Introduction to Rethinking Social Inquiry, 2 edn., Henry E. Brady, and David Collier, ed. (Rowman and Littlefield, 2010).

A Sea Change in Political Methodology

David Collier, Henry E. Brady, and Jason Seawright

We begin with rival claims about the "science" in social science. In our view, juxtaposing these claims brings into focus a sea change in political science methodology.

King, Keohane, and Verba's (KKV) 1994 book, *Designing Social Inquiry*,[2] proposes a bold methodological agenda for researchers who work in the qualitative tradition. The book's subtitle directly summarizes the agenda: "*scientific* inference in qualitative research" (italics added). To its credit, the book is explicit in its definition of science. It draws on what we and many others have viewed as a "quantitative template," which serves as the foundation for the desired scientific form of qualitative methods. In KKV's view, standard research procedures of qualitative analysis are routinely problematic, and ideas drawn from conventional quantitative methods are offered as guideposts to help qualitative researchers be scientific.

---

1. For our own work, we share Freedman's view of plurality in scientific methods, and we recognize social versus natural science as partially different enterprises. Yet the two can and should strive for careful formulation of hypotheses, intersubjective agreement on the facts being analyzed, precise use of data, and good research design. With this big-tent understanding of science, we are happy to be included in the tent.

2. As explained above in the preface, in the second edition we use the abbreviation KKV to refer to the book, rather than *DSI*, as in the first edition.

A starkly different position has been emerging over a number of years, forcefully articulated by the statistician David A. Freedman in chapter 11 of the present volume. He reviews the central role of qualitative analysis in six major breakthroughs from the history of epidemiology—a field highly relevant to political science because it faces many of the same challenges of doing large-N analysis with observational data and because, as Freedman insists, one does indeed find interesting opportunities for qualitative insight. He argues, in fact, that in epidemiology as well as the social sciences, qualitative analysis is indeed a "type of *scientific* inquiry" (italics added), within the framework of recognizing multiple types. In characterizing this form of quantitative analysis, Freedman employs the expression "causal-process observation" (CPO—a term of central importance to the present volume).[3] In his view, such strategically selected pieces of evidence play a critical role in disciplined causal inference. Freedman comments pointedly on the contributions of CPOs.

> Progress depends on refuting conventional ideas if they are wrong, developing new ideas that are better, and testing the new ideas as well as the old ones. The examples show that qualitative methods can play a key role in all three tasks . . . . (chap. 11, this volume)

Relatedly, Freedman underscores the fragility of the quantitative template.

> Indeed, far-reaching claims have been made for the superiority of a quantitative template that depends on modeling—by those who manage to ignore the far-reaching assumptions behind the models. However, the assumptions often turn out to be unsupported by the data. . . . If so, the rigor of advanced quantitative methods is a matter of appearance rather than substance. (chap. 11, this volume)

In this Introduction, against the backdrop of these starkly contrasting views of appropriate methods, we examine new developments in methodology that have framed our approach to the second edition of *Rethinking Social Inquiry*. The discussion focuses on: (1) ongoing controversy regarding KKV's legacy; (2) growing criticism of the standard quantitative template, including regression modeling, significance tests, and estimates of uncertainty; and (3) emerging arguments about both qualitative and quantitative methods that hold the promise of greatly strengthening tools for causal inference.

3. We define a causal-process observation as an insight or piece of data that provides information about context, process, or mechanism, and that contributes distinctively to causal inference. A data-set observation (DSO), by contrast, is the standard quantitative data found in a rectangular data set. See Glossary.

A further initial point should be underscored. The focus in both editions of *Rethinking Social Inquiry* is on the study of causes and consequences—and specifically on causal inference. Of course, this focus is just one facet of methodology. In our own work we have written extensively on conceptualization and measurement, and indeed, assessing causes and consequences emphatically calls for careful attention to concept formation and operationalization. Yet the central concern here is with causal inference.

ONGOING CONTROVERSY OVER KKV

The methodological positions adopted by KKV continue to be of great importance in political science and well beyond. The book has an exceptionally high level of citations, and year after year it has impressive sales rankings with online book sellers.

In the period since the publication of our first edition in 2004, quantitative and qualitative methodologists alike have underscored KKV's importance. Philip A. Schrodt, a quantitative methodologist, argues that it has been the "canonical text of the orthodox camp" among political methodologists. In many graduate programs, it is considered "the complete and unquestionable truth from on high" (Schrodt 2006: 335). On the qualitative side, James Mahoney notes the book's striking importance and remarkable impact in political science (2010: 120).

Ironically, achieving "doctrinal status was not necessarily the intention of KKV's authors" (Schrodt 2006: 336), and their perspectives have doubtless evolved in the intervening years. Yet notably, in 2002—eight years after the book's original publication—King published an extended, programmatic statement on methodology, nearly the length of a short book, entitled "The Rules of Inference" (Epstein and King 2002). This publication departs little from the arguments of KKV.[4]

KKV is controversial, as well as influential, and its continuing importance is of great concern to scholars disturbed by its narrow message. Our first edition already contained strong critiques, and new commentaries—some extremely skeptical—have continued to appear. These more recent arguments merit close examination.

Schrodt presents a bruising critique:

---

4. We were grateful for King, Keohane, and Verba's willingness to contribute their article "The Importance of Research Design" to our first edition, and we are very pleased to include it in this new edition. It contributes important ideas to the debate among authors who have commented on their original book. However, we do not see it as a substantial departure from their book.

> KKV establishes as the sole legitimate form of social science a set of rather idiosyncratic and at times downright counterintuitive frequentist statistical methodologies that came together . . . to solve problems quite distant from those encountered by most political scientists. . . . (2006: 336)

Schrodt views the book as promoting "a statistical monoculture" that is "not even logically consistent" (2006: 336). In his view, this raises the concern that

> one of the reasons our students have so much difficulty making sense of [KKV] is that in fact it does not make sense. (2006: 336)

Mahoney (2010), in his comprehensive essay "After KKV: The New Methodology of Qualitative Research," argues that KKV has "hindered progress in political science" by "controversially and perhaps unproductively promoting a singular quantitative approach" (2010: 121). Weyland, with obvious annoyance, suggests that the authors of KKV "offered to help out their inferentially challenged qualitative brethren," proposing that their work should be "as similar as possible to quantitative studies." The book in effect makes claims of "quantitative superiority" that "rest on problematic assumptions" (2005: 392), thereby reinforcing the mindset in which "qualitative research was often seen as lacking precision and rigor and therefore undeserving of the 'methods' label" (2005: 392).

These and other scholars have also noted the sharp contrast in views between KKV and our own book. For example, Benoît Rihoux sees a "polarized" discussion that reflects a "fierce methodological debate which cuts across the whole of empirical social science in North America" (2006: 333, 334).

In discussing our book, Schrodt suggests that in this polarized context, "adherents of the [methodological] orthodoxy consider the heresies proposed therein to be a distraction at best; a slippery slope . . . at worst" (2006: 335). To take one example, what we would view as one of the orthodox commentaries is found in Nathaniel Beck (2006, passim), who entitles his article "Is Causal-Process Observation an Oxymoron?"—thereby essentially dismissing a basic concept in our book. He repeatedly acknowledges that scholars should "understand their cases" (e.g., 350) and that qualitative evidence contributes to this background knowledge, but he questions the idea that causal-process observations meet acceptable standards for causal inference (352).

Schrodt views elements of the response to *Rethinking Social Inquiry* among mainstream quantitative methodologists as reflecting an unfortunate, defensive reaction. He argues that

> many in the statistical community have taken criticism of any elements of the orthodox approach as a criticism of all elements and circled the wagons rather than considering seriously the need for some reform. (Schrodt 2006: 338)

He also notes that when the editor of the methodology journal *Political Analysis* announced at the 2005 summer methodology meetings that the journal planned a symposium on *Rethinking Social Inquiry*, the room responded as if to express concern that "there are traitors in our midst!" (2006: 338). Schrodt comments that this resistance reflects "a worrisome contentment with the status quo" among quantitative methodologists (2006: 338).

Based on this discussion, it seems clear that major controversies over methods stand behind these criticisms. We now explore two of these controversies.

CRITICISM OF THE STANDARD QUANTITATIVE TEMPLATE

Our discussion here focuses on two facets of current criticism of the standard quantitative template, concerning basic ideas about statistical modeling and regression analysis, and alternative approaches to the important task of estimating uncertainty.

Statistical Modeling and Regression Analysis

In the past few years, the standard quantitative template centered on regression analysis has come under even heavier criticism. This development has two implications here. First, given KKV's reliance on this template, it further sharpens concern about the book's influence. Second, looking ahead, this development greatly extends the horizon of methodological approaches that should be—and in fact are being—discussed and applied, both among methodologists and consumers of alternative methods.

Much of this discussion centers on the enterprise of statistical modeling that stands behind regression analysis. In important respects, the precariousness of work with regression derives from the extreme complexity of statistical models. A statistical model may be understood as "a set of equations that relate observable data to underlying parameters" (Collier, Sekhon, and Stark 2010: xi—see Glossary). The values of these parameters are intended to reflect descriptive and causal patterns in the real world.

Constructing a statistical model requires assumptions, which often are not only untested, but largely untestable. These assumptions come into play "in choosing which parameters to include, the functional relationship between the data and the parameters, and how chance enters the model" (Collier, Sekhon, and Stark 2010: xi). Thus, debates on the precariousness of regression analysis are also debates on the precariousness of statistical models. It is unfortunate that more than a few quantitative researchers believe that when the model is estimated with quantitative data and results emerge that appear interpretable, it validates the model. This is not the case.

We agree instead with the political scientist Christopher H. Achen, who argues that with more than two or three independent variables, statistical models will "wrap themselves around any dataset, typically by distorting what is going on" (2002: 443). Thus, what we might call a "kitchen sink" approach—one that incorporates numerous variables—can routinely appear to explain a large part of the variance without yielding meaningful causal inference. Relatedly, Schrodt states that with just small modifications in the statistical model, estimates of coefficients can

> bounce around like a box of gerbils on methamphetamines. This is great for generating large bodies of statistical literature . . . but not so great at ever coming to a conclusion. (2006: 337)

The econometrician James J. Heckman emphasizes that "causality is a property of a model," not of the data, and "many models may explain the same data" (2000: 89). He observes that "the information in any body of data is usually too weak to eliminate competing causal explanations of the same phenomenon" (91).[5]

Sociologists have expressed related concerns, and Richard A. Berk concisely presents key arguments:

> Credible causal inferences cannot be made from a regression analysis alone. . . . A good overall fit does not demonstrate that a causal model is correct. . . . There are no regression diagnostics through which causal effects can be demonstrated. There are no specification tests through which causal effects can be demonstrated. (2004: 224)

Berk amusingly summarizes his views in section headings within the final chapter of his book on regression analysis: "Three Cheers for Description," "Two Cheers for Statistical Inference," and "One Cheer for Causal Inference" (2004: chap. 11).[6]

Mathematical statisticians have likewise confronted these issues. Freedman's skepticism about regression and statistical modeling has already been noted above, and his incisive critiques of diverse quantitative methods have now been brought together in an integrated volume that ranges across a broad spectrum of methodological tools (Freedman 2010).

---

5. From the standpoint of econometrics, see also Leamer (1983, 36–38).

6. Related arguments of sociologists have been advanced by Morgan and Winship (2007: passim), Hedström (2008: 324), and many other authors who have developed these themes. Statements by psychometricians include Cliff (1983, 116–18) and Loehlin (2004, 230–34).

Also from the side of mathematical statistics, Persi Diaconis argues that "large statistical models seem to have reached epidemic proportions" (1998: 797), and he laments the harm they are causing. He states that "there is such a wealth of modeling in the theoretical and applied arenas that I feel a sense of alarm" (804). Given these problems, methodologists should take more responsibility for the epidemic of statistical models by advocating "defensive statistics" (1998: 805). Thus, it should be a professional obligation to proactively warn scholars about the host of methodological problems summarized here.

In sum, many authors are now expressing grave concern about methods that have long been a mainstay of political and social science, and that are foundational in KKV's approach.

Estimating Uncertainty

Standard practices in mainstream quantitative methods for estimating the uncertainty of research findings have also been challenged. The quest to estimate uncertainty is quite properly a high priority, prized as a key feature of good research methods. KKV views understanding and estimating uncertainty as one of four fundamental features of scientific research (1994: 9). In its discussion of "defining scientific research in the social sciences," the book states that "without a reasonable estimate of uncertainty, a description of the real world or an inference about a causal effect in the real world is uninterpretable" (9). The received wisdom on these issues is central to mainstream quantitative methods.

Unfortunately, KKV presumes too much about how readily uncertainty can be identified and measured. In conjunction with the original debate over KKV, for example, Larry M. Bartels (chap. 4, this volume: 86–87) argues that these authors greatly overestimate the value of the standard insight that random error on an independent variable biases findings in knowable ways, whereas such error on the dependent variable does not. Bartels demonstrates that this would-be insight is incorrect.

A more pervasive problem involves significance tests. Any scholar acquainted with conventional practice in reporting regression results is well aware of the standard regression table with "tabular asterisks" scattered throughout.[7] The asterisks indicate levels of statistical significance, calculated on the basis of the standard errors of the coefficients in the table. Too often, when researchers report their causal inferences they simply identify the coefficients that reach a specified level of statistical significance. This is a dubious research practice.

A central problem here is that findings reported in regression tables are

7. Meehl (1978), cited in Freedman and Berk (2010: 24).

routinely culled from numerous alternative specifications of the regression model, which obviates the standard meaning and interpretation of the asterisks. Once again, Schrodt states the objection with particular clarity:

> The ubiquity of exploratory statistical research has rendered the traditional frequentist significance test all but meaningless. (2006: 337)

Freedman and Berk (2010: 24) underscore the dependence of significance tests on key assumptions. For descriptive inference (external validity), they assume a random sample, rather than the convenience sample common in political science. Even with a random sample, missing data—including the problem of non-respondents—can make it more like a convenience sample.[8] Another assumption requires a well-defined—rather than ill-defined or somewhat arbitrarily defined—population. For causal inference (internal validity), avoiding data snooping is crucial if significance tests are to be meaningful. Here, the presumption is that the researcher has begun with a particular hypothesis and tested it only once against the data, rather than several times, adjusting the hypothesis and model specification in the search for results deemed interesting. This inductive approach is *definitely* a valuable component of creative research, but it muddies the meaning of significance tests.

Against this backdrop, Freedman, Pisani, and Purves (2007) are blunt and—as usual—entertaining in their warnings on significance tests.

1. "If a test of significance is based on a sample of convenience, watch out" (556).
2. "If a test of significance is based on data for the whole population, watch out" (556).
3. "Data-snooping makes P-values hard to interpret" (547).
4. "An 'important' difference may not be statistically significant if the N is small, and an unimportant difference can be significant if the N is large" (553).[9]

A key point should be added. In his various single-authored and coauthored critiques of significance tests, Freedman does not turn to the alternative of Bayesian analysis. Rather, as in his other writings on methodology (see, e.g. chap. 11, this volume), he advocates common sense, awareness

8. See Freedman (2008b: 15). Thus, starting with a random sample, in the face of problems such as resource constraints that limit tracking down respondents, the researcher can end up with what is in effect a type of convenience sample.

9. I.e., if assumptions are not met, "significance" level depends on the sample size, without reflecting the real meaning of statistical significance.

that statistical tools have major limitations, and substantive knowledge of cases as an essential foundation for causal inference.

WHERE DO WE GO FROM HERE?

The practical importance of these problems is quickly seen in the fact that, to a worrisome degree, a great deal of quantitative research in political science has proceeded as if regression-based analysis, including associated measures of uncertainty, yields reliable causal inference. A vast number of journal articles have sought to make causal inferences by estimating perhaps half a dozen related (though quite typically under-theorized) model specifications, picking and choosing among these specifications, and offering an ad hoc interpretation of a few selected coefficients—generally, quite inappropriately, on the basis of significance levels. These failings have been further exacerbated by the readily available statistical software that makes it easy for researchers with virtually no grasp of statistical theory to carry out complex quantitative analysis (Steiger 2001).

In the face of these grave problems, we explore two avenues of escape: first, new developments in quantitative analysis; and second, continuing innovation in qualitative methods, which offer a very different means of addressing these difficulties. In our own work, and in scholarship more broadly, quantitative methods are of course deemed to be of enormous importance in their own right, and this continuing innovation certainly contributes more broadly to strengthening these tools.

Quantitative Methods

One hope has been that solutions can be found in refinements on regression analysis. This aspiration has motivated the new chapters by Jason Seawright and Thad Dunning (chaps. 13 and 14), which explore both some disasters of causal inference in quantitative research, and also potential solutions. They consider, for example, matching designs and the family of techniques associated with natural experiments—including regression discontinuity designs and instrumental variables. In some substantive domains, as Seawright shows, these tools are of little help, especially in macro-comparative analysis. He urges scaling down to more modest frameworks of comparison that potentially incorporate a substantial use of qualitative evidence.

Dunning points to the potentially large contribution of natural experiments—which, in his examples, focus entirely on much smaller-scale comparisons. At the same time, Dunning underscores severe trade-offs that may arise in employing these research designs, and both he and Seawright make clear that perhaps too often, these methodological tools do not escape the confines of regression analysis to the degree that many methodologists hope they will.

Qualitative Methods

Another avenue is opened by further refinements in qualitative tools. A familiar, traditional option here is typically called the small-N comparative method, a strategy common in research that entails both cross-national comparisons and comparison of political units within nations—whether they be regions, provinces or states, or metropolitan areas. Here, the analyst juxtaposes two, or four, or perhaps six cases, with a central idea often being to set up matching and contrasting cases in a way that is seen as "controlling" for extraneous factors and allowing a focus on the principal variables of concern. This approach is often identified with J. S. Mill's (1974 [1843]) methods of agreement and difference, and with Przeworski and Teune's (1970) most similar and most different systems designs.

In our view, this small-N comparative approach is truly invaluable in concept formation and in formulating explanatory ideas (see chap. 1 and online chapters of this book). It is much weaker as a basis for causal inference. It involves, after all, what is in effect a correlation analysis with such a small N that it is not an appropriate basis for evaluating causal claims. The matching and contrasting of cases employed probably cannot succeed, by itself, in controlling for variables that the researcher considers extraneous to the analysis.

Rather, as is well known, the key step is to juxtapose this comparative framing with carefully-executed analysis carried out within the cases. The challenge, therefore, is to find strong tools of within-case analysis.

Correspondingly, the objective of chapters 10, 11, and 12, by Andrew Bennett, David A. Freedman, and Henry E. Brady, is to systematize and refine the tools of process tracing and causal-process observations. Through a new typology of process tracing, along with many examples, both macro and micro, we seek to place these procedures of qualitative analysis on a more secure foundation, thereby strengthening their value and legitimacy as procedures for causal inference. To reiterate, these chapters are accompanied by exercises posted with the online materials for this book.

In sum, our objective in the second edition is to sustain a clear-eyed awareness of limitations inherent in standard inferential tools; and to push forward in strengthening these tools, both quantitative and qualitative.