bantam Books.

Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *American Political Science Review* 64:1033–53.

Sartori, Giovanni. 1976. *Parties and Party Systems: A Framework for Analysis*. Cambridge: Cambridge University Press.

Scuibba, Roberto, and Rossana Schiubba Pace. 1976. *Le comunita di base in Italia*. 2 vols. Rome: Coines.

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. New York: Cambridge University Press.

Skocpol, Theda, and Margaret Somers. 1980. "The Uses of Comparative History in Macrosocial Inquiry." *Comparative Studies in Society and History* 22:174–97.

Smelser, Neil J. 1976. *Comparative Methods in the Social Sciences*. Englewood Cliffs, NJ: Prentice-Hall.

Stinchcombe, Arthur L. 1968. *Constructing Social Theories*. New York: Harcourt, Brace & World.

Stolzenberg, Ross M., and Daniel A. Relles. 1990. "Theory Testing in a World of Constrained Research Design: The Significance of Heckman's Censored Sampling Bias Correction for Nonexperimental Research." *Sociological Methods and Research* 18:395–415.

Tarrow, Sidney. 1988a. "Old Movements in New Cycles of Protest: The Career of an Italian Religious Community." In *From Structure to Action*, ed. B. Klandermans, et al. International Social Movement Research Series, no 1. Greenwich, CT: JAI

Tarrow, Sidney. 1988b. *Democracy and Disorder: Protest and Politics in Italy, 1965–1975*. New York: Oxford University Press.

Tarrow, Sidney. 1994. *Power in Movement: Collective Action, Social Movements, and Politics*. New York: Cambridge University Press.

Tilly, Charles. 1990. *Coercion, Capital, and European States, 990–1990 A.D.* Cambridge, MA: Blackwell.

Tilly, Charles. 1993. *European Revolutions, 1492–1992*. Oxford: Blackwell.

Tilly, Charles. 1994. "State and Nationalism in Europe, 1492–1992." *Theory and Society* 23:131–46.

Truman, David Bicknell. 1951. *The Governmental Process: Political Interest and Public Opinion*. New York: Knopf.

Walker, Henry A., and Bernard P. Cohen. 1985. "Scope Statements: Imperatives for Evaluating Theory." *American Sociological Review* 50:288–301.

Wallerstein, Immanuel. 1974. *The Modern World System*. Vol. 1. New York: Academic.

Weber, Marianne. 1988. *Max Weber: A Biography*. Trans. Harry Zohn. New Brunswick: Transaction.

THE QUALITATIVE-QUANTITATIVE DISPUTATION: GARY KING, ROBERT O. KEOHANE, AND SIDNEY VERBA'S *DESIGNING SOCIAL INQUIRY: SCIENTIFIC INFERENCE IN QUALITATIVE RESEARCH*

Designing Social Inquiry: Scientific Inference in Qualitative Research. By Gary King, Robert O. Keohane, and Sidney Verba. Princeton: Princeton University Press, 1994. 238p. $55.00 cloth, $19.95 paper

Gary King, Robert O. Keohane, and Sidney Verba (KKV) have provoked much impassioned debate at conventions and over the information superhighway with a simple but controversial argument: the logic of good qualitative and good quantitative research is essentially the same. Their book shows how to design qualitative (small-n) studies so that they satisfy the canons of scientific inference. We asked five senior scholars, each of whose work mixes qualitative and quantitative data and methods, to evaluate the success of KKV's attempt to unify political science. David Laitin is a skeptic who wonders whether anyone can "discipline" us unruly political scientists. James Caporaso offers cautious reminders about the many varieties of qualitative research and the many meanings of falsification. David Collier examines KKV's treatment of selection bias, arguing that many of their recommendations correspond to conventional understandings that are already well-established in the field of comparative method, and that qualitative researchers sometimes have a different perspective on basic trade-offs involved in research design. Ronald Rogowski throws down the gauntlet: political scientists who have a strong theory may properly ignore some of KKV's pet "canons." Finally, Sidney Tarrow suggests that triangulating qualitative and quantitative approaches involves much more than considerations of research design. In a rejoinder, KKV reaffirm their belief that political scientists who slight design considerations ultimately hurt their own work. They conclude with a message to the discipline: good design—assuming there is good theory—produces good qualitative and quantitative political science.

DISCIPLINING POLITICAL SCIENCE

DAVID D. LAITIN *University of Chicago*

If political science is ready to be disciplined, King, Keohane and Verba's *Designing Social Inquiry* (KKV) can do that disciplining. By this I mean that the book contains a set of concepts, rules of inference, and methodological precepts that apply to all researchers who seek a generalized and systematic understanding of politics. This does not mean that we all should be doing the same sort of research. Indeed, the rules elucidated in this book have relevance to statistically minded scholars, formal modelers, comparativists, thick describers, and interpretivists. What it does mean is that we all must remain conscious about the degree to which our own research answers an important question, so that we can accurately signal to fellow members of our discipline how much of the picture we have filled in. If we all share a common vocabulary and common standards for evaluation of evidence in light of a theory, we can become a community of scholars in common pursuit of valid knowledge. More bluntly, if we could agree upon standards of scientific inference, we could better identify our colleagues who are guilty of scientific malpractice—which, if regularly done, is a good operational indicator of a discipline. We need not, as Almond (1990) has suggested, eat at "separate tables" any longer; it is now possible productively to consume across cuisines.

Designing Social Inquiry is not itself a methodological breakthrough. Very little in it will be new or surprising to moderately well trained students in political science. What is truly innovative about this book is its catholicity. Its goal is not to exclude the "soft" side of political science from a discipline controlled by "hard-line" statisticians. Rather, its central thesis is that at root, quantitative and qualitative research in political science share a "unified logic" (p. 3). With that viewpoint, KKV's critiques of the methodological problems faced in actual qualitative research show a generosity of spirit. The book has high praise for qualitative work containing elements of good scientific practice. It also has feasible suggestions that would have improved other work that failed to meet reasonable scientific standards. Indeed, the primary goal of the book is to demonstrate to those of us on the soft side that we can approximate the standards of our brethren on the hard side if we make such an attempt. It must be noted that those veterans in quantitative work who on principle ignore "soft" research as unscientific will be disabused from their narrow viewpoint. The achievement of this book, then, is that it sets a reasonable disciplinary standard without using the young A. J. Ayer's (1946) tactic of calling all work on the soft side "metaphysics."

But disciplining has its down side, as Foucault insists in his analysis of those "discursive formations" that transcend or deconstruct disciplines (1972, 178–81). Many political scientists eagerly entered our field, perhaps from unfortunate or boring experiences in the disciplines of economics or psychology, precisely because we have been eclectic, undisciplined, and willing to tolerate a multitude of discursive formations. Brilliant young students who wish to travel to exotic places, read the classics, work out a personal utopia, or promote a political cause enter political science programs. These students may find the disciplining constraints imposed by rules of inference to be an unnecessary burden. Sensitive colleagues are willing to indulge these students, in large part because they themselves were to some extent attracted to political science because its lack of discipline was so inviting.

Students of Bakhtin would make a complementary critique. Bakhtin argued that linguistic assimilation, which is part and parcel of disciplining, leads to the emergence of "canonical" or "authoritative utterances," which themselves are capable of undermining dissent (1986, 88). A common vocabulary, from a Bakhtinian viewpoint, is never neutral. Accepting KKV's "statistical" vocabulary as bedrock could consequently lock us into a cultural framework. Indeed, their call to engage in disciplinary discourse in a language most qualitativists see as "foreign" is surely the source of anger felt by many practitioners who have read this book.

Although I am personally sympathetic with the Foucauldians and Bakhtinian's amongst us, it must be remembered that sharing a language promotes not only effective communication but also focused debate across subdisciplines. While the language of statistics does have its biases, KKV provides a conceptual apparatus that has referents in virtually all domains of our discipline. Scholars writing an article for disciplinary journals in the narrative mode, for example, will be able to use KKV's apparatus to justify its scientific merit in political science's division of labor and thereby raise its chances of appearing in these august and prestigious pages, with concomitant disciplinary rewards. Even more: scholars who strongly disagree with the statistician's bias will, after reading this book, have the tools to show its limitations to practitioners in all of political science. This book is surely the icon that iconoclasts should lust after.

It could still be objected—(along the lines I argued in Laitin 1994)—that maybe it will bring higher expected utility if statisticians learned the language of nonquantitative researchers, rather than the other way around. To this I reply that as of now, there is no contending universal vocabulary for ascertaining whether our research findings are valid. However, I would welcome a counterhegemonic project along the lines of the present one, with an alternative critical language of scientific evaluation that would be applicable in all domains of our discipline. But my welcome of alternatives in no way diminishes my admiration of the three authors of this volume for having centrally positioned their own hegemonic design.

Causes and Concepts

KKV's hegemonic project is to highlight the making of valid causal inferences as the highest goal for social inquiry. To make such inferences, researchers need to combine theory with observations in such a way as to demonstrate a causal effect. With a disciplinary division of labor, the search for valid causal inferences invites participation of scholars on both sides of our present disciplinary divide. On the one hand, the discipline is open to pure describers. Historical and anthropological interpretation are potentially fundamental for us, just so long as researchers in this mode seek to distinguish what is systematic—and what, random—in the particular events they are interpreting. Assessments of this nature will help other scholars use those studies to construct more general theory. On the other hand, the discipline must include formal modelers, if only to demonstrate through the use of mathematics the internal inconsistencies in proposed theories. But within political science, these modelers must subject their stories to systematic and unbiased tests and alter assumptions or set parameter conditions for their models when data do not confirm their theories. Historians need not make general theory; modelers need not collect systematic data; but if both are members of a common discipline, they will do their work in such a way that scholars on the other side of the divide will be able to make reasonable and productive use of their work to ensure that science advances.

The summum bonum of political science, despite KKV's admirable formulation, has never been valid causal inferences. The founders of modern social theory indeed thought otherwise. Max Weber has suggested that the essence of social theory is in the "creation of clear concepts" (Weber 1988, 278). And Emile Durkheim (1938) was especially concerned with the identification of "social facts" Indeed, concepts such as "charisma" and "the division of labor" have been longer-lasting than any valid claims about the causal effects of these concepts. It is hard to think about the political world without them, even if their causal role in any political process remains obscure. And many other such concepts guide our thinking and theorizing today, such as *cross-cutting cleavages, social mobility, prisoner's dilemma, exit/voice/loyalty, social mobilization, political culture, median voter,* and *hegemony*. Such concepts are theoretical in the sense that they combine discrete facts common to our daily life into a category, helping us to see the confusing universe in which we live in a more patterned way.

Although *Designing Social Inquiry* is at its weakest in analyzing the role of concept formation in political science, there is every reason to maintain that the development of formal criteria for such an endeavor is consistent with the authors' goals. Suppose, for example, *prisoner's dilemma* captures elements of reality in everyday life, in international relations, and in congressional committees; but we can make no useful inferences for what people will do—cooperate or defect—if they find themselves in a prisoner's dilemma. Does this mean that the concept has failed and should not be included in our theories of conflict and cooperation? Probably not. Implied by the framework provided in KKV, what we should require of researchers is that they set clear criteria to identify a prisoner's dilemma and continue to search for regularities in outcome depending on the context in which the game is played. Compelling concepts need not be part of a valid causal inference to be powerful; but to remain powerful, these concepts must be part of a research agenda that seeks to identify their systematic implications, revealing their link on a causal chain. KKV may have undervalued the crucial role of conceptual formulation in social inquiry; but this by no means is an argument to reject the disciplining that their work demands.

Critiques from within the Discipline

At a symposium devoted to *Designing Social Inquiry* held at the 1994 annual conference of the American Political Science Association, leading scholars in our discipline were far less enthusiastic than I in regard to the success of this book. Larry Bartels pointed out that the authors treated many statistical conventions (which, in reality, cover over unresolved issues) as solutions to complex epistemological problems. Reliance on these conventions, Bartells inferred, is hardly a solution to the related problems that qualitative researchers have long been addressing with their own conventions. Peter Lange argued that researchers in the area-studies tradition do not seek generality of explanation, because they hold that the "context" in which politics get played out is highly determinative of outcomes yet itself not subject to variable analysis. And Ronald Rogowski argued that some of the best work in the comparative field ignored KKV's injunctions (e.g., on never choosing cases based on codings on the dependent variable); and yet those works' high scientific status can still be justified.

The criticisms made at the APSA convention have merit. I believe two of the critiques are so fundamental as to require future revision of the text. First, KKV focus too much attention on selection criteria within a single study and undervalue the scientific practice of strategically choosing observations based upon knowledge of cases from parallel studies. If the community of scientists, rather than the individual researcher, is the unit of evaluation, some of the selection problems that King, Keohane, and Verba identify in particular studies would be partially washed away. Second, in undervaluing theory, they do not address the issue that selection criteria may be different when theory is strong as opposed to when theory is weak.

The judgment of APSA panelists was harsh indeed. But it ought to be remembered that the criticisms came from scholars who share an understanding of the bedrock concepts of our discipline that are elaborated fully in KKV. This made the criticisms powerful and interesting and allowed for focused debate. Their critiques confirmed, rather than undermined, the importance of this material for the construction of a scientific discipline.

A Plea for Utopia

This review has become something of a plea, or to use Henry Brady's label at the APSA symposium in regard to KKV, a "homily". I would hope that all of our political science curricula include the material developed in *Designing Social Inquiry*. Assigning the book in a required "logic of research" course is only one route to this goal. An alternative is to present the material in lectures, while assigning important articles and books to the students, with the goal of scrutinizing these studies to see how their authors dealt with fundamental issues of descriptive or causal inference. However presented, the concepts and precepts outlined in this book ought to become part of what Bourdięu (1984) would call our intellectual *habitus*. Mutual acknowledgement of work transcending the quantitative and qualitative divide should ensue. This can only spawn—and need not stifle—creativity.

And there are additional rewards for living in such a *habitus*. Suppose it became common practice at job talks, reviews for journals, and panels at disciplinary meetings to ask authors how they addressed issues of endogeneity, of multicollinearity, of possible missing variable bias, of alternative observable implications of their theory, or of their judgment concerning the number of observations necessary for valid causal inference. Such a disciplinary practice would impel all researchers to think systematically about these issues in the course of their research. They need not follow all the rules in this book. KKV recognize that in most real-world research environments, this would be impossible. But all researchers must have good scientific reasons for disregarding or modifying a particular rule. And these reasons must be made available to potential critics. The goal of making political science a discipline seems utopian, but KKV show that it is within our reach.

There is little reason, however, to be sanguine. The reaction to this book at the APSA convention gives me the impression that there is little interest in—and great opposition to—our becoming a discipline. This book will stand, then, as merely a useful exposition of statistical solutions to epistemological questions for those of us who are not statisticians. A pity that a book with such potential will play such a limited role!

THE ROLE OF THEORY AND ANOMALY IN SOCIAL-SCIENTIFIC INFERENCE

RONALD ROGOWSKI *University of California, Los Angeles*

Designing Social Inquiry, by King, Keohane, and Verba (KKV), deserves praise for many reasons. It attempts, seriously and without condescension, to bridge the gap between qualitative and quantitative political science. It reminds a new generation of students, in both traditions, of some main characteristics of good theory (testability, operationalizability, and "leverage," or *deductive fertility*). It clarifies, even for the profoundly mathematically challenged, some of the central strictures of statistical inference (why one cannot have more variables than cases or select on the dependent variable, or why it biases results if measurement of the independent variable is faulty). It abounds with practical wisdom on research design, case selection, and complementary methodologies. Perhaps most important, it opens a dialogue between previously isolated practitioners of these two forms of analysis and provokes worthwhile discussion.

For all of these reasons and more, the book should be, will be, and—indeed, even in its *samizdat* forms—already has been widely assigned and read. It is, quite simply, the best work of its kind now available; indeed, it is very likely the best yet to have appeared.[1] At the same time, I think, *Designing Social Inquiry* falters in its aim of evangelizing qualitative social scientists; and it does so, paradoxically, because it attends insufficiently to the importance of problemation and deductive theorizing in the scientific enterprise.

As natural scientists have long understood (see Hempel, 1966), inference proceeds most efficiently by three complementary routes: (1) making clear the essential model, or process, that one hypothesizes to be at work; (2) teasing out the deductive implications of that model, focusing particularly on the implications that seem *a priori* least plausible; and (3) rigorously testing those least plausible implications against empirical reality.[2] The Nobel physicist and polymath Richard Feynman may have put it best:[3]

> Experimenters search most diligently, and with the greatest effort, in exactly those places where it seems most likely that we can prove our theories wrong. In other words we are trying to prove ourselves wrong as quickly as possible, because only in that way can we find progress. (1965, 158)

The classical example is Einstein's Theory of Relativity, which (1) uniquely provided an overarching model that could explain both the anomalies and the enduring validities of classical Newtonian mechanics, indeed could subsume it as a special case; (2) had, among its many other implications, a quite specific, rather implausible, and previously untested one about how light reflected from the planet Mercury would be deflected by the sun's gravitation; and (3) appeared at the time to be precisely accurate in this specific and implausible implication.[4] To test, however rigorously, hypotheses that challenge no deeper theory or that themselves lack deductive implications is an inefficient route of scientific inference; while theories that are precise and deductively fertile enough can often be sustained or refuted by surprisingly unelaborate tests, including ones that involve few observations or that violate normally sacrosanct principles of selection.

KKV, I contend, emphasize the third part of scientific inquiry, the rigorous testing of hypotheses, almost to the exclusion of the first two—the elaboration of precise models and the deduction of their (ideally, many) logical implications—and thus point us to a pure, but needlessly inefficient, path of social-scientific inquiry.

I can best illustrate these points by applying their strictures to some landmark works in comparative politics, often cited as worthy of emulation. Each work, as it seems to me, would fail KKV's tests and would be dismissed as insufficiently scientific. Yet in each case, I contend, the dismissal would be incorrect: the works illustrate—indeed, epitomize—valid and efficient social-scientific inquiry; and the ways in which they do so illuminate the shortcomings in the analysis under review.

Three of the classical works that I have in mind are single-observation studies; one involves three cases but all within a single region; one selects chiefly on the independent—but also on the dependent—variable, in ways deprecated by KKV; and one selects on the dependent variable. I propose (1) to sketch each briefly; (2) to argue that the conventional wisdom is right and KKV are wrong with regard to these works' worth; and (3) to reflect on the deficiencies that these works reveal in the KKV analysis.

The single-observation studies are Arend Lijphart's 1968 study of the Netherlands, *The Politics of Accommodation*; William Sheridan Allen's single-city examination, *The Nazi Seizure of Power*; and Peter Alexis Gourevitch's critique of Immanuel Wallerstein's *Modern World-System*. Each involves disconfirmation of a prevailing theory, by what Eckstein called the strategy of the "most likely" case (1975, 119).

Lijphart rightly saw in the Netherlands a serious empirical challenge to David Truman's (1951) then widely accepted theory of "cross-cutting cleavages." Truman had argued, plausibly enough, that mutually reinforcing social cleavages (class coterminous with religious practice, or religion with language) impeded social agreement and made conflict more likely. Only where each deep cleavage was orthogonal to another (e.g., Switzerland, where many Catholics are German-speaking, many Francophones Protestant) was social peace likely to endure. About the Netherlands, however, two things were abundantly clear: (1) it had virtually no cross-cutting cleavages; and (2) it had about as stable and amicable a democracy as one could find. Lijphart's study was taken at the time (I believe correctly) as having refuted Truman's theory.[5]

In attempting to explain popular support for such totalitarian movements as fascism, many social scientists had, by the 1950s, accepted a theory whose roots went back to Montesquieu and Tocqueville but whose modern version had been shaped chiefly by Lederer (1940), Arendt (1958), and—the great synthesizer of this genre—Kornhauser (1959). Again simplifying it to the point of caricature, this theory held that societies were opened to totalitarianism's Manichean zealotries by the waning (e.g., through rapid modernization) of associational life—the disappearance of those "natural" groups that afforded meaning, balance, and a sense of efficacy. Totalitarian followers were "atomized" or "mass" individuals.

Tracing the growth of the national socialist cause in a single mid-sized German town where it had prospered earlier and better than average, however, Allen (1965) found, if anything, a superabundance of associational life: singing and shooting societies, card clubs, fraternal orders, religious associations, drinking groups, and *Stammtische* of long standing, to the point that one could hardly imagine a free evening in these protofascists' lives. Neither could he observe any waning of this associational activity before or during the Nazi expansion, nor were Nazis drawn disproportionately from the less active (if anything, the contrary).[6] Only *after* Hitler came to power, with the Nazi *Gleichschaltung* of all associations, did activity decline. Allen's results were read, (again, I think, rightly) as having strongly impugned an otherwise plausible theory.

A central assertion of Immanuel Wallerstein's *Modern World-System,* volume 1 (1974), was that the "core" states of the world economy, from the sixteenth century onward, had been likeliest to develop strong states (in order to guarantee capitalist property rights and to protect trade routes) and to pursue linguistic and cultural homogeneity (in order to lower administrative and transaction costs). Yet as Gourevitch and others quickly observed, it was, in fact, a central European state of what Wallerstein had called the "semiperiphery" (i.e., Prussia) that developed arguably the strongest state in the early modern world and that came earliest to mass education and the pursuit of linguistic homogeneity (1978, esp. 423–27). The case seriously undermined this aspect of Wallerstein's theory; but Gourevitch went on to speculate—and Charles Tilly (1990) has subsequently advanced considerable argument and evidence to show—that in fact, the correlation was the reverse: the economically most advanced early modern states were often the least powerful, and vice-versa.

Against the record amassed by these and other single-observation studies, KKV contend that in general, "the single observation is not a useful technique for testing hypotheses or theories" (p. 211), chiefly because measurement error may yield a false negative, omitted variables may yield an unpredicted result, or social-scientific theories are insufficiently precise.[7] They would have us accept that the Lijphart, Allen, and Gourevitch studies—and even more the sweeping inferences that most comparativists drew from them—were bad science; as KKV state explicitly, falsification from a single observation "'is not the way social science is or should be conducted" (p. 103).

Rudolf Heberle's (1963, 1970) justly famous exploration of Nazi support in Schleswig-Holstein is exemplary in doing what KKV call "making many observations from few" (p. 217); yet it, too, would presumably fail to meet their standard. Long before Barrington Moore, Jr. (1967), solidified the thesis, analysts had conjectured a close link between labor-repressive agriculture and susceptibility to fascism. It occurred to Heberle that the north German state of Schleswig-Holstein offered an ideal test of the thesis, containing, as it did, three distinct agricultural regions, characterized respectively by (1) plantation agriculture on the East Elbian, or "Junker," model (the Hill district); (2) prosperous family farms like those of western and southwestern Germany (the Marsh); and (3) hardscrabble, quasi-subsistence farming (the *Geest*). The asserted link to feudalism would predict the earliest and strongest Nazi support in the first of these regions; but in fact, the fascist breakthrough occurred in the *Geest,* among the marginalized subsistence farmers; the family farmers came along only considerably later, and the feudal region resisted almost to the end. This brilliantly designed little study thus seriously undermined, even before its precise formulation, what has since come to be known as the "Moore thesis" about the origins of fascism.

Like Atul Kohli's (1987) three-state study of poverty policy in India, Heberle's examination inventively exploits within-country—in Heberle's case, within-region—variation. Yet KKV dismiss precisely this aspect of Kohli's analysis, on the ground that the values of both the explanatory and the dependent variables were known in advance; "Selection, in effect, is on both the explanatory and dependent variables," so that "the design provides no information about his causal hypothesis" (KKV, 145). Of course, Heberle, by confining his attention to a single state, partially constrained himself against biased selection; but Schleswig-Holstein itself might represent only random variation, and so (they would surely say) could not be taken as refuting the hypothesized causal link between feudalism and fascism. Again, I think, their strictures, taken literally, would dismiss a brilliant study as bad (or at least inadequate) social science.

My final two examples raise the stakes considerably; for they represent, by common consent, the very best of recent work in comparative politics. Yet Peter Katzenstein's *Small States in World Markets* (1985), by KKV's lights, inadmissibly restricts variation on the independent and dependent variables;

and Robert Bates's *Markets and States in Tropical Africa* (1981) impermissibly selects on the dependent variable.

Katzenstein, contesting the conventional wisdom that only large states were independent enough to be worth studying, deliberately restricted his focus to the smaller European states and, within that set, to the smaller states that were "close to the apex of the international pyramid of success," thus "excluding Ireland, Finland, and some of the Mediterranean countries" (1985, 21). His reasons were straightforward: (1) the cases that he did study were *anomalous*, for small, price-taking countries were widely supposed to face particular challenges in an uncertain international environment; and (2) they were *forerunners*, in the sense that all countries were rapidly becoming as dependent on international markets as these small ones had long been. To examine why countries that theoretically should not succeed in fact did so (reminiscent of Lijphart's strategy) and to attempt to discern a possible path of adaptation of larger states, seemed, both to Katzenstein and to his generally enthusiastic readership, a sensible strategy. Yet KKV, at least as I read them, must hold Katzenstein guilty of two cardinal sins that largely vitiate his analysis: (1) instead of choosing his cases to guarantee some range of variation on the independent variable, he restricts his analysis to small (and therefore quite trade-dependent) states; and (2) more seriously, taking economic success or failure as his dependent variable, he looks only at instances of success.

Bates's book is an even clearer case of selection on the dependent variable. Exactly as Michael Porter's *Competitive Advantage of Nations* (1990) examines only cases of economic success and thus draws withering fire from KKV (pp. 133–34), Bates focuses almost entirely on cases of economic failure or, more precisely, on the remarkably uniform *pattern* of economic failure among the states of postindependence Africa. He nonetheless develops an account that most readers have found compelling: (1) that the failures all resulted from an economic policy that heavily taxed agricultural exports to subsidize investment in heavily protected manufactures; and (2) that this self-destructive economic policy was the inevitable result of a political constellation in which urban groups were organized and powerful, rural ones scattered and weak. While Bates supports his analysis by observing that the two African cases of relative economic success (i.e., Kenya and Côte d'Ivoire) were characterized by export-friendly policy and politically more powerful farmers, this part of his discussion is brief and clearly tangential to his main argument.

Why, despite their seemingly egregious sins,[8] are all of these works believed by most comparativists—rightly, in my judgment—to have provided convincing inferences about their topics of study? Chiefly, I submit, for two reasons, which shed much light on the problems of KKV's account: (1) all of them tested, relied on, or proposed, clear and precise *theories*; and (2) all focused on *anomalies*, either in prevailing theories or in the world—cases that contradicted received beliefs or unexpected regularities that were too pronounced to be accidental.

The theories of cross-cutting cleavages (Truman 1951), atomization (e.g., Kornhauser 1959), world-systems (Wallerstein 1974), and feudal legacy (Moore 1967) had the great advantage of being precise enough to yield implications for single, or for very few, observations: Lijphart, Sheridan Allen, Gourevitch, and Heberle, respectively, took brilliant scholarly advantage of that precision: (1) to seek out anomalous cases and, usually, (2) to conjecture intelligently about a more satisfactory general theory that could avoid such anomalies.

About small states and heavy reliance on external markets there was less a prevailing theory than a prevailing prejudice—that puniness entailed constraint, insecurity, and (barring extraordinary good luck) economic trouble. By adducing seven cases of small states that had consistently prospered, Katzenstein demonstrated that insecurity and poverty were far from inevitable; by showing that their strategies, in similar circumstances, had differed, he proved that they retained considerable freedom of policy; and by analyzing their marked similarities of historical development and present-day governance, he advanced a plausible (if in this work still conjectural)[9] theory of situational requisites for highly trade-dependent states.

The African economic devastation that Bates studied was usually "explained" by a mélange of misunderstood Marxism and economic illiteracy that stressed the "dependence" of the Third World on the First. By invoking standard, simple economics, Bates easily showed that local policy, not First World plots, must be to blame: if domestic agricultural prices were systematically suppressed, one would expect to see smuggling and rural flight; if domestic industry was protected and subsidized, cartels, uncompetitive goods, and an overvalued currency; if taxes and controls poured power and resources into the hands of bureaucrats, a bloated public sector and vicious competition for place and favor. In each African case, all of these in fact prevailed, and no amount of external "dependence" could so easily explain this particular concatenation of disasters.

Yet this left a riddle no less profound than the original one: Why should almost all governments of the region have deliberately chosen policies so inimical to aggregate welfare and to long-term growth? Just as a psychologist might become intrigued if all but one or two of the people on a certain street began suddenly to mutilate themselves, Bates pursued a "cluster analysis" (see pp. 148–49) of perverse African policies and reached his highly plausible conjecture that rural weakness produced a fatal "urban bias" (see Lipton 1976) in policy.[10]

In the works of Katzenstein and Bates, then, no less than in those cited earlier, the crucial ingredient was clear, precise, powerful ("high-leverage") theory with what Lave and March (1975) tellingly called a "sense of process," that is, intuitively plausible causal links. In both accounts, universally accepted economic theory underpinned the critique of received wisdom: if small, price-taking firms survived in uncertain markets, why not small, price-taking countries; if all of the symptoms of the African cases were consistent with systematic price distortions, what other diagnosis was possible? The core of Katzenstein's alternative account was a story about how democratic corporatism facilitated flexible adjustment to external markets; the core of Bates's, a hypothesized link between power and policy. That both arguments were so clear, plausible, and precise contributed crucially to their persuasiveness.

KKV, in contrast, frequently choose as examples hypotheses that seem obvious or that lack deductive fertility. To prove, for example, that declining communist societies were likelier to spawn mass movements of opposition the less repressive the old regime was (p. 127) neither contravenes received wisdom nor carries broader implications for other cases.

The aspects of larger theory and of "sense of process," consequently seem to be sorely absent from the KKV prescriptions for social inquiry. While the authors are right to fear our natural tendency to see patterns where none exist (p. 21), they emphasize insufficiently the centrality of patterns—indeed, of "paradigms" (Kuhn, 1962)—to efficient scientific inquiry. A powerful, deductive, internally consistent theory can be seriously undermined, at least in comparative politics, by even one wildly discordant observation (Lijphart's Netherlands). On the positive side, a powerful theory can, by explaining an otherwise mysterious empirical regularity (European small-state corporatism, African economic failure), gain provisional acceptance at least as a highly plausible conjecture worthy of further research. As most discussions of spurious correlation make clear, we gain confidence in a proposed explanation to the extent that it *both* (1) fits the data *and* (2) "makes sense" in terms of its consistency with other observations and its own deductive implications. KKV, it seems to me, emphasize the former at the expense of the latter. In consequence, their advice to area specialists focuses almost entirely on "increasing the number of observations" (chap. 6). Many comparativists, I think, would instead counsel, "Choose better theory, which can make better use of few or single observations."[11]

Valuable as KKV's strictures are, I fear that devout attention to them may paralyze, rather than stimulate, scientific inquiry in comparative politics. They write eloquently and insightfully about the trade-offs between close observation of a few cases and more cursory measurement of many (chap. 2, esp. pp. 66–68); I wish they had as perceptively discussed how better theory permits inference from fewer cases, allows restriction on the independent variable, and may even profit from judicious selection on the dependent variable.

In short, I suspect they do not mean quite as stern a message as they send; or perhaps they view the studies I have discussed here in a different and more redeeming light. They would, however, have spoken more clearly to comparativists if they had specifically addressed the major literature of the less quantitative tradition.

Notes

1. The only competition, long out of print and aimed more at the advanced undergraduate level, is probably Lave and March 1975.

2. Eckstein characterized this as the strategy of the "least likely" case 1975, 118–19. See Hempel 1966, 37–38.

3. I owe this citation to Mark Lichbach.

4. "Dear Russell: Einstein's theory is completely confirmed. The predicted displacement was 1". 72 and the observed 1" .75 ± .06. Yours, J.E.L." (J. E. Littlewood to Bertrand Russell, 1919, quoted in Russell 1969, 149).

5. Lijphart went on to conjecture, on the basis of the Dutch case, about the precise circumstances in which non-cross-cutting cleavages were compatible with civic peace; but that is secondary to the point I am arguing here.

6. To be sure, KKV distinguish between *cases* and *observations*; and Allen's study could be read as a single case that by examining a variety of groups and individuals, encompasses many observations. Such a reading, in my view, would fundamentally misunderstand the underlying theory, whose central independent variable is the level of association that individuals encounter. Given the theory, the town (or, at most, the class within the town) is the relevant observation; and Allen's study is therefore a single case *and* a single observation.

7. Their strictures on the first two points are so sweeping that they must, by implication, include theories and hypotheses in the physical sciences. Hence I take it that they would also reject the confirmation of the theory of relativity and other cases alluded to by Hempel (1966, 77), which rested on single observations.

8. As regards selection on the dependent variable, KKV take a particularly Draconian stand: "We can. . .learn nothing about a causal effect from a study which selects observations so that the dependent variable does not vary" (p. 147).

9. To be sure, by looking only at successful small European states, Katzenstein had to leave open the possibilities (1) that unsuccessful small states were also governed corporatively and (2) that small non-European states had discovered quite different recipes for success.

10. It is worth noting that Bates has pursued this conjecture *not* through any large-n study, but by close analysis of an apparently anomalous case: Colombia, where dispersed coffee farmers of modest means prevailed politically not only against citydwellers but over concentrated plantation owners of considerable wealth.

11. As I note at the outset, KKV do discuss—at some length and quite sensibly—some major characteristics of good theory (sec. 3.5). They seem, however to despair that social-scientific theories can ever be precise enough to permit valid inference from few cases (pp. 210–11); and they explicitly reject parsimony as an inherently desirable property of social-scientific theory (pp. 20, 104–5). On neither point, I suspect, will most comparativists find their arguments persuasive; and they seem to me to be refuted by the examples I adduce here.

Review: Bridging the Quantitative-Qualitative Divide in Political Science
Reviewed Work(s): Designing Social Inquiry: Scientific Inference in Qualitative Research by Gary King, Robert O. Keohane and Sidney Verba
Review by: Sidney Tarrow
Source: *The American Political Science Review*, Vol. 89, No. 2 (Jun., 1995), pp. 471-474
Published by: American Political Science Association
Stable URL: http://www.jstor.org/stable/2082444
Accessed: 08-03-2017 17:19 UTC

BRIDGING THE QUANTITATIVE-QUALITATIVE DIVIDE IN POLITICAL SCIENCE

SIDNEY TARROW *Cornell University*

In *Designing Social Inquiry*, Gary King, Bob Keohane and Sidney Verba (KKV) have performed a real service to qualitative researchers. I, for one, will not complain if I never again have to look into the uncomprehending eyes of first-year graduate students when I enjoin them (pace Przeworski and Teune) to "turn proper names into variables." The book is brief and lucidly argued and avoids the weighty, muscle-bound pronouncements that are often studded onto the pages of methodological manuals.

But following KKV's injunction that "a slightly more complicated theory will explain vastly more of the world" (p. 105), I will praise them no more but focus on an important weakness in the book. Their central argument is that the same logic that is "explicated and formalized clearly in discussions of quantitative research methods" underlies—or should—the best qualitative research (p. 4). If this is so, then they really ought to have paid more attention to the *relations* between quantitative and qualitative approaches and what a rigorous use of the latter can offer quantifiers. But while they offer a good deal of generous (at times patronizing) advice to qualitatively oriented scholars, they say very little about how qualitative approaches can be combined with quantitative research. Especially with the growth of choice-theoretic approaches, whose users often illustrate their theories with stories, there is a need for a set of ground rules on how to make intelligent use of qualitative data.

KKV do not address this issue. Rather, they use the model of quantitative research to advise qualitative researchers on how best to approximate good models of descriptive and causal inference. (Increasing the number of observations is their cardinal operational rule.) But in today's social science world, how many social scientists can be simply labeled "qualitative" or "quantitative"? How often, for example, do we find support for sophisticated game-theoretic models resting on the use of anecdotal reports or on secondary evidence lifted from one or two qualitative sources? More and more frequently in today's social science practice, quantitative and qualitative data are interlarded within the same study. A recent work that KKV warmly praise illustrates both that their distinction between quantitative and qualitative researchers is too schematic and that we need to think more seriously about the interaction of the two kinds of data.

Marinating Putnam

In Robert Putnam's (1993) analysis of Italy's creation of a regional layer of government, *Making Democracy Work*, countless elite and mass surveys and ingenious quantitative measures of regional performance are arrayed for a 20-year period of regional development. On top of this, he conducted detailed case studies of the politics of six Italian regions, gaining, in the process, what KKV recommend as "an intimate knowledge of the internal political manoeuvering and personalities that have animated regional politics over the last two decades" (p. 5) and Putnam calls "marinating yourself in the data" (Putnam 1993, 190). KKV use *Making Democracy Work* to praise the virtues of "soaking and poking," in the best Fenno tradition (p. 38).

But Putnam's debt to qualitative approaches is much deeper and more problematic than this; for after spending two decades administering surveys to elites and citizens in the best Michigan mode, he was left with the task of explaining the sources of the vast differences he had found between Italy's north-central and southern regions. To find them, his quantitative evidence offered only indirect evidence; and he turned to history, repairing to the halls of Oxford, where he delved deep into the Italian past to fashion a provocative interpretation of the superior performance of the northern Italian regional governments vis-à-vis the southern ones. This he based on the civic traditions of the (northern) Renaissance city-states, which, according to him, provide "social capital" that is lacking in the traditions of the South (chap. 5). A turn to qualitative history (probably not even in Putnam's mind when he designed the project) was used to interpret cross-sectional, contemporary quantitative findings.

Putnam's procedure in *Making Democracy Work* pinpoints a problem in melding quantitative and qualitative approaches that KKV's canons of good scientific practice do not help to resolve. For in delving into the qualitative data of history to explain our quantitative findings, by what rules can we choose the *period* of history that is most relevant to our problem? And what *kind* of history are we to use; the traditional history of kings and communes or the history of the everyday culture of the little people? And how can the effect of a particular historical period be separated from that of the periods that precede or follow it? In the case of *Making Democracy Work*, for example, it would have been interesting to know (as Suzanne Berger asked at the 1994 APSA roundtable devoted to the book) by what rules of inference Putnam chose the Renaissance as determining of the North's late twentieth-century Italian civic superiority. Why not look to its sixteenth-century collapse faced by more robust monarchies, its nineteenth-century military conquest of the South, or its 1919–21 generation of fascism (not to mention its 1980s corruption-fed pattern of economic growth)? None of these are exactly "civic" phenomena; by what rules of evidence are they less relevant in "explaining" the northern regions' civic superiority over the South than the period of the Renaissance city-states? Putnam does not tell us; nor do KKV.

To generalize from the problem of Putnam's book, qualitative researchers have much to learn from the model of quantitative research. But their quantitative cousins who wish to profit from conjoining their findings with qualitative sources need, for the selection of qualitative data and the intersection of the two types, rules just as demanding as the rules put forward by KKV for qualitative research on its own. I shall sketch some useful approaches to bridging the quantitative-qualitative gap from recent examples of comparative and international research.

Tracing Processes To Interpret Decisions

One such rule that KKV cite favorably is the practice of *process tracing*, in which the researcher looks closely at 'the decision process by which various initial conditions are translated into outcomes' " (p. 226, quoting George and McKeown 1985, 35). But even here, KKV interpret the advantages of process tracing narrowly, assimilating it to their favorite goal of increasing the number of theoretically relevant observations (p. 227). As George and McKeown actually conceived it, the goal of process tracing was not to increase the number of discrete decision stages and aggregate them into a larger number of data points but to *connect* the phases of the policy process and enable the investigator to identify the reasons for the emergence of a particular decision through the dynamic of events (George and McKeown 1985, 34–41). Process tracing is different *in kind* from observation accumulation and is best employed in conjunction with it—as was the case, for example, in the study of cooperation on economic sanctions by Lisa Martin (1992) that KKV cite so favorably.

Systematic and Nonsystematic Variable Discrimination

KKV give us a second example of the uses of qualitative data but, once again, underestimate its particularity. They argue that the variance between different phenomena "can be conceptualized as arising from two separate elements: *systematic* and *nonsystematic* differences," the former more relevant to fashioning generalizations than the latter (p. 56). For example, in the case of conservative voting in Britain, systematic differences include such factors as the properties of the district, while unsystematic differences could include the weather or a flu epidemic at the time of the election. "Had the 1979 British elections occurred during a flu epidemic that swept through working-class houses but tended to spare the rich," they conclude, "our observations might be rather poor measures of underlying Conservative strength" (pp. 56–57).

Right they are, but this piece of folk wisdom hardly exhausts the importance of nonsystematic variables in the interpretation of quantitative data. A good example comes from how the meaning and extension of the strike changed as systems of institutionalized industrial relations developed in the nineteenth century. At its origins, the strike was spontaneous, uninstitutionalized, and often accompanied by whole-community "turnouts." As unions developed and governments recognized workers' rights, the strike broadened to whole sectors of industry, became an institutional accompaniment to industrial relations, and lost its link to community collective action. The systematic result of this change was permanently to affect the patterns of strike activity. Quantitative researchers like Michelle Perrot (1986) documented this change. But had she regarded it only as a case of "nonsystematic variance" and discarded it from her model, as KKV propose, Perrot might well have misinterpreted the changes in the form and incidence of the strike rate. Because she was as good a historian as she was a social scientist, she retained it as a crucial change that transformed the relations between the strike incidence and industrial relations.

To put this more abstractly, distinct historical events often serve as the tipping points that explain the interruptions in an interrupted time-series, permanently affecting the relations between the variables (Griffin 1992). Qualitative research that turns up "nonsystematic variables" is often the best way to uncover such tipping points. Quantitative research can then be reorganized around the shifts in variable interaction that such tipping points signal. In other words, the function of qualitative research is not only, as KKV seem to argue, to peel away layers of unsystematic fluff from the hard core of systematic variables but also to assist researchers to understand shifts in the value of the systematic variables.

Framing Qualitative Research within Quantitative Profiles

These two uses of qualitative data pertain largely to aiding quantitative research. But this is not the only way in which social scientists can combine quantitative and qualitative approaches. Another is to focus on the qualitative data, using a systematic quantitative data base as a frame within which the qualitative analysis is carried out. Case studies have been validly criticized as being based on often dramatic but frequently unrepresentative cases. Studies of successful social revolutions often possess characteristics that may also be present in unsuccessful revolutions, rebellions, riots, and ordinary cycles of protest (Tilly 1993, 12–14). In the absence of an adequate sample of revolutionary episodes, no one can ascribe particular characteristics to a particular class of collective action.

The representativity of qualitative research can never be wholly assured until the cases become so numerous that the analysis comes to resemble quantitative research (at which point the qualitative research risks losing its particular properties of depth, richness and process tracing). But framing it within a quantitative data base makes it possible to avoid generalizing on the occasional "great event" and points to less dramatic—but cumulative—historical trends.

Scholars working in the "collective action event" history tradition have used this double strategy with success. For example, in his 1993 study of over seven hundred revolutionary years in over five hundred years of European history, Charles Tilly assembled data that could have allowed him to engage in a large-N study of the correlates and causes of revolution. Tilly knows how to handle large time-series data sets as well as anybody. But he did not believe that the concept of *revolution* had the monolithic quality that other social scientists had assigned to it (1993, chap. 1). So he resisted the temptation for quantification, using his data base, instead, to frame a series of regional time-series narratives that depended as much on his knowledge of European history as on the data themselves. When a problem cried out for systematic quantitative analysis (e.g., when it came to periodizing nationalism), Tilly (1994) was happy to exploit the quantitative potential of the data. But the quantitative data set served mainly as a frame for qualitative analysis of representative regional and temporal revolutionary episodes and series of episodes.

Putting Qualitative Flesh on Quantitative Bones

These examples are possibly exotic to the traditions of much of American social science practice. But an American sociologist, Doug McAdam, has shown how social science can be enriched by combining quantitative and qualitative approaches to the same data base. McAdam's 1988 study of Mississippi Freedom Summer participants was based on a treasure-trove of quantifiable data—the original questionnaires of the prospective Freedom Summer volunteers. While some of these young people eventually stayed home, others went south to register voters, teach in "freedom schools" and risk the dangers of Ku Klux Klan violence. Two decades later, both the volunteers and the no-shows could be interviewed by a researcher with the energy and the imagination to go beyond the use of canned data banks.

McAdam's main analytic strategy was to carry out a paired comparison between the questionnaires of the participants and the stay-at-homes and to interview a sample of the former in their current lives. This systematic comparison formed the analytical spine of the study and of a series of technical papers. But except for a table or two in each chapter, the texture of *Freedom Summer* is overwhelmingly qualitative. McAdam draws on his interviews with former participants, as well as on secondary analysis of other people's work, to get inside the Freedom Summer experience and to highlight the effects that participation had on their careers and ideologies and their lives since 1964. With this combination of quantitative and qualitative approaches, he was able to tease a convincing picture of the effects of Freedom Summer activism from his data.

As I write this, I imagine KKV exclaiming, "But this is *precisely* the direction we would like to see qualitative research moving—toward expanding the number of observations and respecifying hypotheses to allow them to be tested on different units!" (see chap. 6). But would they argue, as I am, that it is the *combination* of quantitative and qualitative methods trained on the same problem (not a move toward the logic of quantitative analysis alone) that is desirable? Two more ways of combining these two logics illustrate my intent.

Sequencing Quantitative and Qualitative Research

The growth industry of qualitative case studies that followed the 1980–81 Solidarity movement in Poland largely took as given the idea that Polish intellectuals had the most important responsibility for the birth and ideology of this popular movement. There was scattered evidence for this propulsive role of the intellectuals; but since most of the books that appeared after the events were written by them or by their foreign friends, an observer bias might have been operating to inflate their importance in the movement vis-à-vis the working class that was at the heart of collective action in 1980–81 and whose voice was less articulate.

Solid quantitative evidence came to the rescue. In a sharp attack on the "intellectualist" interpretation and backed by quantitative evidence from the strike demands of the workers themselves, Roman Laba showed that their demands were overwhelmingly oriented toward trade union issues and showed little or no effect of the proselytizing that Polish intellectuals had supposedly been doing among the workers of the Baltic coast since 1970 (1991, chap. 8). This finding dovetailed with Laba's own qualitative analysis of the development of the workers' movement in the 1970s and downplayed the role of the Warsaw intellectuals who had been at the heart of a series of books by their foreign friends.

The response of those who had been responsible for the intellectualist interpretation of Solidarity was predictably violent. But there were also more measured responses that shed new light on the issue. For example, prodded by Laba's empirical evidence of worker self-socialization, Jan Kubik returned to the issue with both a sharper analytical focus and better qualitative evidence than the earlier intellectualist theorists had employed, criticizing Laba's conceptualization of class and reinterpreting the creation of Solidarity as "a multistranded and complicated social entity . . . created by the contributions of various people" whose role and importance he proceeded to demonstrate (1994, 230–38). Moral: a sequence of contributions using different kinds of evidence led to a clearer and more nuanced understanding of the role of different social formations in the world's first successful confrontation with state socialism.

Triangulation

I have left for last the research strategy that I think best embodies the strategy of combining quantitative and qualitative methods—the *triangulation* of different methods on the same problem. Triangulation is particularly appropriate in cases in which quantitative data are partial and qualitative investigation is obstructed by political conditions. For example, Valerie Bunce used both case methodology and quantitative analysis to examine the policy effects of leadership rotation in Western and socialist systems. In *Do New Leaders Make a Difference?*, she wrote, "I decided against selecting one of these approaches to the neglect of the other [the better] to test the impact of succession on public policy by employing *both* methodologies" (1981, 39).

Triangulation is also appropriate in specifying hypotheses in different ways. Consider the classical Tocquevillian insight that regimes are most susceptible to a political opportunity structure that is partially open. The hypothesis takes shape in two complementary ways: (1) that liberalizing regimes are more susceptible to opposition than either illiberal or liberal ones; and (2) that within the same constellation of political units, opposition is greatest at intermediate levels of political opportunity. Since there is no particular advantage in testing one version of the hypothesis over the other, testing both is optimal (as can be seen in the recent social movement study, Kriesi et al. 1995).

My final example of triangulation comes, with apologies, from my own research on collective action and social movements in Italy. In the course of a qualitative reconstruction of a left-wing Catholic "base community" that was active in a peripheral district of Florence in 1968, I found evidence that linked this movement discursively to the larger cycle of student and worker protest going on in Italy at the same time (Tarrow 1988). Between 1965 and 1968, its members had been politically passive, focusing mainly on neighborhood and educational issues. But as the worker and student movements exploded around it in 1968, their actions became more confrontational, organized around the themes of autonomy and internal democracy that were animating the larger worker and student movements around them.

Researchers convinced of their ability to understand political behavior by interpreting "discourse" might have been satisfied with these observations; but I was not. If nothing else, Florence was only one case among potential thousands. And in today's global society, finding thematic similarity among different movements is no proof of direct diffusion, since many movements around the world select from the same stock of images and frames without the least connection among them (Tarrow 1994, chap. 11).

As it happened, quantitative analysis came to the rescue to triangulate on the same problem. For a larger study, I had collected a large sample of national collective action events for a period that bridged the 1968 Florentine episode. And as it also happened, two Italian researchers had collected reliable data on the total number of religious "base communities" like the Florentine one throughout the country (Sciubba and Pace 1976). By reoperationalizing the hypothesis cross-sectionally, I was able to show a reasonably high positive correlation ($R = .426$) between the presence of Catholic base communities in various cities and the magnitude of general collective action in each city (Tarrow 1989, 200). A longitudinal, local, and qualitative case study triangulated with the results of cross-sectional, national, and quantitative correlations to turn my intuitive hunch that Italy in the 1960s underwent an integrated cycle of protest into a more strongly supported hypothesis.

KKV are not among those social scientists who believe that quantification is the answer to all the problems of social science research. But their single-minded focus on the logic of quantitative research (and of a certain *kind* of quantitative research) leaves underspecified the particular contributions that qualitative approaches make to scientific research, especially when combined with quantitative research. As quantitatively trained researchers shift to choice-theoretic models backed up by illustrative examples (often containing variables with different implicit metrics), the role of qualitative research grows more important. We are no longer at the stage when public choice theorists can get away with demonstrating a theorem with an imaginary aphorism. We need to develop rules for a more systematic use of qualitative evidence in scientific research. Merely wishing that it would behave as a slightly less crisp version of quantitative research will not solve the problem.

This is no plea for the veneration of historical uniqueness and no argument for the precedence of "interpretation" over inference. (For an excellent analysis of the first problem, see KKV pp. 42–3 and of the second, pp. 36–41.) My argument, rather, is that a single-minded adherence to *either* quantitative or qualitative approaches straightjackets scientific progress. Whenever possible, we should use qualitative data to interpret quantitative findings, to get inside the processes underlying decision outcomes, and to investigate the reasons for the tipping points in historical time-series. We should also try to use different kinds of evidence together and in sequence and look for ways of triangulating different measures on the same research problem.

KKV have given us a spirited, lucid, and well-balanced primer for training our students in the essential unity of social science work. Faced by the clouds of philosophical relativism and empirical nominalism that have recently blown onto the field of social science, we should be grateful to them. But their theoretical effort is marred by the narrowness of their empirical specification of qualitative research and by their lack of attention to the qualitative needs of quantitative social scientists. I am convinced that had a final chapter on combining quantitative and qualitative approaches been written by these authors, its spirit would not have been wildly at variance with what I have argued here. As it is, someone else will have to undertake that effort.

Notes

I wish to thank Henry Brady, Miriam Golden, Peter Katzenstein, David Laitin, Peter Lange, Doug McAdam, Walter Mebane, Robert Putnam, Shibley Telhami and Charles Tilly for their comments on drafts of this review.